

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

На правах рукопису

МИТНИК Олег Юрійович

УДК 004.855:681.518

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ СИНТЕЗУ РОБАСТНИХ
НЕЙРОНЕЧІТКИХ МОДЕЛЕЙ СТОХАСТИЧНИХ ПРОЦЕСІВ

Спеціальність: 05.13.06. — Інформаційні технології

Д и с е р т а ц і я
на здобуття наукового ступеня
кандидата технічних наук

Науковий керівник:
Бідюк Петро Іванович
доктор технічних наук,
професор

Київ–2008

ЗМІСТ

ВСТУП	8
1 ПРОБЛЕМИ ПРОГНОЗУВАННЯ СТОХАСТИЧНИХ ПРОЦЕСІВ	22
1.1. Задача прогнозування стохастичних процесів	22
1.1.1. Поняття емпіричного прогнозування стохастичних процесів	22
1.1.2. Представлення стохастичного процесу як нелінійного динамічного процесу з шумом	25
1.1.3. Помилки ідентифікації нелінійних динамічних систем з шумом	27
1.2. Методи розв'язання некоректних задач навчання	30
1.2.1. Принцип мінімізації структурного ризику	30
1.2.2. Регуляризація Тихонова	33
1.2.3. Байєсівський підхід до навчання	36
1.2.4. Перевірка моделей на адекватність	44
1.3. Структурованість моделі. Розклад аналізу варіацій	49
1.3.1. Описання динамічних процесів за допомогою ряду Вольтера	50
1.3.2. Метод групового врахування аргументів	52
1.4. Висновки і постановка задач роботи	57
2 РОЗВИТОК НЕЙРОНЕЧІТКИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ АНАЛІЗУ СТОХАСТИЧНИХ ПРОЦЕСІВ	59
2.1. Нечітка логіка в нейронних моделях	59
2.2. Нейронечітка модель у формі Бернштейна	62
2.2.1. Алгоритм побудови нейронечіткої моделі у формі Бернштейна	62
2.2.2. Класичний алгоритм Кастельжо	65
2.2.3. Обернений алгоритм Кастельжо	66
2.3. Швидке навчання нейронечіткої моделі у формі Бернштейна	69
2.3.1. Швидке обернене відображення Кастельжо	69
2.3.2. Аналітичний метод визначення барицентричних координат	72
2.4. Оптимальне навчання нейронечіткої моделі у формі Бернштейна	74
2.4.1. Регуляризація поліномів у формі Бернштейна	74
2.4.2. Застосування еволюційних методів для визначення оптимального розподілу базових контрольних точок	76
2.4.3. Деякі оцінки оптимальностей	78
2.5. Збалансовані нейронечіткі моделі	83
2.6. Висновки до розділу	87

3	ТЕХНОЛОГІЯ ОПОРНИХ ВЕКТОРІВ. ПОБУДОВА БАЙЄСІВСЬКОЇ РЕГРЕСІЇ ОПОРНИХ ВЕКТОРІВ У ФОРМІ БЕРНШТЕЙНА	89
3.1.	Регресія опорних векторів. Робастність	89
3.1.1.	Лінійна регресія опорних векторів	90
3.1.2.	Нелінійна регресія опорних векторів	91
3.1.3.	Зведення РОВ до задачі квадратичного програмування	95
3.2.	Розвиток і застосування механізму опорних векторів	96
3.3.	Індуктивний метод побудови байєсівської моделі регресії опорних векторів у формі Бернштейна. Алгоритм ПРИАМ	97
3.3.1.	Байєсівська регресія опорних векторів	97
3.3.2.	Критерій байєсівського підтвердження адекватності моделі	100
3.3.3.	Схема алгоритму ПРИАМ. Характеристичний простір поліноміальних функцій Без'є-Бернштейна.	109
3.3.4.	Апостеріорний прогнозовий розподіл. Довірчі інтервали	111
3.4.	Висновки до розділу	114
4	РОЗРОБКА ТА ЗАСТОСУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ПРОГНОЗУВАННЯ СТОХАСТИЧНИХ ПРОЦЕСІВ	115
4.1.	Розробка інформаційної системи обробки даних спостережень	115
4.1.1.	Концептуальний рівень інформаційної системи	115
4.1.2.	Логічний рівень інформаційної системи	117
4.1.3.	Фізичний рівень інформаційної системи. Програмна реалізація	121
4.2.	Апробація ПРИАМ на еталонних моделях	122
4.2.1.	Модель Лонглі (Longley)	123
4.2.2.	Модель Філіппеллі (Filippelli)	125
4.2.3.	Модель Фрідмана (Friedman)	126
4.2.4.	Модель АМРГ	128
4.3.	Апробація ПРИАМ на спрощеній моделі макроекономіки України	128
4.3.1.	Моделювання індексу споживчих цін	129
4.3.2.	Моделювання реального споживання	130
4.4.	Апробація ПРИАМ на гідрометеорологічній моделі вітрових хвиль	131
4.5.	Апробація ПРИАМ на екологічній моделі біомаси макрозообентосу	134
4.6.	Висновки до розділу	135
	ЗАГАЛЬНІ ВИСНОВКИ	138
	СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	140
А	РЕЗУЛЬТАТИ ПОБУДОВИ ПРОГНОЗІВ ЗА ДОПОМОГОЮ ПРИАМ	164
Б	ІНФОРМАЦІЙНА СИСТЕМА ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ	174
В	МАТЕМАТИЧНИЙ ДОДАТОК	177
Г	ДОКУМЕНТИ, ЯКІ ПІДТВЕРДЖУЮТЬ АПРОБАЦІЮ ОСНОВНИХ ВИСНОВКІВ І РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ	181

ПЕРЕЛІК ІЛЮСТРАЦІЙ

1	Підкріплення експертних знань технічним аналізом.	9
2	Розвиток інтелектуальних інформаційних технологій.	13
3	Гібридизація інтелектуальних інформаційних технологій.	17
1.1	Процес і модель.	27
1.2	Помилки моделювання.	28
1.3	Верхня межа справжнього ризику.	32
1.4	Контури правдоподібності і власні вектори в масштабі власних чисел для $\mathbf{A}^\top \mathbf{A}$ з центром \mathbf{w}_{ml} , для $\mathbf{\Omega}_a^{-1}$ з центром \mathbf{w}_0 , для \mathbf{H} з центром \mathbf{w}_{reg}	35
1.5	Порівняння величини байєсівського підтвердження для різних моделей.	37
1.6	Байєсівське навчання.	39
1.7	Якість моделі в залежності від гіперпараметрів.	43
1.8	Циклічна структура побудови моделі.	49
2.1	Структура нейронечіткої моделі К.Харріса.	64
2.2	Побудова кривої Без'є методом Кастельжо.	66
2.3	Робота алгоритму Кастельжо для поверхні Без'є.	67
2.4	Обчислення контрольних точок для побудови проекції поверхні Без'є на площині.	68
2.5	Симетричне розташування контрольних точок проекції поверхні Без'є на площині.	71
2.6	Оцінка базисних поліномів Бернштейна 4-го порядку зверху.	79
2.7	Оцінка сліду інформаційної матриці Фішера.	81
3.1	ϵ -нечутлива смуга регресії опорних векторів.	92
3.2	Графік першої похідної для $\epsilon = 0.2$, $\eta = 0.05$	101
3.3	Вибір відхилення $\Delta\epsilon$ для заданої вибірки. Опорні вектори зображені зафарбованими квадратами.	103
3.4	Порівняння оцінок підтвердження методом Монте-Карло (МК) та методом Лапласа (КБП).	108
3.5	Порівняння моделі шуму БРОВ (1) з гіперпараметрами $\beta = 15$, $\epsilon = 0.06$ та моделі нормального шуму (2) з нульовим середнім та дисперсією $\sigma_N^2 = 0.0114$	113
4.1	Концептуальна модель інформаційної системи. Блоки у вигляді прямокутників відображають ручні процеси або процедури. Овальна форма блоків відповідає автоматичним операціям.	116
4.2	Інформаційні потоки системи.	117
4.3	Схема взаємодії моделей інформаційної системи.	118

4.4	Діаграма пакетів (бібліотек) системи.	119
4.5	Діаграма взаємодії користувача з аплетом.	120
4.6	Діаграма топологій інформаційної системи.	121
4.7	Схема мережі інформаційної системи.	123
4.8	ПРІАМ для моделі Лонглі: 1 — прогноз, 2 — реальні дані, 3 — опорні вектори, 4 — 95% довірчий інтервал.	124
4.9	ПРІАМ для моделі Філіппеллі.	125
4.10	ПРІАМ для спрощеної моделі Фрідмана.	127
4.11	ПРІАМ для індексу споживчих цін.	130
4.12	ПРІАМ для реального споживання.	131
A.1	Залишки для моделі Лонглі: 1 — ПРІАМ, 2 — МГВА, 3 — РНМ, 4 — НМГВА, 5 — смуга нечутливості.	165
A.2	Залежність (а) КБП і (б) СКПн моделі Лонглі від ширини смуги нечутли- вості ϵ для різних значень гіперпараметру β	165
A.3	Залишки для моделі Філіппеллі.	167
A.4	Залежність (а) КБП і (б) СКПн моделі Філіппеллі від ширини смуги нечу- тливості ϵ для різних значень гіперпараметру β	167
A.5	Залишки для спрощеної моделі Фрідмана.	169
A.6	Залежність (а) КБП і (б) СКПн спрощеної моделі Фрідмана від ширини смуги нечутливості ϵ для різних значень гіперпараметру β	169
A.7	Залишки для моделі ІСЦ.	171
A.8	Залежність (а) КБП і (б) СКПн моделі ІСЦ від ширини смуги нечутливості ϵ для різних значень гіперпараметру β	171
A.9	Залишки для моделі РСГ.	173
A.10	Залежність (а) КБП і (б) СКПн моделі РСГ від ширини смуги нечутливості ϵ для різних значень гіперпараметру β	173
B.1	Приклад побудови моделі (скріншот моделі у формі Бернштейна).	174
B.2	Приклад побудови моделі (скріншот ПРІАМ).	175
B.3	Приклад побудови моделі (скріншот аналізу залежності КБП і СКП).	176

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

НБЗ	Нечітка База Знань
ННМ	Нечітка Нейронна (Нейронечітка) Модель (Мережа)
РНМ	Рекурентна Нейронна Модель (Мережа)
МОВ	Механізм (Метод, Машина) Опорних Векторів
(Б)РОВ	(Байєсівська) Регресія Опорних Векторів
ν -РОВ	ν -Регресія Опорних Векторів
ПРІАМ	Поліноміальної РОВ Індуктивний Алгоритм Моделювання
МАП	Максимум АПостеріорного розподілу
ASMOD	Adaptive Spline Modelling of Observational Data
ANFIS	Adaptive Network Based Fuzzy Inference System
ANOVA	ANalysis Of VAriance
SUPANOVA	SUPport vector Parsimonious ANOVA
(Н)МГВА	(Нечіткий) Метод Групового Врахування Аргументів
(О)МНК	(Ортогональний) Метод Найменших Квадратів
СБКТ	Система Базових Контрольних Точок
УПП	Узагальнена Помилка Прогнозування
ІКМ	Інформаційний Критерій Мережі
ІКА	Інформаційний Критерій Акаїке
БІК	Байєсівський Інформаційний Критерій
БСМЗ	Байєсівська Статистична Міра Значущості
ОПП	Остаточна Помилка Прогнозування
МДО	Мінімальна Довжина Описання
ІСЦ	Індекс Споживчих Цін
РСГ	Реальне Споживання Господарств
КБП	Критерій Байєсівського Підтвердження
СКП(н)	Середньо Квадратична Помилка (в нормованому просторі)
САП(н)	Середня Абсолютна Помилка (в нормованому просторі)
СНАР	Система Нелінійних Алгебраїчних Рівнянь

R_{emp}	емпіричний ризик
R_{reg}	регуляризований ризик
$f \circ g$	згортка функцій f та g
$\langle \cdot, \cdot \rangle_{\mathcal{X}}$	скалярний добуток в просторі \mathcal{X}
$\ \cdot \ _{\mathcal{X}}$	норма в просторі \mathcal{X}
$\ \cdot \ $	евклідова норма
$(\cdot, \dots, \cdot)^{\top}$	вектор
\tilde{a}	оцінка параметру чи показника a в нормованому просторі
$ \cdot _{\epsilon}$	ϵ -нечутлива функція втрат Вапника
$E[\cdot]$	математичне сподівання
$D[\cdot]$	дисперсія
Ω_a	апріорна матриця коваріації параметрів
Ω_p	апостеріорна матриця коваріації параметрів
$a \longleftarrow b$	присвоєння змінній a значення b

ВСТУП

Прогнозування поведінки технічних процесів успішно виконується на основі диференційних рівнянь. Тоді як прогнозування економічних, екологічних, соціальних процесів або так зване інтелектуальне прогнозування виконується на основі правил, оскільки досвідчені менеджери приймають ефективні рішення на рівні практичних міркувань. Для побудови інтелектуального прогнозу доцільно використовувати експертні системи на основі нечітких баз знань. Такі бази знань використовують апарат нечіткої логіки, сила якого полягає у здатності створювати кількісне представлення для лінгвістичних змінних, а також ефективно відображати залежності між цими змінними у вигляді нечітких правил. Однак нечіткі бази знань, які побудовані виключно на основі експертної інформації, є суб'єктивними і повністю залежать від кваліфікації експертів. Тому необхідно підкріплювати висновки і рішення експертів технічним аналізом процесів. Для цього, після побудови апіорної моделі процесу у формі нечіткої бази знань, проводять навчання цієї моделі на емпіричних даних. Здатність навчатися на даних притаманна нейроматематичним технологіям. Тож гібридизація нечітких та нейроматематичних технологій в так звані нейронечіткі технології є важливим інструментом інтелектуального прогнозування. Нейронечіткі технології дозволяють перетворювати апіорні нечіткі правила в аналітичні моделі, навчати ці моделі, і перетворювати їх знову в нечіткі правила, які обґрунтовані технічним аналізом. Цей процес показано на рис. 1. Проблема навчання моделей процесів з людським фактором це проблема відновлення стохастичних

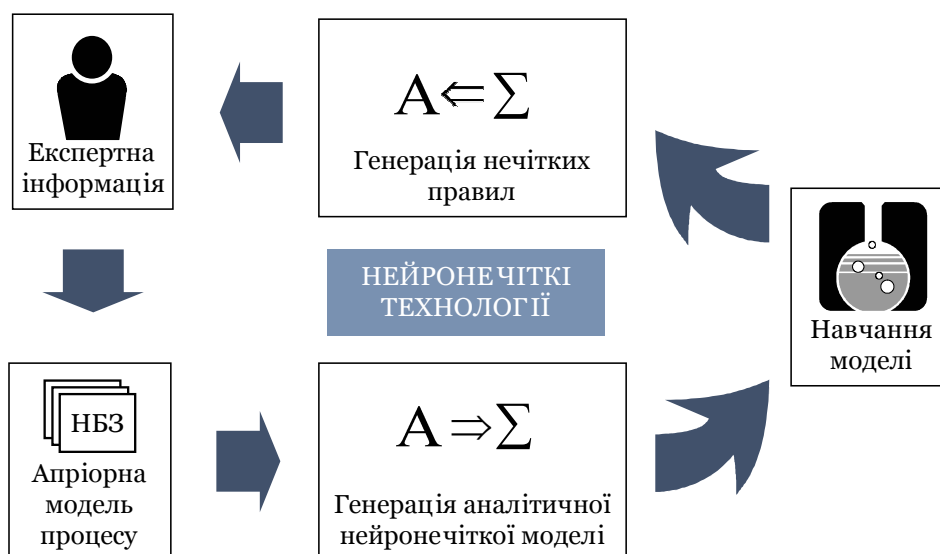


Рис. 1. Підкріплення експертних знань технічним аналізом.

залежностей. Тож дисертаційна робота присвячена дослідженню некоректних задач¹ відновлення стохастичних залежностей на основі емпіричних даних, побудові нейронечітких моделей за допомогою інтелектуальних інформаційних технологій, та практичному застосуванню нових моделей для синтезу нечітких баз знань та прогнозування стохастичних процесів.

Можна виділити три основні групи інформаційних технологій синтезу моделей стохастичних процесів та інтелектуального аналізу даних. З точки зору теорії суб'єктивної ймовірності — байєсівські технології. З точки зору теорії статистичного навчання — технологія опорних векторів. З точки зору теорії адаптивного моделювання — нейронечіткі технології. Розвиток ймовірнісного підходу до аналізу процесів, засновниками якого вважаються Т.Байєс [1] і П.С.Лаплас [2], привів до появи ряду видатних праць. Серед них фундаментальна робота Л.Дж.Севіджа [3], написана під впливом теорії ігор Джона фон Неймана, дала поштовх теорії суб'єктивної ймовірності, на якій ґрунтується сучасна байєсівська статистика і так званий байєсівський індуктивний висновок або байєсівська індукція. Прикладами сучасних ін-

¹Під некоректною задачею (англ. ill-posed) слід розуміти задачу некоректно поставлену за Адамаром.

дуктивних алгоритмів є простий байєсівський класифікатор Ч.Елкана [4] і байєсівські мережі [5, 6]. Історія становлення байєсівської школи, яка розглядає ймовірність як ступінь впевненості по відношенню до випадкової події і пов'язана з іменами Дж.Кейнса, Г.Джефрі, Ф.Рамсея, А.Вальда, досить повно описана у Г.Райфи [7]. Подальші дослідження Е.Джейнса [8] в цьому напрямку стосуються проблеми “об’єктивного” вибору апіорних ймовірностей байєсівського висновку, зокрема на основі принципу максимальної ентропії. Аналогічний байєсівський підхід до ідентифікації систем, основи якого викладені, наприклад, в роботі В.Петерки [9], знайшов своє продовження в роботах С.Гала і Дж.Скілінга [10, 11], де вперше була сформульована концепція байєсівського підтвердження параметрів апіорних розподілів ймовірностей для задач регресійного аналізу. Потім Дж.Хопфілд і Д.Маккей в своїх роботах [12, 13, 14, 15] використали і розвинули байєсівський принцип підтвердження з метою регуляризації штучних нейронних мереж. В наш час байєсівські технології активно використовуються пошуковими системами мережі Інтернет та системами фільтрації електронної пошти.

Початком принципово іншого підходу до аналізу процесів стала статистична теорія навчання В.Вапника та А.Червоненкіса [16, 17, 18, 19], також відома як ВЧ теорія, яка намагається пояснити процес навчання зі статистичної точки зору. Головною метою статистичної теорії навчання є вивчення проблем індуктивного висновку, тобто отримання знань, прогнозування, передбачення, прийняття рішень, а також побудова моделей на основі заданої вибірки даних. В рамках цієї теорії сформульований принцип мінімізації структурного ризику, який став передумовою створення Б.Босером, І.Гійон, В.Вапником, К.Кортес [20, 21] машини опорних векторів для задач класифікації. Незабаром метод опорних векторів був поширений А.Смолюю, С.Головічем, С.Гуном, Б.Шолькопфом [22, 23, 24, 25] на задачі побудови робастних регресійних моделей і розроблена так звана регресія опорних векторів. Для автоматичного вибору гіперпараметрів запропонована ν -регресія

опорних векторів [26, 27]. Пізніше К.-Р.Мюллер, А.Смола, Б.Шолькофф [28] застосували метод опорних векторів для прогнозування часових рядів.

Теорія опорних векторів розвивалася як частина більш загальної теорії гаусівських процесів [29, 30, 31, 32, 33], де апріорна інформація про модель задається у вигляді багатовимірного нормального розподілу вектору невідомих параметрів. Сучасний стан цього напрямку представлений в роботах М.Кусса, К.Расмуссена, К.Вільямса [34, 35, 36]. Так, в рамках теорії гаусівських процесів отримано ряд важливих результатів. Зокрема М.Сігер [37] розробив варіаційний байєсівський метод для вибору оптимальної моделі машини опорних векторів. Використавши байєсівську індукцію Маккея для підтвердження гіперпараметрів методу опорних векторів, М.Лоу та Дж.Квок [38] започаткували байєсівську регресію опорних векторів. Незалежно від них, на основі тих самих ідей Маккея, М.Тіпінг [39] розробив машину значущих векторів. Для подолання проблеми гладкості ϵ -нечутливої функції втрат і забезпечення існування гессіану регресії при підтвердженні гіперпараметрів, В.Чу та С.Кіртті [40] запропонували використовувати згладжену функцію втрат. На основі ідей машини опорних векторів в останні роки активно розвиваються локально регуляризована регресія найменших квадратів Ш.Чена [41] та узагальнені ядерні моделі [42].

Нейротехнології беруть свій початок з фундаментальної роботи В.МакКаллока і В.Піттса [43] в області моделювання нервової системи. Саме з цієї роботи почався розвиток штучного інтелекту та нейронних мереж, який пов'язаний з іменами Джона фон Неймана, В.М.Глушкова, О.А.Павлова, О.Г.Івахненка, Н.Вінера, Б.Уїдроу, Ф.Розенблата, Д.Хебба, М.Мінського, Дж.Альбана Т.Кохонена, Д.Румельхарта та Дж.Хопфілда. Сучасний огляд цього напрямку висвітлений в монографіях С.Хайкіна [44], К.Харріса [45], Д.Маккея [46], Г.Г.Малинецького [47]. Альтернативні дослідження в області штучного інтелекту призвели до появи експертних систем [48], які віддають перевагу накопиченню спеціальних знань про предметну область, а не пошуку загальних

методів розв'язання широкого класу задач. Новим етапом розвитку штучного інтелекту стала концепція нечіткої логіки, запропонована Л.Заде [49] і пізніше обґрунтована М.Брає та Д.Разерфордом. На основі цих ідей Е.Мамдані та його учні [50] розробили адаптивний нечіткий контролер, який пізніше удосконалили Т.Такагі і М.Сугено [51]. Після того як Д.Брумхед застосував радіальні базисні функції в узагальнених лінійних адаптивних мережах, став очевидним аналітичний потенціал нейронних моделей з асоціативною пам'яттю. Невдовзі з'явилась робота М.Брауна і К.Харріса [52], де вперше було встановлено зв'язок між нейронними моделями і нечіткою логікою в так званих нейронечітких адаптивних мережах, які володіють лінгвістичною прозорістю нечіткої логіки в поєднанні з аналітичною зручністю нейронних моделей. Типовими представниками таких нейронечітких архітектур стали багатовимірні адаптивні сплайни (MARS) Дж.Фрідмана [53], система адаптивного нечіткого виводу (ANFIS) Р.Джанга [54] на основі нечіткого контролеру Такагі-Сугено, адаптивний сплайновий алгоритм моделювання (ASMOD) Т.Кавлі [55, 56] на основі В-сплайнів, метод опорних векторів на основі аналізу варіацій² (SUPANOVA) С.Гуна [57]. Оскільки мережі з асоціативною пам'яттю потерпають від прокляття вимірності, то подальші дослідження адаптивних алгоритмів спрямовані на вирішення саме цієї проблеми. Зокрема, К.Харріс і К.Хонг [58, 59] запропонували узагальнений нейронечіткий алгоритм моделювання, який використовує поліноми у формі Бернштейна для адитивної декомпозиції аналізу варіацій. Виникнення і розвиток цього напрямку частково зображено на рис. 2. Сучасні вітчизняні дослідження нечітких нейронних мереж представлені в дисертаційних роботах Д.В.Новицького [60], Ю.І.Мітюшкіна [61], В.В.Колодяжного [62], В.А.Тазетдінова [63], В.В.Трипольської [64]. Приклад практичного застосування вище згаданих технологій штучного інтелекту в проектуванні інтелектуальних систем підтримки прийняття рішень можна знайти у Г.Сетлак [65] а також в

²У вітчизняній літературі зустрічається також термін дисперсійний аналіз

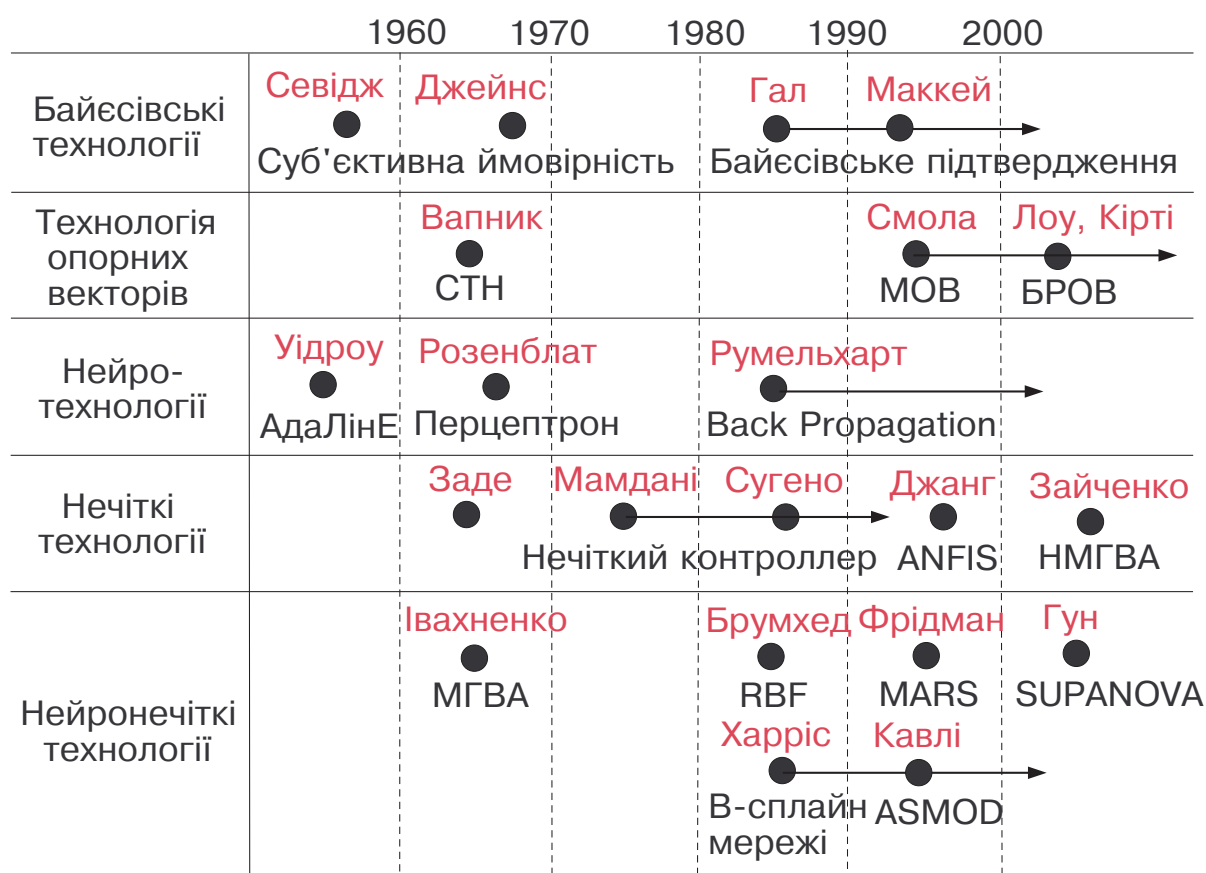


Рис. 2. Розвиток інтелектуальних інформаційних технологій.

роботі С.М.Побережника [66].

Розвиток теорії вейвлетів (сплесків) [67, 68], теоретичні основи і практичне застосування якої викладені в оглядових статтях і монографіях [69, 70, 71, 72, 73], також сприяв появі нових підходів в системній ідентифікації. Так, в роботах Й.Кітади [74], Р.Праденіча [75], З.Хасієвича [76] розглянуті проблеми ідентифікації нелінійних динамічних систем за допомогою вейвлетів. М.Туїлард в своїх роботах [77, 78] застосував нечітку логіку в контексті теорії вейвлетів і ввів так звані нечіткі вейвлет-мережі. Пізніше А.Попула [79] запропонував метод нечітких вейвлетів для аналізу нестационарних часових рядів, в якому за допомогою багатомасштабного вейвлет-аналізу генерується прогнозуюча система нечітких правил.

Робота виконана в навчально-науковому комплексі “Інститут прикладного системного аналізу” Національного технічного університету України

“Київський політехнічний інститут” відповідно до плану науково-дослідних робіт: “Розробка якісно-кількісного методу ситуаційного аналізу на основі байєсівських мереж” (№ДР 0103U000528), “Розробка та впровадження автоматизованої інформаційної системи підтримки прийняття рішень в управлінні проектами” (№ДР 0102U000245).

Мета роботи — підвищити якість прогнозування стохастичних процесів за допомогою нових інформаційних технологій та методів розв’язання некоректних задач відновлення залежностей. Конкретні завдання дослідження, які треба вирішити для досягнення поставленої мети, сформульовані в кінці першого розділу. Об’єкт дослідження — стохастичні процеси. Предмет дослідження — нові інформаційні технології та методи розв’язання некоректних задач відновлення стохастичних залежностей.

В дисертаційній роботі отримані наступні нові результати:

1. Розвинуто метод побудови нейронечітких моделей у формі Бернштейна. Для визначення барицентричних координат новий прискорений метод використовує швидке обернене відображення Кастельжо. Вперше запропоновано еволюційний метод побудови робастних нейронечітких моделей у формі Бернштейна, де для визначення барицентричних координат використовується оптимальне обернене відображення Кастельжо.
2. Створена нейронечітка інформаційна технологія синтезу нечітких баз знань квадратичної складності на основі так званих збалансованих нейронечітких моделей у формі Бернштейна.
3. На основі наближення маргінальної правдоподібності методом Лапласа отримано новий критерій адекватності байєсівської регресії опорних векторів, який, на відміну від відомих критеріїв, не порушує робастності моделей та має просту аналітичну форму.
4. Розроблено новий індуктивний метод побудови збалансованих робастних нейронечітких моделей у формі Бернштейна (ПРІАМ) на основі байєсів-

ської регресії опорних векторів у характеристичному просторі поліноміальних функцій Без'є-Бернштейна, який відрізняється від аналогів підвищеною якістю прогнозування стохастичних процесів з домінуючою випадковою складовою і можливістю задавати апріорну структуру моделі.

5. Створено нову інформаційну систему обробки статистичних даних, математичний апарат якої реалізує запропоновані методи. Унікальність системи полягає в її здатності оперувати в онлайн режимі (в мережі Інтернет).

Результати дисертації отримані на основі теорії ймовірності, статистичного оцінювання і навчання, оптимального планування, нечітких множин, математичного програмування, чисельних методів. Програмну реалізацію запропонованих методів та алгоритмів виконано в середовищах MATLAB 6.1, JDK 1.5.0.4. Для порівняння отриманих результатів використано пакети “NeuroShe” “GMDH Modeler 0.9.37” та “NeuroSolutions 5”.

Практичне значення одержаних результатів полягає в тому, що запропоновані автором інформаційні технології та методи можуть бути використані в системах підтримки прийняття рішень при математичному моделюванні і прогнозуванні стохастичних процесів з домінуючою випадковою складовою. Зокрема отримані результати впроваджені в Севастопольську гідрометеорологічну обсерваторію МНС України.

Достовірність отриманих нових наукових положень та висновків обґрунтована аналітичними доведеннями теорем та тверджень, а також порівняльним аналізом експериментальних результатів із оцінкою їх статистичної значимості.

Всі наукові результати, викладені в дисертації, отримані автором самостійно. В написаних у співавторстві роботах здобувачеві належать: швидке обернене відображення Кастельжо [88]; еволюційний метод визначення положення максимумів першої похідної спектру відбиття листків рослин-

ності [89, 90]; новий математичний метод аналізу даних гідробіологічного моніторингу [93].

Основні результати дисертаційної роботи доповідалися і обговорювалися на міжнародних наукових конференціях “Системний аналіз та інформаційні технології” (Київ, 2003-2004 рр.) [80, 81], “Стан та перспективи розвитку новітніх науково-освітніх комп’ютерних технологій” (Миколаїв, 2003 р.) [82], “Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій” (Євпаторія, 2005 р.) [83], “Моделювання та дослідження стійкості динамічних систем” (Київ, 2005, 2007 рр.) [84, 85], “Интеллектуальный анализ информации” (Київ, 2006 р.) [86], на одинадцятій міжнародній науковій конференції імені академіка М.Кравчука (Київ, 2006 р.) [87].

Основні положення дисертації викладені в 14 друкованих роботах, з яких 5 у фахових виданнях [88, 89, 90, 91, 92] ([93] — біологічні науки).

Дисертація складається із вступу, чотирьох розділів, висновків, списку літератури з 212 джерел на 24 сторінках, чотирьох додатків. Загальний обсяг роботи становить 183 сторінки, з яких 134 сторінки основного тексту.

Зміст дисертаційної роботи

Основна ідея дисертаційного дослідження — об’єднати переваги сучасних інформаційних технологій інтелектуального аналізу даних в новій гібридній технології (рис. 3). Зокрема, нейронечіткі технології забезпечують прозору інтерпретацію моделей на мові нечіткої логіки і синтез нечітких баз знань. Для знаходження адекватної структури нейронечіткої моделі використаний байєсівський механізм регуляризації. А технологія опорних векторів дає робастні оцінки параметрів нейронечіткої моделі.

В першому розділі розглянуті основні проблеми емпіричного прогнозування, серед яких некоректність постановки самої задачі навчання за даними спостережень і слабка структурованість представлення моделей. Значна увага приділяється механізму регуляризації задачі навчання як засобу для



Рис. 3. Гібридизація інтелектуальних інформаційних технологій.

розкриття невизначеності відношення зміщення-варіації моделі. Розглянутий розклад аналізу варіацій як засіб забезпечення структурованості представлення моделі. Наведені типові приклади використання цього розкладу. Зокрема викладені основні положення методу групового врахування аргументів (МГВА) [94, 95] і його нечіткої модифікації (НМГВА) [96]. Перший розділ закінчується постановкою задач дисертаційного дослідження.

Другий розділ присвячений дослідженню запропонованих К.Харрісом нейронечітких моделей у формі Бернштейна, які мають за основу розклад аналізу варіацій від поліномів у формі Бернштейна. Розклад обмежений парною взаємодією факторів і використовується при визначенні індивідуальних впливів одного або пари вхідних факторів на вихід системи. Цей підхід позбавлений прокляття вимірності нечітких контролерів і гарантує поліноміальну складність нечіткої бази знань, оскільки нечіткі правила генеруються

на основі окремих нейронечітких підмоделей. Такими підмоделями є поліноми у формі Бернштейна. При цьому функції належності це базисні поліноми Бернштейна від барицентричних координат. Барицентричні координати знаходяться за допомогою оберненого відображення Кастельжо.

В результаті дослідження нейронечітких моделей у формі Бернштейна виявлені такі недоліки. Обернене відображення Кастельжо використовує ітераційний чисельний метод зворотного розповсюдження помилки, що значно сповільнює процес навчання. Конфігурація контрольних точок не визначена однозначно. Не обґрунтована можливість незалежного використання нечітких правил на основі окремих підмоделей. В роботі розроблено прискорений метод навчання нейронечітких моделей у формі Бернштейна і виведена чітка схема розміщення контрольних точок, яка забезпечує швидке навчання нейронечітких моделей у формі Бернштейна. Досліджено також питання про оптимальне, за заданим критерієм, розміщення контрольних точок. Відповідно введено ряд означень оптимальних систем контрольних точок, які підвищують робастність нейронечітких моделей у формі Бернштейна. Для розв'язання таких оптимізаційних задач в загальному вигляді запропоновано використовувати еволюційні методи, зокрема генетичний алгоритм. Для обґрунтування незалежного використання нечітких правил сформованих на основі окремих підмоделей автором вводяться так звані збалансовані нейронечіткі моделі. Виведено достатню умову збалансованості нейронечіткої моделі. Показано, що нейронечіткі моделі у формі Бернштейна можна звести до збалансованих. Клас збалансованих нейронечітких моделей дозволяє робити прогноз в умовах часткової або навіть повної невизначеності вхідних факторів.

В третьому розділі викладені основи РОВ. Розглянута байєсівська РОВ, де основною проблемою є обчислення підтвердження моделі. Асимптотично точну оцінку підтвердження моделі можна отримати за допомогою методу Монте-Карло. Досить точну оцінку підтвердження дає метод розповсюджен-

ня сподівань. В роботі використано метод Лапласа, який хоч і є неточним в порівнянні з вище згаданими методами, проте потребує значно менші обчислювальні витрати. Більш глибокий аналіз методу Лапласа показує, що цей метод систематично переоцінює підтвердження і відтак не може бути використаний для точного підрахунку підтвердження. Але, оскільки метод Лапласа переоцінює підтвердження систематично для всіх моделей, він може бути використаний для швидкого порівняння моделей. Проте для застосування методу Лапласа необхідно подолати недиференційовність функції втрат в критичних точках, які лежать на верхній та нижній межах ϵ -нечутливої смуги. Тому, як правило, застосовують різноманітні згладжені аналоги. Але вони порушують розрідженість та робастність оцінок параметрів РОВ і призводять до складного байєсівського висновку. В роботі запропоновано використовувати локально згладжену в нескінченно малому околі критичних точок функцію втрат. І отримано критерій байєсівського підтвердження (КБП) адекватності моделей РОВ. Задача максимізації підтвердження зводиться до задачі мінімізації КБП. Оскільки КБП залежить від гіперпараметрів, то в межах тієї самої моделі можна знайти оптимальні значення гіперпараметрів. Цю задачу розв'язано рефлексивним методом Ньютона, де на кожному кроці будується РОВ і знаходиться МАП оцінка для вектору параметрів. Розроблений метод побудови довірчих інтервалів, який ґрунтується на оцінюванні дисперсії апостеріорного прогнозного розподілу.

Основні переваги байєсівської РОВ полягають в наступному. Розклад опорних векторів не залежить від розмірності вхідного простору. Задача квадратичного програмування дає єдиний розв'язок. Модель РОВ робастна. Байєсівський підхід дозволяє робити оцінку для довірчих інтервалів. КБП забезпечує високу швидкість перебору моделей. Серед недоліків слід відзначити неточне обчислення підтвердження, що змушує нас вибирати модель із значно меншим КБП. Також відзначимо, що при виборі моделі з найбільшою апостеріорною ймовірністю та виборі її параметрів за МАП оцінкою,

порушується класичний байєсівський висновок. Далі в роботі досліджені питання вибору простору моделей і характеристичного простору. З метою поєднання переваг байєсівської РОВ із зручністю нейронечіткого підходу Харріса в якості характеристичного простору використано простір поліноміальних функцій Без'є-Бернштейна, де для обчислення барицентричних координат використовується швидке обернене відображення Кастельжо. Введено поняття конфігурації характеристичного простору поліноміальних функцій Без'є-Бернштейна у вигляді верхньотрикутної матриці, де кожен елемент відображає ступінь впливу фактору або пари факторів на вихідну змінну і визначає порядок базисних поліномів Бернштейна для відповідної поліноміальної функції. Основним результатом третього розділу є індуктивний метод побудови байєсівської РОВ в характеристичному просторі поліноміальних функцій Без'є-Бернштейна (або більш коротко — байєсівської моделі РОВ у формі Бернштейна), який визначає порядок перебору моделей в просторі. Приведена схема алгоритму ПРИАМ, який реалізує даний метод. Алгоритм проаналізований на збіжність, і виведена оцінка його складності. Показано, що ПРИАМ генерує збалансовані нейронечіткі моделі у формі Бернштейна.

Четвертий розділ присвячено розробці архітектури нової інформаційної системи обробки статистичних даних, яка використовує розроблені в роботі нові алгоритми та методи. Так, інформаційна система представлена на трьох рівнях: концептуальному, логічному та фізичному. На концептуальному рівні визначені інформаційні процеси та процедури. На логічному рівні описана взаємодія формалізованих моделей інформаційних процесів. На фізичному рівні створені підсистеми, які реалізують процеси обробки, накопичення даних та представлення знань. В роботі приведені діаграми потоків, пакетів, послідовності роботи, топології, та схема мережі інформаційної системи. Також наведені результати практичного застосування реально створеного прототипу такої інформаційної системи для відомих штучних та реальних еталонних моделей, для спрощеної моделі макроекономіки України, для гід-

рометеорологічної моделі вітрових хвиль та гідробіологічної моделі біомаси макрзообентосу. Для кожної з моделей визначена оптимальна конфігурація характеристичного простору, побудований прогноз і довірчі інтервали, виконано порівняльний аналіз з такими провідними в області прогнозування алгоритмами як МГВА, НМГВА, ANFIS та рекурентні нейронні мережі (РНМ). Порівняльний аналіз підтверджує високу ефективність розробленого автором індуктивного методу.

У висновках сформульовані основні наукові результати дисертаційної роботи та наведені рекомендації щодо практичного використання отриманих результатів.

В додатку А зведені детальні результати побудови прогнозів за допомогою ПРИАМ та їх порівняльний аналіз із МГВА, НМГВА та РНМ. А саме, для кожної з моделей наведено її повний розв'язок у вигляді коефіцієнтів розкладу опорних векторів, ітерації байєсівського виводу оптимальних значень гіперпараметрів, залишки навчальної та прогнозовної вибірок, залежності КБП та середньо квадратичної помилки від ширини смуги нечутливості. Додаток Б містить скріншоти роботи розробленого прототипу інформаційної системи. В додатку В наведені відомі математичні означення, теореми та деякі доведення. Додаток Г містить акти впровадження результатів дисертаційної роботи.

РОЗДІЛ 1

ПРОБЛЕМИ ПРОГНОЗУВАННЯ СТОХАСТИЧНИХ ПРОЦЕСІВ

1.1. Задача прогнозування стохастичних процесів

1.1.1. Поняття емпіричного прогнозування стохастичних процесів

Під емпіричним моделюванням будемо розуміти статистичне узагальнення і наближення математичними формулами даних спостережень. Результатом емпіричного моделювання є емпірична модель. Відповідно емпіричне прогнозування це побудова прогнозу на основі емпіричної моделі. Якщо мова йде про деякий процес, то будемо говорити про емпіричне моделювання і прогнозування процесу, якщо про систему, яка представляє цей процес, то про ідентифікацію системи. Під системою будемо розуміти деяку сукупність взаємодіючих елементів, які доцільно розглядати як єдине ціле. Все, що не входить в дану систему, є по відношенню до неї зовнішнім середовищем. Система підлягає впливу зі сторони середовища і, в свою чергу, сама впливає на середовище. Перші впливи будемо називати вхідними факторами (сигналами, змінними), другі — вихідними факторами. Емпірична математична модель динамічної, тобто функціонуючої в часі, системи будується за результатами спостереження за вхідними і вихідними сигналами. Процедура такої побудови і називається ідентифікацією динамічних систем [97]. Класичне означення ідентифікації запропоноване Л.Заде у 1962 році: “Ідентифікація полягає у відшуванні за вхідними і вихідними сигналами деякої системи еквівалентної

до неї системи із деякого заданого класу” [9]. Більш строгі математичні означення динамічних систем можна знайти в збірнику робіт по математичній теорії систем [98]. Слід зазначити, що задача побудови модельних рівнянь заданими спостережень належить до класу обернених задач [99]¹. Як правило обернені задачі є некоректно поставленими (за Адамаром), оскільки для них не існує єдиного розв’язку.

В залежності від рівня апріорної інформації про модель процесу говорять про ідентифікацію в широкому розумінні або ж у вузькому. З метою охоплення широкого класу задач в роботі досліджується ідентифікація в широкому розумінні, коли точна структура моделі не відома. В роботі [101] ідентифікація в широкому розумінні визначається як отримання або уточнення на основі експериментальних даних математичної моделі цієї системи, яка виражена на мові того чи іншого математичного апарату. Оскільки дослідження будуть обмежені пасивною ідентифікацією, тобто ідентифікацією систем, для яких неможливо поставити експеримент, то мова буде йти не про експериментальні дані, а про дані спостережень. Типовий приклад таких систем — економічні, екологічні системи. Як правило такі системи є нестационарними, тобто властивості перетворення вхідних сигналів змінні в часі. Реакція нестационарної системи залежить як від величини вхідного сигналу, так і від моменту його проходження. До того ж значення вихідних функцій в деякий момент часу залежать не тільки від поточних, але й від попередніх значень вхідних функцій. Подібні системи називають інерційними, або системами з пам’яттю. Період часу пам’яті може бути різним. В залежності від величини часу пам’яті інерційні системи можна розділити на системи зі скінченною та нескінченною пам’яттю. Очевидно, що стаціонарна залежність між входом та виходом може бути нелінійною, тож досліджуються нелінійні моделі систем. За типом зв’язків системи також поділяють на стохастичні (ймовірнісні) і детерміновані (регулярні). Як правило для складних систем

¹ Будемо відрізняти обернені задачі побудови модельних рівнянь від задач відновлення вхідного сигналу по відомому вихідному, які також називають оберненими [100]

з багаторівневою ієрархічною структурою, які функціонують в умовах невідзначеності під впливом зовнішніх факторів, суттєвою є випадкова складова (надалі вважаємо, що складова шумів є адитивною і гомоскедастичною). Такі системи є стохастичними. Виділення систематичної (закономірної, детермінованої) складової стохастичного процесу полягає у відновленні функціональної залежності між вхідними та вихідними факторами. Надалі будемо розглядати стохастичні системи з одним вихідним фактором.

Узагальнена задача відновлення стохастичної залежності

Узагальнена задача відновлення стохастичної залежності формально поставлена у наступному вигляді. Нехай задані дані спостережень $\mathcal{D} = \{(y_j, \mathbf{x}_j) : j = 1, \dots, N\}$, де $\mathbf{x} = (x^1, \dots, x^n)^\top \in \mathcal{X} \subseteq \mathbb{R}^n$, $y \in \mathcal{Y} \subseteq \mathbb{R}$. Висувається гіпотеза про існування стохастичної залежності, відповідно до якої кожному вектору \mathbf{x} ставиться у відповідність число y , отримане за допомогою випадкового випробування за законом $P(y|\mathbf{x})$. Задача відновлення стохастичної залежності полягає у відшуванні умовної щільності розподілу $P(y|\mathbf{x})$ на основі скінченої вибірки даних \mathcal{D} . Дана обернена задача некоректно поставлена і зводиться до пошуку апостеріорного прогнозного розподілу $P(y|\mathbf{x}, \mathcal{D})$, який є наближенням реальної стохастичної залежності. Наскільки близьким буде таке наближення залежить від повноти даних спостережень. Емпіричне моделювання а також навчання в цьому випадку це процес побудови апостеріорного прогнозного розподілу $P(y|\mathbf{x}, \mathcal{D})$, або його математичного сподівання, тобто функції регресії $y(\mathbf{x}) = \mathbb{E}y_x = \int yP(y|\mathbf{x}, \mathcal{D})dy$. Процес побудови функції регресії також будемо називати відновленням регресії.

Постановка задачі в такому вигляді поєднує ідентифікацію систем в теорії керування та машинне навчання в теорії штучного інтелекту. Це дає нам можливість одночасно застосовувати методи досліджень обох теоретичних розділів кібернетики.

1.1.2. Представлення стохастичного процесу як нелінійного динамічного процесу з шумом

Як правило, детермінована складова стохастичного процесу це нелінійний динамічний процес. Розглянемо клас нелінійних динамічних систем, які в загальному випадку можна представити в просторі станів як

$$\begin{cases} \mathbf{z}(t+1) = \mathbf{g}(\mathbf{z}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) = \mathbf{h}(\mathbf{z}(t), \mathbf{u}(t)) \end{cases}, \quad (1.1)$$

де t — змінна часу, $\mathbf{z}(t) \in \mathcal{Z}$ — внутрішній стан системи, $\mathbf{u}(t) \in \mathcal{U}$ — вхідний сигнал, $\mathbf{y}(t) \in \mathcal{Y}$ — вихід системи, $\mathbf{g} : \mathcal{Z} \times \mathcal{U} \rightarrow \mathcal{Z}$ — нелінійна перехідна функція, $\mathbf{h} : \mathcal{Z} \times \mathcal{U} \rightarrow \mathcal{Y}$ — нелінійна функція виходу. Дослідження обмежуються випадком, де $\dim \mathcal{Y} = 1$, тобто $\mathbf{y}(t) \equiv y(t)$ — скалярна функція. Такі моделі ще називають моделями зі скалярним відгуком [97]. Для побудови моделей складних систем, для яких апріорі невідомі фізичні взаємозв'язки типу рівнянь Максвелла або ж інші причиноутворюючі процеси, безпосереднє використання (1.1) не є зручним. Якщо відомі інформація або дані про вхідні та вихідні сигнали в достатній кількості і якості (повноті), тоді невідомі взаємозв'язки можна наблизити емпіричними парадигмами, які є основою класичної статистичної теорії оцінювання. Показано [102], що якщо система (1.1) може бути описана у скінченновимірному просторі станів і коли вона знаходиться достатньо близько до своєї точки рівноваги, то її можна представити у вигляді моделі вхід-вихід:

$$y(t) = f(y(t-1), \dots, y(t-n_y), \mathbf{u}(t), \mathbf{u}(t-1), \dots, \mathbf{u}(t-n_u), \\ e(t-1), \dots, e(t-n_e), \mathbf{w}) + e(t),$$

де f — деяке нелінійне відображення, n_y , n_u , n_e — натуральні числа, які визначають запізнення реакції системи, \mathbf{w} — невідомий вектор параметрів.

Хоча був отриманий більш строгий критерій взаємних перетворень представлень системи між моделлю у просторі станів і моделлю вхід-вихід

для випадку нескінченно гладких функцій \mathbf{g} , \mathbf{h} [103], проте подальші дослідження в дисертаційній роботі не розглядають представлення у просторі станів. Альтернативним напрямком є цікава стаття [104], де запропоновано використання саме моделі простору стану для моделювання нейронної мережі як чорного ящика. Широкий спектр методів ідентифікації нелінійних систем в просторі станів представлено в роботі В.Вердульт [105].

Надалі, як правило, будемо мати справу з представленням нелінійних динамічних систем у вигляді наступного відображення:

$$y(t) = f(\mathbf{x}(t), \mathbf{w}) + e(t), \quad (1.2)$$

де $\mathbf{x}(t) = (y(t-1), \dots, y(t-n_y), \mathbf{u}(t), \dots, \mathbf{u}(t-n_u), e(t-1), \dots, e(t-n_e))^T$ — вектор регресорів. Це представлення узагальнює широкий спектр відомих моделей:

- Нелінійні регресійні моделі зі скінченим відгуком на імпульс, які використовують тільки регресори $\mathbf{u}(t-k)$. До цього класу зокрема відносяться моделі у вигляді непараметричного ряду Вольтера [106]:

$$y(t) = g^m(u(t), u(t-1), \dots, u(t-n_u)) + e(t),$$

де g^m — нелінійна функція степеню m .

- Нелінійні авторегресійні моделі, які використовують регресори $\mathbf{u}(t-k)$, $y(t-k)$. Типова модель — параметричний ряд Вольтера вигляду:

$$y(t) = g^m(u(t), u(t-1), \dots, u(t-n_u)) - a_1 y(t-1) - \dots - a_{n_y} y(t-n_y) + e(t).$$

До цього класу відносяться і нелінійні диференціальні рівняння [106]:

$$y(t) = b_0 u(t) + b_1 u(t-1) + \dots + b_n u(t-n_u) + g^m(y(t-1), \dots, y(t-n_y)) + e(t).$$

А також нелінійні моделі Вінера-Гамерштейна з нелінійним каскадом $z = g(u)$ у вигляді:

$$y(t) = -a_1 y(t-1) - \dots - a_{n_y} y(t-n_y) + \\ + b_1 g(u(t-1)) + \dots + b_{n_u} g(u(t-n_u)) + e(t).$$

- Нелінійні авторегресійні моделі з ковзним середнім, які використовують регресори $\mathbf{u}(t-k)$, $y(t-k)$, $e(t-k)$. На практиці для збільшення гнучкості таких моделей шум виноситься в залишки наступним чином:

$$y(t) = f(\mathbf{x}(t), \mathbf{w}) + C(q^{-1})e(t),$$

де C — деякий лінійний фільтр, q^{-1} — оператор запізнення реакції системи.

1.1.3. Помилки ідентифікації нелінійних динамічних систем з шумом

Нехай задана вибірка спостережень $\mathcal{D} = \{(y(t), \mathbf{x}(t)) : t = 1, \dots, N\}$ об'ємом N . В процесі ідентифікації нелінійна параметризована динамічна модель $\mathcal{M}(\mathbf{x}(t), \mathbf{w})$ намагається перетворити вхідний сигнал $\mathbf{u}(t)$ у вихідний $y(t)$, який спотворений шумовою складовою $e(t)$, за допомогою наближення $f(t) = \mathcal{M}(\mathbf{x}(t), \mathbf{w})$. Загальна концепція ідентифікації нелінійної динамічної системи зображена на рис. 1.1. Задачу такого наближення слід розділити на

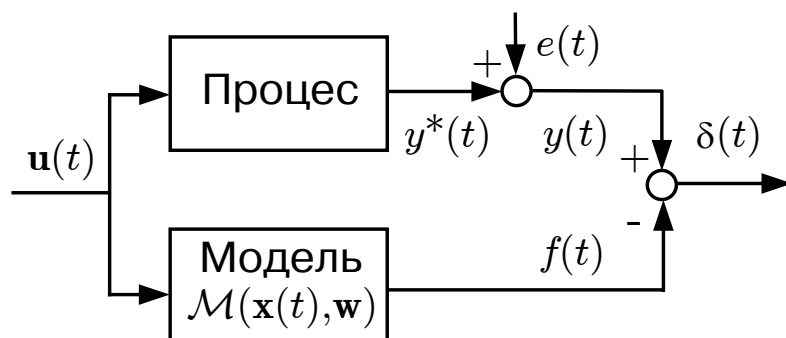


Рис. 1.1. Процес і модель.

дві категорії:

- ✓ Структурна ідентифікація. Це пошук оптимальної структури моделі \mathcal{M} , в тому числі оптимального числа параметрів моделі, що визначається розмірністю вектора \mathbf{w} . Як правило задача структурної ідентифікації пов'язана з проблемами комбінаторної оптимізації, складність яких швидко зростає із числом вхідних змінних.

- ✓ Параметрична ідентифікація. Якщо відома структура моделі і кількість параметрів, задача наближення зводиться до відшукування таких оптимальних значень цих параметрів, які дають найкраще за заданим критерієм наближення оцінки $f(t)$ реального процесу $y(t)$. Для цього мінімізується так званий емпіричний ризик $R_{emp}(f)$, про який буде йти мова далі.

Припустимо, що за допомогою структурної ідентифікації була вибрана деяка структура моделі \mathcal{M} , яка визначає простір (ансамбль) потенційних функцій $\mathcal{M} = \{f\}$. Кожній функції відповідає свій вектор параметрів \mathbf{w} . Оскільки простір, в якому знаходиться реальний процес $y(t)$, не співпадає і, як правило, містить в собі простір потенційних функцій, то виникає помилка наближення моделі, пов'язана з неточним вибором структури моделі. Крім цього виникає помилка оцінювання моделі пов'язана з вибором неоптимальних параметрів моделі. Таким чином узагальнена помилка складається з помилки наближення та помилки оцінювання. На рис. 1.2 зображені взаємозв'язки помилок. Наявність таких помилок пов'язана з дилемою зміщення-варіації (точності-

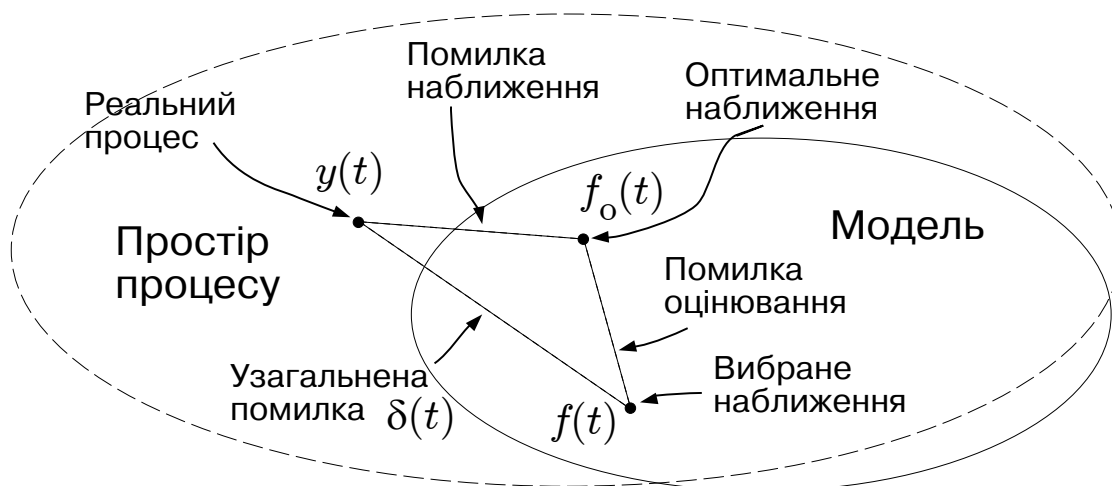


Рис. 1.2. Помилки моделювання.

гладкості) моделі, де зміщення моделі призводить до помилки наближення, а варіація моделі призводить до помилки оцінювання.

Паритет зміщеності та варіації моделі

Розглянемо математичне сподівання квадрату похибки наближення:

$$\begin{aligned} \mathbb{E}[f - y]^2 &= \mathbb{E}[f - \mathbb{E}f + \mathbb{E}f - y]^2 = \\ &= \mathbb{E}[f - \mathbb{E}f]^2 + \mathbb{E}[\mathbb{E}f - y]^2 + 2\mathbb{E}[(f - \mathbb{E}f)(\mathbb{E}f - y)]. \end{aligned}$$

Оскільки функція y детермінована і $\mathbb{E}[f - \mathbb{E}f] = 0$, то остаточно маємо:

$$\mathbb{E}[f - y]^2 = \underbrace{\mathbb{E}[f - \mathbb{E}f]^2}_{\text{варіація}} + \underbrace{\mathbb{E}[\mathbb{E}f - y]^2}_{\text{зміщення}}.$$

Перший доданок характеризує варіацію моделі, тобто показує наскільки модель чутлива до різних вибірок, або ж іншими словами — відображає коваріацію між вектором параметрів і оптимальним вектором параметрів. Другий доданок характеризує зміщення моделі, тобто показує як середня модель заданої структури відрізняється від реальної системи. Модель є незміщеною, якщо вона збігається до реальної системи при $N \rightarrow \infty$. В той час як варіація моделі пов'язана з її точністю, зміщення моделі пов'язано з її складністю [107].

Для зменшення середньо-квадратичної помилки необхідно зменшувати одночасно зміщення і варіацію моделі. Однак збільшення числа параметрів моделі приводить з одного боку до зменшення зміщення моделі, а з іншого до збільшення варіації моделі і можливо погіршує її прогностуючі властивості. При цьому модель стає перевизначеною надмірною кількістю параметрів. Звідси слідує відомий принцип ощадливості моделі: найкраща модель — це модель з найменшою кількістю ступенів свободи, яка дозволяє із заданою точністю генерувати вибірку даних.

З метою розкриття невизначеності зміщення-варіації моделі застосовуються різноманітні методи регуляризації некоректних задач ідентифікації, які будуть розглянуті в наступному підрозділі.

1.2. Методи розв'язання некоректних задач навчання

З точки зору теорії ймовірності задачу ідентифікації можна сформулювати як задачу навчання, де необхідно відтворити імплікації за їх наслідками. В подальшому, задачі ідентифікації нелінійних динамічних систем також будемо називати задачами (машинного) навчання. При цьому для стаціонарних систем очевидна повна відповідність, якщо дискретну змінну часу представити індексом у впорядкованій вибірці даних, а для нестационарних систем змінну часу можна віднести до додаткового вхідного фактору. Тож надалі позначення залежності змінних від часу ми вживати не будемо.

1.2.1. Принцип мінімізації структурного ризику

Якість наближеної моделі отриманої на основі вхідних та вихідних даних може визначатися як наступний L^2 функціонал справжнього ризику [108]:

$$R_{true}(f) = \iint_{\mathcal{X} \times \mathcal{Y}} (y - f(\mathbf{x}))^2 P(\mathbf{x}, y) d\mathbf{x} dy, \quad (1.3)$$

де $P(\mathbf{x}, y)$ — сумісна щільність розподілу, яка апіорі невідома. Основна задача — знайти таку функцію f_{true} серед деякого скінченного класу функцій \mathcal{M} , для якої ризик $R_{true}(f)$ мінімальний:

$$f_{true} = \arg \min_{f \in \mathcal{M}} R_{true}(f).$$

Справжній ризик R_{true} має зміст узагальненої помилки.

Мінімізація емпіричного ризику

Згідно з принципом мінімізації емпіричного ризику, задачу мінімізації невідомого справжнього ризику R_{true} можна замінити на мінімізацію відомого емпіричного ризику R_{emp} , який може бути визначений наступним чином [18]:

$$R_{emp}(f) = \frac{1}{N} \sum_{j=1}^N (y_j - f(\mathbf{x}_j))^2. \quad (1.4)$$

При цьому $R_{emp}(f) < R_{true}(f)$. Емпіричний ризик є спроможним, якщо

$$\lim_{N \rightarrow \infty} \min_{f \in \mathcal{M}} R_{emp}(f) = \min_{f \in \mathcal{M}} R_{true}(f).$$

Мінімізація емпіричного ризику відпрацьовує бездоганно у випадку коли функція f_{true} , яка точно відображає реальний процес, належить простору \mathcal{M} . Але на жаль, складні процеси не можуть бути описані за допомогою тривіальних залежностей. Для дослідження таких питань як: за яких умов припустима мінімізація емпіричного ризику і наскільки гарні наближення можна отримати з її допомогою, розвинена спеціальна статистична теорія навчання [109, 110], започаткована В.Н.Вапником і А.Я.Червоненкісом [16]. Цікавий розвиток статистична теорія навчання знайшла також в роботах К.В.Воронцова [111, 112, 113].

Мінімізація структурного ризику

В [114] можна знайти означення розмірності Вапника-Червоненкіса для простору значень. Як правило це означення поширюється на простір функцій наступним чином [115]:

Означення 1.1. Розмірністю Вапника-Червоненкіса (ВЧ) деякого класу функцій \mathcal{M} називається максимальний об'єм вибірки даних h такий, що для будь-яких значень вибірки існує така функція f з цього класу, яка дає нульову помилку.

Очевидно, що для лінійних моделей розмірність ВЧ співпадає з числом ступенів свободи моделей (числом параметрів моделі). Наприклад, для найпростішої лінійної стаціонарної моделі вигляду $y(x) = ax + b$ ВЧ-розмірність $h = 2$, оскільки через будь-які дві точки можна провести пряму.

Показано, що з ймовірністю $(1 - \kappa)$ верхня межа справжнього ризику складає:

$$R_{true}(f) \leq R_{emp}(f) + \sqrt{\frac{h(\ln(2N/h) + 1) - \ln(\kappa/4)}{N}}.$$

З цього можна зробити очевидний висновок, що для стовідсоткової впевненості в структурі моделі необхідна нескінченна вибірка даних. Верхню межу справжнього ризику можна інтерпретувати як суму емпіричного ризику і інтервалу довіри до моделі, який виражає ступінь її складності (рис. 1.3). Принцип мінімізації структурного ризику є формальним методом для вибору

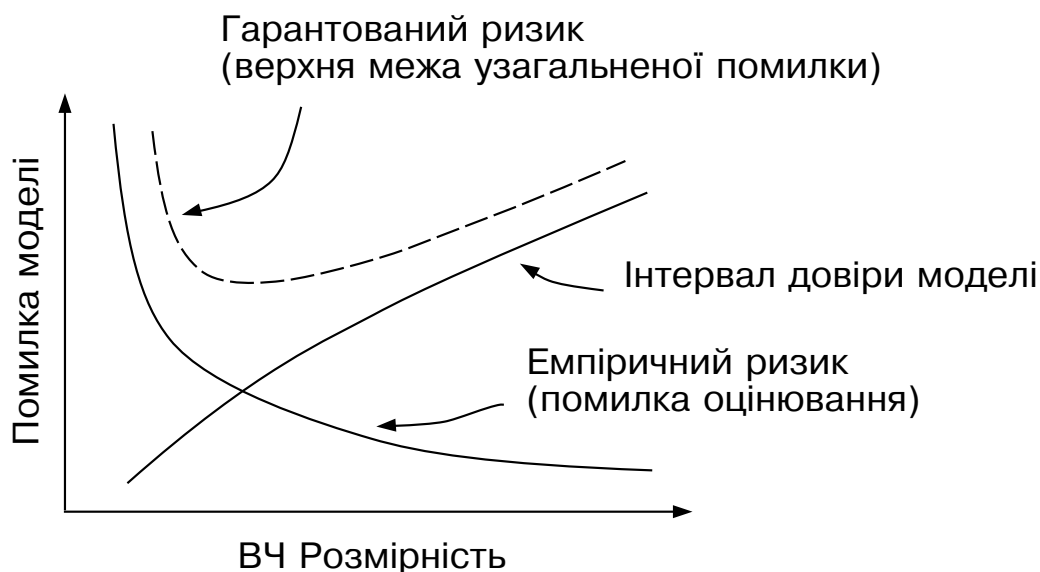


Рис. 1.3. Верхня межа справжнього ризику.

оптимальної структури моделі для скінченної вибірки даних. Він полягає в наступному.

Задається структура вкладених класів потенційних моделей:

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \subset \dots \subset \mathcal{M}_k \subset \dots,$$

де модель $\mathcal{M}_k = \{f_k(\mathbf{x}, \mathbf{w}) : \mathbf{w} \in \mathcal{F}_k, \dim \mathcal{F}_k = q_k\}$ має скінчену ВЧ розмірність h_k (наприклад модель \mathcal{M}_k може визначати множину поліномів ступеню не менше ніж k). Таким чином множина потенційних моделей ранжується в залежності від складності (ємності) тієї чи іншої моделі, оскільки:

$$h_1 \leq h_2 \leq h_3 \leq \dots \leq h_k \leq \dots$$

Для заданої вибірки даних вибирається така модель \mathcal{M}_k , для якої досягається мінімальна верхня межа справжнього ризику.

Оскільки в роботі розглядаються складні нелінійні процеси, а статистична теорія навчання обмежена простими (лінійними) моделями, де ВЧ розмірність може бути обчислена аналітично, то для застосування цієї теорії необхідно або робити наближення для ВЧ розмірності, або лінеаризувати нелінійні моделі.

1.2.2. Регуляризація Тихонова

Для визначення псевдооберненої матриці \mathbf{A}^+ до матриці \mathbf{A} над дійсними числами можна використовувати границю обернених:

$$\mathbf{A}^+ = \lim_{\alpha \rightarrow 0} (\mathbf{A}^\top \mathbf{A} + \alpha \mathbf{I})^{-1} \mathbf{A}^\top = \lim_{\alpha \rightarrow 0} \mathbf{A}^\top (\mathbf{A}^\top \mathbf{A} + \alpha \mathbf{I})^{-1}.$$

Аналогічна ідея використовується в тихоновській регуляризації некоректних задач [116, 117], зокрема для розв'язання погано обумовлених систем лінійних рівнянь $\mathbf{A}\mathbf{w} = \mathbf{y}$ відносно вектора невідомих змінних \mathbf{w} . Метод регуляризації Тихонова зводить початкову задачу до задачі мінімізації наступного функціоналу:

$$\|\mathbf{A}\mathbf{w} - \mathbf{y}\|^2 + \alpha \|\mathbf{w} - \mathbf{w}_0\|^2, \quad (1.5)$$

де $\|\cdot\|$ — евклідова норма. Тоді регуляризований розв'язок можна знайти у формі:

$$\mathbf{w}_{reg} = (\mathbf{A}^\top \mathbf{A} + \alpha \mathbf{I})^{-1} (\mathbf{A}^\top \mathbf{y} + \alpha \mathbf{w}_0).$$

Таким чином в основі регуляризації лежить залучення додаткової апріорної інформації про розв'язок — шукають розв'язок близький до \mathbf{w}_0 . Постановка може мати більш загальний характер, якщо задати (апріорну) матрицю коваріації $\mathbf{\Omega}_a$ для вектору невідомих змінних. В нашому випадку $\mathbf{\Omega}_a = \mathbf{I}$, що відповідає регуляризації нульового порядку [118]. В контексті некоректних задач навчання функціонал (1.5) має зміст регуляризованого ризику:

$$R_{reg}(f) = NR_{emp}(f) + \alpha R_f(f), \quad (1.6)$$

де $\alpha > 0$ — коефіцієнт регуляризації (Тихонова), а складова $R_f(f)$ виражає апріорні сподівання щодо очікуваної функції. Коефіцієнт регуляризації контролює баланс зміщення-варіації. Так, при збільшенні значення коефіцієнту регуляризації зменшується варіація моделі, але збільшується зміщення. Для розв'язання такого класу задач в роботі Н.В.Панової [119] запропонований ефективний регуляризований рекурентний МНК, де складова $R_f(f)$ виступає в ролі стабілізуючого функціоналу. Інший регуляторний підхід для побудови стійких алгоритмів відновлення залежностей на основі робастних методів згладжуючої сплайн-регресії розроблений В.М.Остропицьким [120, 121]. Задача мінімізації ризику (1.6) також відома як задача відновлення гребеневої або реберної (ridge) регресії [122], де коефіцієнт регуляризації α називають параметром гребеня.

Нехай невідома залежність f описується лінійною в параметрах моделлю в деякому просторі \mathcal{F} :

$$y = f(\mathbf{x}) = \mathbf{a}^\top \mathbf{w}, \quad \mathbf{a} = \Phi(\mathbf{x}), \quad \Phi : \mathcal{X} \rightarrow \mathcal{F}.$$

Для заданої вибірки даних $\mathcal{D} = \{(y_j, \mathbf{x}_j) : j = 1, \dots, N\}$ позначимо $\mathbf{y} = (y_1, \dots, y_N)^\top$, де $y_j = f(\mathbf{x}_j)$. Тоді задача навчання зводиться до розв'язання системи лінійних рівнянь $\mathbf{y} = \mathbf{A}\mathbf{w}$, де матриця $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_N)^\top$. Як правило припускають, що $\mathbf{w}_0 = \mathbf{0}$, тоді регуляризований розв'язок має вигляд:

$$\mathbf{w}_{reg} = \mathbf{H}^{-1} \mathbf{A}^\top \mathbf{y}, \quad \mathbf{H} = \mathbf{A}^\top \mathbf{A} + \alpha \mathbf{I}.$$

де матриця \mathbf{H} є гессіаном ризику, оскільки її можна отримати після подвійного диференціювання ризику $\mathbf{H} \propto \nabla_{\mathbf{w}}^2 R_{reg}(\mathbf{w})$. Оскільки кожному вектору параметрів відповідає певна функція, то позначення $R_{reg}(\mathbf{w})$ і $R_{reg}(f)$ є рівноцінними. Складові регуляризованого вектору параметрів відображено на рис. 1.4. Відповідно оцінку відгуку можна представити у вигляді:

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y},$$

де $\mathbf{S} = \mathbf{A}\mathbf{H}^{-1}\mathbf{A}^\top$ — лінійний оператор згладжування. У випадку відсутності механізму регуляризації оператор згладжування вироджується в проекційну

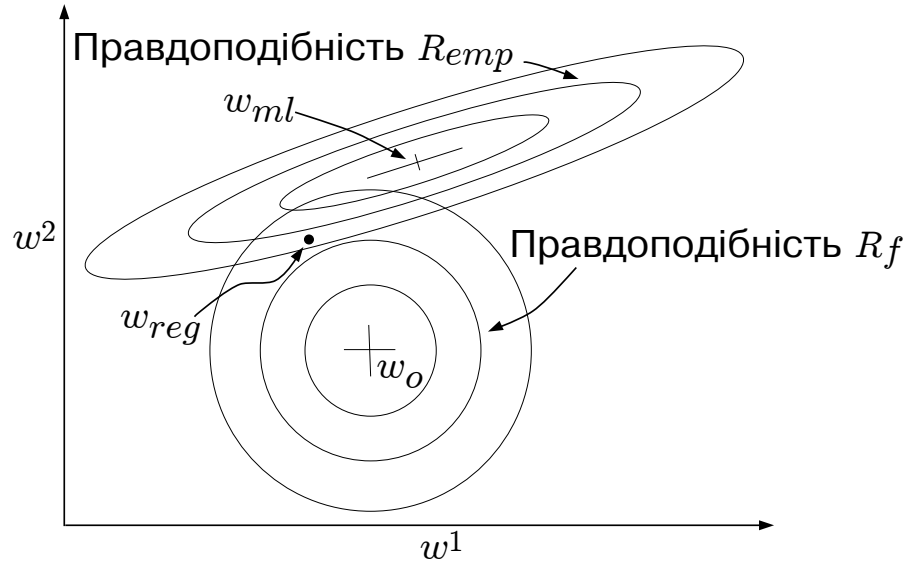


Рис. 1.4. Контури правдоподібності і власні вектори в масштабі власних чисел для $\mathbf{A}^\top \mathbf{A}$ з центром \mathbf{w}_{ml} , для $\mathbf{\Omega}_a^{-1}$ з центром \mathbf{w}_0 , для \mathbf{H} з центром \mathbf{w}_{reg} .

матрицю $\mathbf{S} \equiv \mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$. Проекційну матрицю можна інтерпретувати як оператор проекції, який проектує вектори на підпростір елементів стовпців матриці \mathbf{A} [123]. Як легко показати, кількість ступенів свободи \acute{q} нерегуляризованої моделі, яка співпадає з числом незалежних змінних (кількістю ефективних параметрів за Дж.Муді [124]) дорівнює

$$\acute{q} = \text{rang } \mathbf{P} = \text{tr } \mathbf{P}.$$

Це слідує з того, що власні числа проекційної матриці рівні або 0 або 1. Регуляризація змінює число ступенів свободи моделі:

$$\acute{q} = \text{tr } \mathbf{S} = \text{tr } (\mathbf{A} \mathbf{H}^{-1} \mathbf{A}^\top) = \text{tr } (\mathbf{A}^\top \mathbf{A} \mathbf{H}^{-1}). \quad (1.7)$$

Звідси маємо:

$$\acute{q} = \sum_{i=1}^q \frac{\gamma_i}{\gamma_i + \alpha}, \quad (1.8)$$

де γ_i — власні числа автокореляційної матриці $\mathbf{A}^\top \mathbf{A}$, дійсні та невід'ємні. Якщо γ_i невелике в порівнянні з α , то функція правдоподібності виражає невелику довіру цьому напрямку в просторі параметрів, вважаючи за краще

використовувати апіорно заданий розподіл параметрів. Якщо γ_i істотно більше ніж α , то вибірка даних значно більше впливає на відповідний напрямок в просторі параметрів.

1.2.3. Байєсівський підхід до навчання

Основною задачею байєсівського підходу є знаходження апостеріорного прогнозного розподілу на основі даних спостережень $\mathcal{D} = \{(y_j, \mathbf{x}_j) : j = 1, \dots, N\}$. Нехай систематична складова стохастичної залежності може бути представлена деякою функцією f моделі \mathcal{M} із простору моделей \mathcal{H} . Виникають питання: яким чином визначити простір моделей \mathcal{H} , як визначити оптимальну модель \mathcal{M} з цього простору, і яким чином визначити параметри цієї моделі, тобто конкретну функцію f . Байєсівський аналіз дає відповідь на останні два запитання.

Проблема вибору моделі

Нехай задано простір моделей \mathcal{H} . Тоді, відповідно до байєсівського аналізу, в якості оптимальної вибирається модель з найбільшою апостеріорною ймовірністю $P(\mathcal{M}|\mathcal{D})$. За формулою Байєса:

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})},$$

де $P(\mathcal{M})$ — апіорний розподіл моделей в просторі \mathcal{H} , $P(\mathcal{D})$ — апіорна ймовірність виникнення самої вибірки даних, яка після отримання даних спостережень відіграє роль лише нормуючої константи. Припускаючи рівномірний розподіл моделей, моделі ранжуються за величиною їх підтвердження $P(\mathcal{D}|\mathcal{M})$ (англ. evidence), яке виражає здатність моделі \mathcal{M} генерувати дані \mathcal{D} . На рис. 1.5 показані величини підтверджень даних $\mathcal{D}_1, \mathcal{D}_2$ моделями $\mathcal{M}_1, \mathcal{M}_2$. Підтвердження, також відоме як маргінальна правдоподібність, визначається як:

$$P(\mathcal{D}|\mathcal{M}) = \int_{\mathcal{M}} P(\mathcal{D}, f|\mathcal{M})df, \quad (1.9)$$

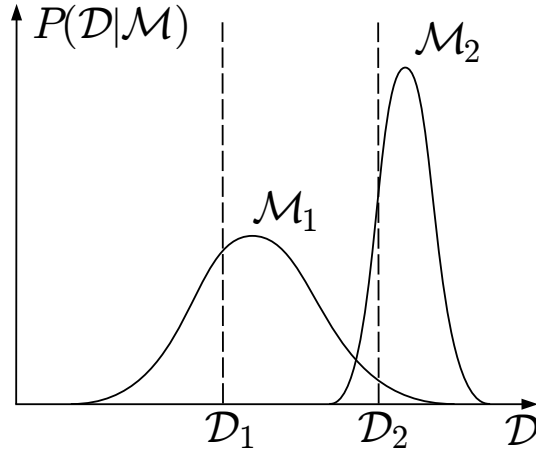


Рис. 1.5. Порівняння величини байєсівського підтвердження для різних моделей.

де сумісна щільність розподілу $P(\mathcal{D}, f|\mathcal{M}) = P(\mathcal{D}|f, \mathcal{M})P(f|\mathcal{M})$. Ймовірність $P(\mathcal{D}|f, \mathcal{M})$ виражає здатність функції f генерувати дані \mathcal{D} і називається правдоподібністю. Надалі вважається, що випадкова складова стохастичної залежності є адитивним шумом і $y_j = f(\mathbf{x}_j) + \delta_j$, де δ_j — незалежні і однаково розподілені за законом $P(\delta|f, \mathcal{M})$ випадкові величини. Правдоподібність у цьому випадку має вигляд:

$$P(\mathcal{D}|f, \mathcal{M}) = \prod_{j=1}^N P(\delta_j|f, \mathcal{M}),$$

Нехай апіорна щільність розподілу функцій f моделі \mathcal{M} , згідно з принципом максимальної ентропії [8]², задана у вигляді багатовимірного нормального розподілу з математичним сподіванням \mathbf{b} і матрицею коваріації \mathbf{K} :

$$P(f|\mathcal{M}) = \mathcal{N}(\mathbf{f}|\mathbf{b}, \mathbf{K}), \quad \mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^\top. \quad (1.10)$$

Оскільки підтвердження (1.9) може бути обчислене аналітично тільки в найпростіших випадках, наприклад у випадку нормального шуму, то застосову-

²Згідно з принципом максимальної ентропії знаходять розподіл, який з одного боку максимізує інформаційну ентропію процесу, а з іншого задовольняє наявній апіорній інформації про процес. В класі розподілів із заданими першими двома моментами таким оптимальним апіорним розподілом є нормальний розподіл.

ють різноманітні методи наближення. Серед них метод Монте-Карло [125], розповсюдження сподівань [126], варіаційні методи, метод Лапласа [15].

Проблема визначення параметрів моделі

Для визначення параметрів моделі байєсівський аналіз розглядає апостеріорний розподіл функцій:

$$P(f|\mathcal{D}, \mathcal{M}) = \frac{P(\mathcal{D}, f|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})}.$$

Словами:

$$\text{Апостеріорний розподіл} = \frac{\text{Правдоподібність} \times \text{Апріорний розподіл}}{\text{Підтвердження}}.$$

В цьому випадку апостеріорний прогнозний розподіл, який виражає апостеріорні сподівання щодо значень y , обчислюється як інтеграл по апостеріорній невизначеності функцій:

$$P(y|\mathbf{x}, \mathcal{D}) = \int_{\mathcal{M}} P(y|\mathbf{x}, f, \mathcal{M})P(f|\mathcal{D}, \mathcal{M})df.$$

Використовуючи так звану оцінку максимальної апостеріорної ймовірності функції (МАП), яка дає наближення математичного сподівання апостеріорного прогнозного розподілу модою апостеріорного розподілу функцій, знаходяться оптимальні значення параметрів моделі:

$$y(\mathbf{x}) \approx f_{mp}(\mathbf{x}) = \arg \max_f P(f|\mathcal{D}, \mathcal{M}).$$

Графічна інтерпретація байєсівського висновку за С.А.Шумським [6] в контексті навчання зображена на рис. 1.6. Зазначимо, що, слідуючи класичній байєсівській теорії прийняття рішень, ризик прийняти оцінку \hat{f} для справжньої функції f_{true} дорівнює:

$$R_L(\hat{f}) = \int_{\mathcal{M}} L(\hat{f}, f)P(f|\mathcal{D}, \mathcal{M})df,$$

де функція втрат $L(\hat{f}, f)$ характеризує втрати пов'язані з різницею між оцінкою і невідомою справжньою функцією [127]. Це слідує з мінімізації

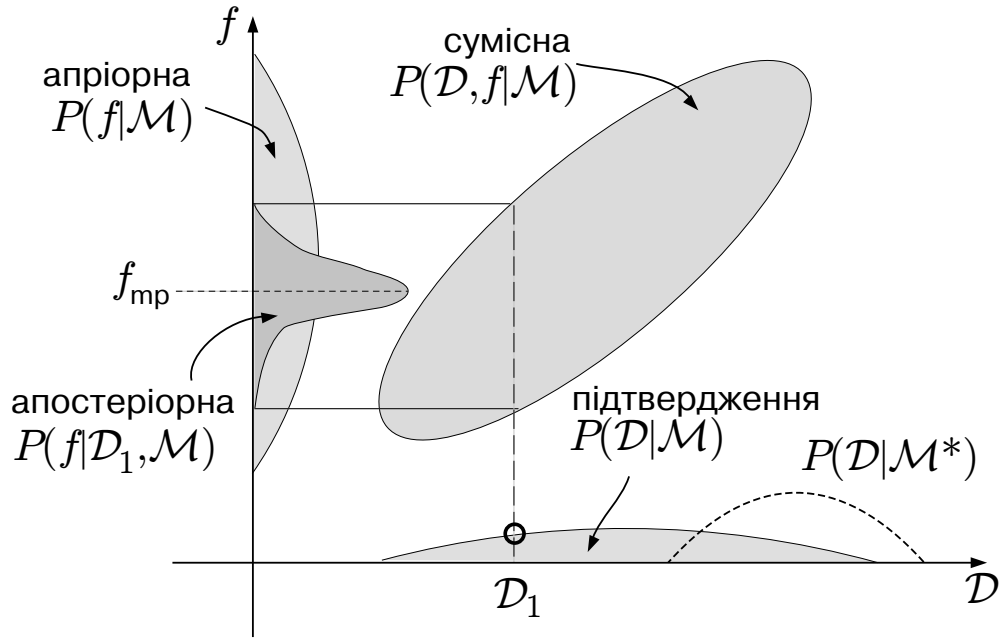


Рис. 1.6. Байєсівське навчання.

справжнього байєсівського ризику, який узагальнює (1.3), як математичного сподівання функції втрат [128]:

$$\begin{aligned}
 R_{true}(\hat{f}) &= \mathbb{E}L(\hat{f}, f) = \iint L(\hat{f}, f)P(f, \mathcal{D}|\mathcal{M})df d\mathcal{D} = \\
 &= \int \left(\int_{\mathcal{M}} L(\hat{f}, f)P(f|\mathcal{D}, \mathcal{M})df \right) P(\mathcal{D}|\mathcal{M})d\mathcal{D}.
 \end{aligned}$$

Тож, для заданої вибірки \mathcal{D} оптимальна байєсівська оцінка мінімізує як справжній ризик R_{true} так і так званий умовний байєсівський ризик R_L . Якщо визначити функцію втрат як абсолютну помилку $L(\hat{f}, f) = |\hat{f} - f|$, то оптимальною байєсівською оцінкою буде значення медіани апостеріорного розподілу функцій. Оскільки медіана це точка, для якої сума абсолютних значень різниці всіх значень менше суми різниць для будь-якої іншої точки. Якщо ж в якості функції втрат взяти середньо-квадратичну помилку $L(\hat{f}, f) = |\hat{f} - f|^2$, то оптимальною байєсівською оцінкою буде значення математичного сподівання апостеріорного розподілу функцій. Як показано в монографії Е.Джейнса [129], МАП оцінка f_{mp} , тобто мода апостеріорного розподілу функцій, буде оптимальною байєсівською оцінкою лише у випадку,

коли функція втрат визначена як:

$$L(\hat{f}, f) = \lim_{k \rightarrow 0} |\hat{f} - f|^k = \begin{cases} 0, & \hat{f} = f \\ 1, & \hat{f} \neq f \end{cases}.$$

Хоч серед прибічників чистого байєсівського висновку такий підхід часто критикують, надалі в роботі буде використовуватися саме МАП оцінка. Змістовна теорія байєсівського оцінювання в теорії статистичних висновків досить детально викладена в роботах Ш.Закса [130] та М.Де Гроота [131].

Відповідно до МАП оцінки найбільш ймовірна функція f_{mp} визначається як:

$$\begin{aligned} f_{true} \sim f_{mp} &= \arg \max_f (\log P(f|\mathcal{D}, \mathcal{M})) = \\ &= \arg \min_f (-\log P(\mathcal{D}|f, \mathcal{M}) - \log P(f|\mathcal{M})). \end{aligned} \quad (1.11)$$

Оскільки підтвердження моделі $P(\mathcal{D}|\mathcal{M})$ не залежить від окремої функції і в даному випадку виконує роль нормуючої константи.

Якщо припустити рівномірність апіорного розподілу функцій $P(f|\mathcal{M})$, то максимізація апостеріорної ймовірності $P(f|\mathcal{D}, \mathcal{M})$ еквівалентна максимізації функції правдоподібності:

$$f_{ml} = \arg \min_f (-\log P(\mathcal{D}|f, \mathcal{M})).$$

За умови, що розподіл шуму є нормальним з нульовим середнім і з дисперсією $D\delta = \sigma^2$, тобто $P(\delta|f, \mathcal{M}) = \mathcal{N}(\delta|0, \sigma^2)$, тоді критерій максимальної правдоподібності в протилежному логарифмічному масштабі записується у вигляді [132]:

$$-\ln P(\mathcal{D}|f, \mathcal{M}) = N \ln(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2} \sum_{j=1}^N \delta_j^2 \xrightarrow{f} \min. \quad (1.12)$$

Мінімізація (1.12) еквівалентна мінімізації емпіричного ризику (1.4). Відзначимо, що метод найменших квадратів знаходить максимум правдоподібності

з нормально розподіленим шумом. Критерій (1.12) можна узагальнити на випадок довільної коваріації помилок $\mathbf{R} \sim r_{ij} = \text{cov}(\delta_i, \delta_j)$ [133]:

$$-\ln P(\mathcal{D}|f, \mathcal{M}) = N \ln(\sqrt{2\pi}\sigma) + \frac{1}{2} \ln \det \mathbf{R} + \frac{1}{2} \delta^\top \mathbf{R}^{-1} \delta,$$

де вектор $\delta = (\delta_1, \dots, \delta_N)^\top$. Оскільки визначник коваріації не залежить від функцій f , то фактично задача знаходження максимальної апостеріорної ймовірності зводиться до мінімізації наступної квадратичної форми:

$$\delta^\top \mathbf{R}^{-1} \delta \xrightarrow{f} \min.$$

Проте, оскільки апріорний розподіл функцій $P(f|\mathcal{M})$ відіграє роль регуляризуючого елементу і забезпечує контроль над відношенням зміщення-варіація моделі, припущення щодо його рівномірності не є коректним. Тож, надалі ми будемо вважати апріорний розподіл нормально розподіленим у формі (1.10).

Наближення підтвердження методом Лапласа. Підтвердження гіперпараметрів.

Опишемо метод Лапласа для наближення підтвердження. Нехай апріорний розподіл функцій $f(\mathbf{x})$ моделі $\mathcal{M}(\mathbf{x}, \mathbf{w})$ визначається нормальним розподілом параметру \mathbf{w} наступним чином: $P(f|\mathcal{M}) = \mathcal{N}(\mathbf{w}|\mathbf{w}_0, \mathbf{\Omega}_a)$. Щільність розподілу шуму також нормальна з дисперсією σ^2 : $P(\delta|f, \mathcal{M}) = \mathcal{N}(\delta|0, \beta^{-1})$, параметр $\beta = \sigma^{-2}$. Тоді, згідно з МАП оцінкою (1.11), отримаємо:

$$\begin{aligned} f_{mp} = \arg \min_f & -\frac{N}{2} \ln \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{j=1}^N \delta_j^2 - \\ & - \frac{q}{2} \ln \frac{1}{2\pi} + \frac{1}{2} \ln \det \mathbf{\Omega}_a + \frac{1}{2} (\mathbf{w} - \mathbf{w}_0)^\top \mathbf{\Omega}_a^{-1} (\mathbf{w} - \mathbf{w}_0). \end{aligned}$$

Зауважимо, що ця задача еквівалентна мінімізації регуляризованого ризику:

$$R_{reg}(\mathbf{w}) = \frac{\beta}{2} N R_{emp}(\mathbf{w}) + \frac{1}{2} R_{\mathbf{w}}(\mathbf{w}), \quad (1.13)$$

де

$$R_{emp}(\mathbf{w}) = \frac{1}{N} \sum_{j=1}^N \delta_j^2, \quad R_{\mathbf{w}}(\mathbf{w}) = (\mathbf{w} - \mathbf{w}_0)^\top \mathbf{\Omega}_a^{-1} (\mathbf{w} - \mathbf{w}_0).$$

У відповідності з (1.6) коефіцієнт регуляризації $\alpha = \beta^{-1}$. Тобто β є оберненим коефіцієнтом регуляризації. Параметр β також називають гіперпараметром. При цьому регуляризований розв'язок \mathbf{w}_{reg} відповідає МАП оцінці $\mathbf{w}_{mp} = \mathbf{w}_{reg}$.

Для знаходження підтвердження (1.9) необхідно обчислити інтеграл:

$$P(\mathcal{D}|\mathcal{M}) = \sqrt{\left(\frac{\beta}{2\pi}\right)^N \left(\frac{1}{2\pi}\right)^q \frac{1}{\det \mathbf{\Omega}_a}} \cdot \int_{\mathcal{F}} \exp(-R_{reg}(\mathbf{w})) d\mathbf{w}. \quad (1.14)$$

Слідуючи методу наближень Лапласа [15], розкладемо регуляризований ризик в ряд Тейлора другого порядку в точці \mathbf{w}_{mp} :

$$R_{reg}(\mathbf{w}) \approx R_{reg}(\mathbf{w}_{mp}) + \frac{\partial R_{reg}(\mathbf{w}_{mp})}{\partial \mathbf{w}} (\mathbf{w}_{mp} - \mathbf{w}) + \frac{1}{2} (\mathbf{w}_{mp} - \mathbf{w})^\top \frac{\partial^2 R_{reg}(\mathbf{w}_{mp})}{\partial \mathbf{w}^2} (\mathbf{w}_{mp} - \mathbf{w}). \quad (1.15)$$

Друга складова в правій частині дорівнює нулю, оскільки точка \mathbf{w}_{mp} є точкою мінімуму регуляризованого ризику за побудовою. Таким чином (1.15) запишеться:

$$R_{reg}(\mathbf{w}) \approx R_{reg}(\mathbf{w}_{mp}) + \frac{1}{2} (\mathbf{w}_{mp} - \mathbf{w})^\top \mathbf{H} (\mathbf{w}_{mp} - \mathbf{w}).$$

де гессіан $\mathbf{H} = \beta \mathbf{A}^\top \mathbf{A} + \mathbf{\Omega}_a^{-1}$. Враховуючи розклад Тейлора підтвердження (1.14) можна записати у наступній формі:

$$P(\mathcal{D}|\mathcal{M}) \approx \sqrt{\left(\frac{\beta}{2\pi}\right)^N \left(\frac{1}{2\pi}\right)^q \frac{1}{\det \mathbf{\Omega}_a}} \cdot e^{-R_{reg}(\mathbf{w}_{mp})} \int_{\mathcal{F}} e^{-\frac{1}{2} (\mathbf{w}_{mp} - \mathbf{w})^\top \mathbf{H} (\mathbf{w}_{mp} - \mathbf{w})} d\mathbf{w}.$$

Обчисливши інтеграл, який відповідає багатовимірному розподілу Гауса, отримуємо:

$$P(\mathcal{D}|\mathcal{M}) \approx \sqrt{\left(\frac{\beta}{2\pi}\right)^N \left(\frac{1}{2\pi}\right)^q \frac{1}{\det \mathbf{\Omega}_a}} \cdot e^{-R_{reg}(\mathbf{w}_{mp})} \cdot \sqrt{\frac{(2\pi)^q}{\det \mathbf{H}}}.$$

Слідуючи Маккею [134], цей результат можна інтерпретувати як:

$$\underbrace{P(\mathcal{D}|\mathcal{M})}_{\text{Підтвердження}} \approx \underbrace{P(\mathcal{D}|\mathbf{w}_{mp}, \mathcal{M})}_{\text{Правдоподібність}} \cdot \underbrace{P(\mathbf{w}_{mp}|\mathcal{M})(2\pi)^{q/2} \det \mathbf{H}^{-1/2}}_{\text{Фактор Окхема}}.$$

Тобто підтвердження дорівнює добутку максимальної правдоподібності на так званий множник Окхема³, який має зміст відношення апостеріорної невізначеності вектору параметрів до апріорної [10]. Це означає, що байєсівський критерій вибору моделі є простим розширенням критерію максимальної правдоподібності. Підтвердження у зворотному логарифмічному масштабі запишеться так:

$$-\ln P(\mathcal{D}|\mathcal{M}) \approx -\frac{N}{2} \ln \frac{\beta}{2\pi} + \frac{1}{2} \ln \det \mathbf{\Omega}_a + \frac{1}{2} \ln \det \mathbf{H} + R_{reg}(\mathbf{w}_{mp}). \quad (1.16)$$

Для знаходження найбільш ймовірного значення гіперпараметру β необхідно знайти максимум підтвердження $P(\mathcal{D}|\mathcal{M})$. Це можна зробити наприклад за допомогою ітеративного методу Маккея [15]. Припустимо що, інтуїтивно задавши довільне значення гіперпараметру β , за допомогою регуляризованого критерію (1.13) було знайдене регуляризоване значення \mathbf{w}_{mp} . Необхідно дати рекомендації щодо того, яким чином необхідно змінити значення гіперпараметру для збільшення підтвердження. Змістова інтерпретація такої процедури оптимізації зображена на рис. 1.7. Зауважимо, що

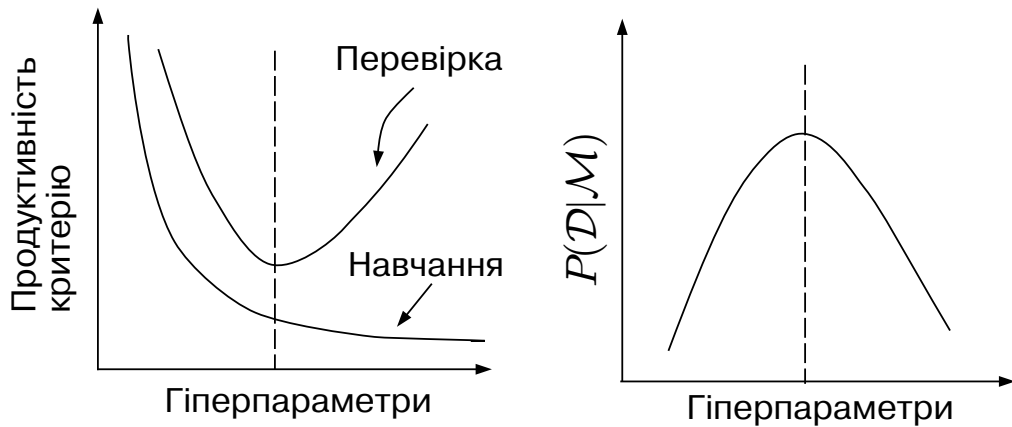


Рис. 1.7. Якість моделі в залежності від гіперпараметрів.

$\ln \det \mathbf{H} = \text{tr} \ln \mathbf{H}$. Тоді, взявши похідну в (1.16) відносно гіперпараметру β і

³Термін походить від відомого в літературі леза Окхема — Occam's Razor principle (William of Ockham): "Numquam ponendo est pluritas sine necessitate. [Latin]". Що буквально означає — не потрібно застосовувати складне, якщо не має необхідності. Більш звичний переклад — маючи дві однаково ефективні теорії, вибираємо простішу; найпростіше пояснення як правило є найкращим.

прирівнявши її нулю, отримаємо:

$$-\frac{N}{2} \cdot \frac{1}{\beta} + \frac{1}{2} \operatorname{tr}(\mathbf{H}^{-1} \mathbf{A}^\top \mathbf{A}) + \frac{N}{2} R_{emp}(\mathbf{w}_{mp}) = 0.$$

Звідси можна визначити переоціночну ітеративну формулу для гіперпараметру у формі:

$$\beta(i+1) = \frac{N - \beta(i) \operatorname{tr}(\mathbf{H}^{-1} \mathbf{A}^\top \mathbf{A})}{N R_{emp}(\mathbf{w}_{mp})},$$

При цьому $\mathbf{w}_{mp} = \mathbf{H}^{-1} \mathbf{A}^\top \mathbf{y}$ і гессіан $\mathbf{H} = \beta(i) \mathbf{A}^\top \mathbf{A} + \boldsymbol{\Omega}_a^{-1}$.

1.2.4. Перевірка моделей на адекватність

Адекватність моделі регресії. Критерій Фішера

В класичній задачі перевірки моделі регресії на значущість розраховується наступна F -статистика [135, 136]:

$$F = \frac{s_x^2}{s_e^2}, \quad s_x^2 = \frac{1}{q} \sum_{i=1}^N (f_i - \bar{y})^2, \quad s_e^2 = \frac{1}{N-q} \sum_{i=1}^N (y_i - f_i)^2$$

де N — об'єм вибірки, q — кількість параметрів регресійної моделі. s_x^2 характеризує факторну (пояснювальну) дисперсію, s_e^2 характеризує дисперсію залишків. Тож, якщо F -статистика більше табличного значення $F_{q, N-q, \alpha}$, то з ймовірністю $1 - \alpha$ можна стверджувати про значимість регресії і наявність тренду, а також про адекватність моделі регресії за критерієм Фішера.

Застосування критерію Фішера для порівняння моделей

Розглянемо дві моделі:

$$\begin{aligned} \mathcal{M}_0 &= \{f_0(\mathbf{x}, \mathbf{w}_0) : \mathbf{w}_0 \in \mathcal{W}_0\}, \quad \dim \mathcal{W}_0 = q_0, \\ \mathcal{M}_1 &= \{f_1(\mathbf{x}, \mathbf{w}_1) : \mathbf{w}_1 \in \mathcal{W}_1 \supset \mathcal{W}_0\}, \quad \dim \mathcal{W}_1 = q_1. \end{aligned}$$

Позначимо простір $\mathcal{W}^\perp : \mathcal{W}_1 = \mathcal{W}_0 \oplus \mathcal{W}^\perp$, $\dim \mathcal{W}^\perp = s = q_1 - q_0$. Висунемо нуль гіпотезу про те, що для заданої вибірки даних \mathcal{D} вектор параметрів \mathbf{w}^\perp із простору \mathcal{W}^\perp дорівнює нулю:

$$H_0 : \mathbf{w}^\perp = \mathbf{0}, \quad \mathbf{w}^\perp \in \mathcal{W}^\perp.$$

Це в свою чергу означає що $R_{emp}(\mathbf{w}_0) = R_{emp}(\mathbf{w}_1)$. І, відповідно, альтернативна гіпотеза $H_1 : \mathbf{w}^\perp \neq \mathbf{0}$. Мета статистичного тесту полягає у визначенні факту: чи підтримує вибірка даних більшу модель. При цьому F -статистика рівна:

$$F = \frac{N - q_1}{N - q_0} \cdot \frac{R_{emp}(\mathbf{w}_0)}{R_{emp}(\mathbf{w}_1)},$$

оскільки $R_{emp}(\mathbf{w}_i)$ відповідає розподілу $\chi^2_{N-q_i}$. Задавши рівень значущості α , тобто ймовірність відкинути гіпотезу H_0 при $\mathbf{w}^\perp = \mathbf{0}$, можна порівняти F -статистику із табличним значенням $F_{N-q_0, N-q_1, \alpha}$ і прийняти рішення про справедливість нуль гіпотези, якщо розраховане значення менше табличного.

Міра кількості інформації

Відповідно до класифікації наведеної в роботі [137] існує декілька підходів до оцінювання кількості інформації: статистичний, семантичний, прагматичний і структурний. Найбільш поширений і розвинений статистичний підхід, засновником якого вважається К.Шенон. К.Шенон ввів поняття кількості інформації як міри невизначеності стану системи, яка знімається при отриманні інформації [138]. Кількісно виражена невизначеність отримала назву ентропії (інформаційної ентропії) за аналогією з подібним поняттям в статистичній механіці. При отриманні інформації зменшується невизначеність, тобто ентропія системи. Однак для оцінювання якості статистичних моделей прийнято дотримуватися поняття ентропії за Больцманом (статистичної ентропії), яка відображає міру ймовірності знаходження системи в однорідному стані. Узагальнюючи поняття статистичної ентропії як логарифму ймовірності отримання вибіркового розподілу з очікуваного, вводиться міра близькості ймовірнісних розподілів із заданими щільностями $P_f(x)$ і $P_g(x)$ відома як дивергенція (відстань) Кульбака-Лейблера, або інформація по Кульбаку [139]:

$$D_{KL}(P_g \| P_f) = \int P_g(x) \log \left(\frac{P_g(x)}{P_f(x)} \right) dx.$$

При цьому $D_{KL}(P_g||P_f) \geq 0$ і $D_{KL}(P_g||P_f) = 0 \Leftrightarrow P_g = P_f$. Величину $-D_{KL}(P_g||P_f)$ називають ентропією $P_g(x)$ по відношенню до $P_f(x)$. Чим більше ентропія, тим краще $P_f(x)$ наближає $P_g(x)$. Саму дивергенцію Кульбака-Лейблера називають негентропією (негативна ентропія) $P_g(x)$ по відношенню до $P_f(x)$. Чим менше негентропія тим краще $P_f(y)$ наближає $P_g(y)$.

Наведемо деякі відомі критерії вибору моделей на основі міри кількості інформації:

§ Інформаційний критерій Акаїке (ІКА):

$$\text{ІКА} = N \ln(R_{emp}(f)) + 2q.$$

В роботі [9] показано, що значення ІКА співпадає із подвійною середньою негентропією, і визначає міру неузгодженості моделі з реальним процесом: $\text{ІКА} = 2D_{KL}(P_{true}||P_f)$.

§ Остаточна помилка прогнозування (ОПП) в логарифмічному масштабі:

$$\ln(\text{ОПП}) = \ln(R_{emp}(f)) + \ln\left(\frac{N+q}{N-q}\right).$$

ОПП визначається як дисперсія помилки прогнозу на один крок вперед за умови використання для прогнозу оцінок параметрів авторегресійної моделі по МНК. Критерій мінімуму ОПП враховує як помилку пов'язану із неточністю оцінювання параметрів так і помилку від заміни системи моделлю [9]. Критерій ОПП, а потім і ІКА були розвинуті Х.Акаїке [140]. Зазначимо, що мінімізація ІКА і мінімізація ОПП це асимптотично еквівалентні процедури.

§ Байєсівська статистична міра значущості (БСМЗ) [141]:

$$\text{БСМЗ} = N \ln(R_{emp}(f)) + q \ln N.$$

Ще одна назва цього критерію — байєсівський інформаційний критерій (БІК), або байєсівський інформаційний критерій Шварца, який вивів

асимптотичну апроксимацію для інтегрованої функції правдоподібності, як ще називають байєсівське підтвердження, для слабо інформативного апріорного розподілу функцій $P(f|\mathcal{M})$ на умовах регулярності у вигляді:

$$\log P(\mathcal{D}|\mathcal{M}) \approx \log P(\mathcal{D}|f_{ml}, \mathcal{M}) - \frac{q}{2} \log N,$$

де q — число вільних параметрів моделі \mathcal{M} , f_{ml} — оцінка максимальної правдоподібності для f , яка рівна:

$$f_{ml} = \arg \max_f P(\mathcal{D}|f, \mathcal{M}).$$

§ Мінімальна довжина описання (МДО) за Дж.Рісаненом [142]:

$$\text{МДО} = \ln(R_{emp}(f)) + q \frac{\ln N}{N}.$$

Цей критерій співпадає з байєсівською мірою значущості [143].

Для порівняння регуляризованих моделей (підрозділ 1.2.2.) розвинуті наступні критерії вибору моделі, де замість кількості вільних параметрів моделі використовується кількість ефективних параметрів.

§ Узагальнена помилка прогнозування (УПП) [124, 118]:

$$\text{УПП} = R_{emp}(f) + 2\sigma^2 \frac{\acute{q}}{N}. \quad (1.17)$$

де σ^2 — варіація шуму, \acute{q} — кількість ефективних параметрів.

§ Інформаційний критерій мережі (ІКМ) [144]:

$$\text{ІКМ} = N \ln(R_{emp}(f)) + 2\acute{q}. \quad (1.18)$$

ІКМ фактично є узагальненням критерію Акаїке, для випадку регуляризації (1.13). Цей критерій мінімізує незміщену оцінку помилки прогнозу.

Перехресна валідація (або перехресне обґрунтування)

Підхід перехресного обґрунтування (англ. cross-validation) полягає у розділенні наявної вибірки даних на дві вибірки: навчальну і перевіірочну. Процедура пошуку моделі перебирає задану кількість структур моделей і вибирає

модель з мінімальною помилкою на перевіірочній вибірці. Якщо перевіірочна вибірка слідує безпосередньо після навчальної, то говорять про випереджаючу валідацію (англ. forward validation). Також застосовують так звану k -кратну перехресну валідацію (k -fold cross-validation). В цьому випадку вибірку ділять на k приблизно рівних підмножин. Модель навчають k разів, кожен раз виключаючи з навчальної вибірки одну з підмножин. У випадку якщо k рівне об'єму вибірки, то має місце перехресна валідація через виключення одного (англ. leave-one-out cross-validation). Але такого роду підходи мають наступні недоліки:

- Як правило вибірка досить обмежена в розмірі і не покриває вхідний простір. Тож, якщо не змінювати перевіірочну вибірку, може бути вибрана гіперчутлива модель. Тобто ідеально мати на кожен тест нову перевіірочну вибірку.
- Великі обчислювальні витрати. Особливо у випадку k -кратної перехресної валідації.
- Одним із суттєвих недоліків випереджаючої валідації є той факт, що безпосередньо останні наявні дані беруть участь лише в пошуку структури моделі і не впливають на сам прогноз вибраної моделі. Якщо припустити, що структура моделі відома, то останні дані будуть просто ігноруватися.
- Перехресна валідація це “зашумлена” міра адекватності моделі. Тож складно обчислити оптимальні значення гіперпараметрів, оскільки необхідно знайти екстремум функції шуму. На відміну, наприклад, від байєсівського підтвердження, яке не є функцією шуму і дозволяє обчислити градієнт відносно гіперпараметрів.

Більш ефективною є узагальнена перехресна валідація (УПВ) [145]:

$$\log(\text{УПВ}) = \log(R_{emp}(f)) - 2 \log\left(1 - \frac{q}{N}\right).$$

Показано, що УПВ асимптотично еквівалентна перехресній валідації.

Зазначимо, що всі наведені критерії асимптотично еквівалентні⁴.

1.3. Структурованість моделі. Розклад аналізу варіацій

На практиці дослідник, як правило, рідко знаходить правильну структуру моделі з першого разу і він вимушений повторювати спроби знову і знову поки не отримає задовільну модель. Цей процес відображено на рис. 1.8. При

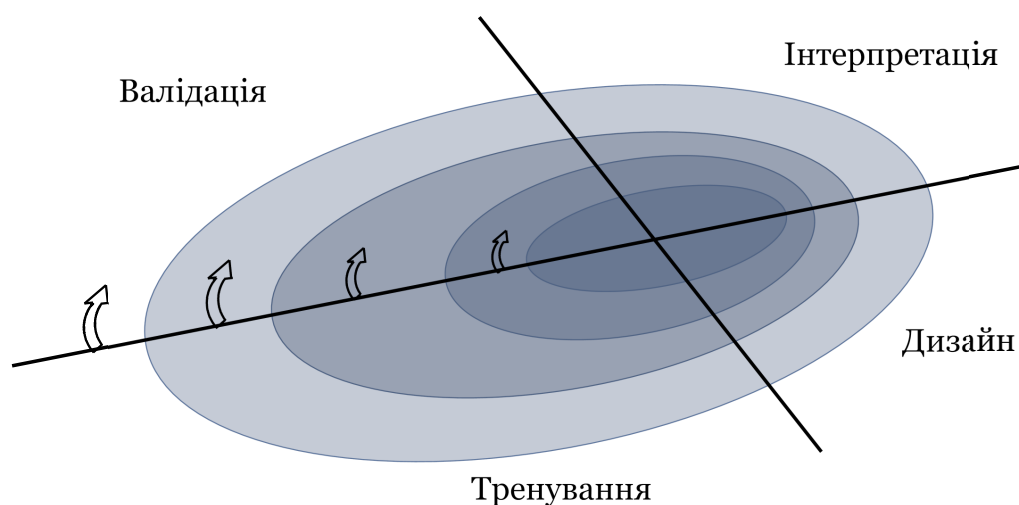


Рис. 1.8. Циклічна структура побудови моделі.

цьому основні складові циклу побудови моделі мають наступний зміст [45]:

- ✓ Дизайн — вибір представлення моделі.
- ✓ Тренування — оцінка параметрів моделі.
- ✓ Валідація — тестування моделі для визначення її адекватності реальному процесу.
- ✓ Інтерпретація — яким чином модель пояснює (розуміє) реальний процес?

Зазначимо, що структурованість моделі, тобто її властивість до легкої інтерпретації, відіграє неабияку роль в правильному уточненні структури моделі.

⁴В тому числі, за деяких додаткових припущень, асимптотично еквівалентні метод перехресної валідації і критерій Акаїке, про що згадується у Х.Акаїке [9]. Асимптотична еквівалентність означає нескінченність вибірки даних ($N \rightarrow \infty$).

Тому в роботі приділяється увага розробці саме структурованих, з точки зору інтерпретації, моделей. Формальне визначення класу структурованих моделей наведено в роботі [146]. Відповідно до нього структурована модель розглядається як сукупність операторів, суперпозиція яких здійснює відображення вхідних факторів (функцій) у вихідний фактор (функцію).

Пошук структури моделі у вигляді розкладу аналізу варіацій:

$$\eta(X) = \sum_j \eta_j(X_j) + \sum_{j>k} \eta_{jk}(X_j, X_k) + \sum_{j>k>l} \eta_{jkl}(X_j, X_k, X_l) + \dots,$$

де вивчається залежність фактору X від факторів X_j , є найбільш поширеним способом забезпечення структурованості. Так, наприклад, робота [147] присвячена пошуку і прийняттю рішення по поліноміальній статистичній n -факторній моделі другого порядку:

$$\eta(\mathbf{x}) = \beta_0 + \sum_{j=1}^n \beta_j x_j + \sum_{j=1}^n \beta_{jj} x_j^2 + \sum_{i<j} \beta_{ij} x_i x_j.$$

В цьому підрозділі розглядається декілька типових застосувань такого розкладу, якими є представлення моделі у вигляді ряду Вольтера і його дискретного аналогу — узагальненого поліному Колмогорова-Габора.

1.3.1. Описання динамічних процесів за допомогою ряду Вольтера

Математичний апарат рядів Вольтера добре вивчений [148, 149, 150]. Також можна знайти багато джерел по методах моделювання на основі функціональних рядів Вольтера (ФРВ) [151, 152], тобто методах, які дозволяють відшукати ядра Вольтера. В.М.Пінчук [153] розробив метод ідентифікації динамічних стаціонарних об'єктів керування, який полягає в ітеративній побудові ФРВ. Інший інтегральний метод ідентифікації динамічних об'єктів за експериментальними даними запропоновано в [154]. В роботі [75] запропоновано метод ідентифікації ядер Вольтера другого порядку (двох змінних) за допомогою вейвлетів з трикутним носієм.

Однією з сучасних розробок на основі рядів Вольтера є метод навчання нейронних мереж А.І.Іванова [155, 156], в якому кожному ряду Вольтера ставиться у відповідність нейронна мережа і, як наслідок, всі розробки для рядів Вольтера переносяться на теорію побудови та навчання нейронних мереж. Так, зокрема, симетризація ядер Вольтера дозволяє знизити складність задачі навчання нейронних мереж з експонентної до поліноміальної (лінійної або квадратичної). Основною ідеєю є використання рядів Вольтера для ідентифікації структури економічної моделі як системи з нескінченною пам'яттю. Саме такі системи пояснюють нелінійні динамічні процеси в економіці: явища гістерезису та наявність ділових циклів.

Ідентифікація ФРВ за допомогою розкладу ядер за системою ортогональних поліномів

В роботі [157] розвинена техніка аналізу ФРВ з допомогою розкладу ядер за системою базисних функцій. Для вхідного сигналу будується спеціальна система функцій, які є лінійними згортками входу і вибраного базису функцій, далі вихід системи розкладається в ряд по цим функціям. Гарні апроксимаційні властивості для широкого класу систем мають наближення вигляду:

$$\hat{y}(t) = h_0 + \int h_1(\tau)x(t - \tau)d\tau + \iint h_2(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2)d\tau_1d\tau_2. \quad (1.19)$$

Це відповідає представленню першими членами функціонального ряду Вольтера. Припустимо, що ядра h_1 , h_2 допускають розклад по деякій системі базисних функцій $\{\varphi_i(\tau, \alpha) : i = 1, \dots, k\}$:

$$h_1(\tau) = \sum_{i=1}^k w_i \varphi_i(\tau, \alpha), \quad h_2(\tau_1, \tau_2) = \sum_{i,j=1}^k w_{ij} \varphi_i(\tau_1, \alpha) \varphi_j(\tau_2, \alpha). \quad (1.20)$$

Тоді вираз (1.19), враховуючи (1.20), приймає вигляд:

$$\hat{y}(t) = w_0 + \sum_{i=1}^k w_i f_i(t) + \sum_{i,j=1}^k w_{ij} f_i(t) f_j(t), \quad (1.21)$$

де $f_i(t) = \int \varphi_i(\tau, \alpha)x(t - \tau)d\tau$ — реакція i -ого фільтра на вхідний сигнал $x(t)$.

В якості системи базисних функцій можна, наприклад, використовувати систему функцій Лагера, що дозволяє найкращим чином апроксимувати імпульсні перехідні характеристики аперіодичних і слабоколивальних об'єктів. Система функцій Лагера має вигляд:

$$\varphi_i(\tau, \alpha) = \sqrt{2\alpha}e^{-\alpha\tau} \sum_{j=0}^{i-1} \frac{(i-1)!(-2\alpha)^j}{(i-j-1)!(j!)^2} \tau^j.$$

Зважаючи на специфіку того чи іншого об'єкта, систему базисних функцій можна формувати з міркувань найкращого наближення ядер і використовувати для цього спеціальні класи поліномів, зокрема ортогональні поліноми Чебишева 2-го роду. Переваги цих поліномів описані в [158].

Рівняння регресії (1.21) також можна узагальнити на випадок системи з багатьма вхідними змінними, якщо представити всі виходи фільтрів для всіх вхідних змінних, як виходи одного узагальненого набору фільтрів:

$$\{f_i(x_r(t)) : i = 1, \dots, k; r = 1, \dots, n\} \mapsto \{f_p(t) : p = 1, \dots, kn\}.$$

Тоді узагальнений вихід запишеться у формі:

$$\hat{y}(t) = w_0 + \sum_{p=1}^{kn} w_p f_p(t) + \sum_{p,q=1}^{kn} w_{pq} f_p(t) f_q(t).$$

1.3.2. Метод групового врахування аргументів

МГВА запропонований в роботах О.Г.Івахненка [159, 94, 95] використовує ідеї самоорганізації та механізми еволюції такі як схрещування аргументів і генерація нащадків, селекція та відбір кращих. Основними перевагами методу є [96]:

- Метод не потребує задавання моделі у явному вигляді, модель будується автоматично в процесі роботи алгоритму.
- Метод працює на коротких вибірках, розмір яких незначно більший за число коефіцієнтів моделі.

При цьому класичний МГВА має досить суттєві недоліки:

- При близьких експериментальних точках ймовірна поява виродженості матриці нормальних рівнянь Гауса, так званий “індуцит”, внаслідок чого виникає необхідність застосування методів регуляризації.
- Метод видає точкову модель прогнозу, хоч іноді бажано мати довірчий інтервал, яких характеризує точність прогнозу.

Повна залежність між входом та виходом шукається в класі поліноміальних моделей де може бути представлена у вигляді узагальненого поліному Колмогорова-Габора:

$$\mathcal{M}(\mathbf{x}, \mathbf{a}) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{j=1}^n \sum_{i \leq j} a_{ij} x_i x_j + \dots \quad (1.22)$$

При побудові моделі і визначенні коефіцієнтів в якості критерію оптимальності використовується наступний критерій:

$$L(f) = \frac{1}{N} \sum_{t=1}^N (y_t - f(\mathbf{x}_t))^2.$$

Тобто мінімізується функціонал $L(f)$, який співпадає з емпіричним ризиком для фіксованої структури моделі (1.4). Для зменшення структурного ризику і подолання неповноти вибірки даних, яка є наслідком теореми неповноти Геделля, використовується принцип зовнішнього доповнення. В якості зовнішнього доповнення виступає додаткова перевірна вибірка, точки якої не використовувалися при навчанні системи, тобто при пошуку оціночних коефіцієнтів поліному (1.22). Це відповідає методу перехресної валідації з підрозділу 1.2.4. Таким чином пошук найкращої моделі здійснюється наступним чином:

1. Вся наявна вибірка даних \mathcal{D} ділиться на навчальну $\mathcal{D}_{N_{\text{навч}}}$ та перевірочну $\mathcal{D}_{N_{\text{перев}}}$: $N = N_{\text{навч}} + N_{\text{перев}}$.
2. На навчальній вибірці $\mathcal{D}_{N_{\text{навч}}}$ визначається вектор параметрів \mathbf{a} .

3. На перевірочній вибірці $\mathcal{D}_{N_{\text{перев}}}$ відбираються кращі моделі.

Складність моделі S визначається розмірністю вектору параметрів \mathbf{a} , тобто кількістю коефіцієнтів поліному. Зазначимо, що складність S є аналогом розмірності Вапника-Червоненкіса.

Схема класичного МГВА включає:

1. Для кожної пари x_i, x_j будуються часткові описи або субмоделі (всього C_n^2) вигляду:

— або лінійні

$$\hat{y}^{(s)} = \phi(x_i, x_j) = a_0 + a_i x_i + a_j x_j, \quad s = 1, \dots, C_n^2.$$

— або квадратичні

$$\hat{y}^{(s)} = \phi(x_i, x_j) = a_0 + a_i x_i + a_j x_j + a_{ii} x_i^2 + a_{ij} x_i x_j + a_{jj} x_j^2.$$

2. Визначаємо коефіцієнти цих моделей, тобто вектор параметрів \mathbf{a} , за допомогою МНК, використовуючи навчальну вибірку.

3. На перевірочній вибірці для кожної з побудованих моделей шукаємо оцінку:

$$L_s = \frac{1}{N_{\text{перев}}} \sum_{t=1}^{N_{\text{перев}}} \left(y_t - \hat{y}_t^{(s)} \right)^2, \quad (1.23)$$

де y_t — дійсне значення виходу системи в момент t , $\hat{y}_t^{(s)}$ — значення виходу моделі s в момент t на перевірочній вибірці. Визначаємо F_1 кращих моделей з найменшим значенням функціоналу L_s . Відібрані моделі подаються на другий ряд:

$$\phi^{(2)}(y_i, y_j) = a_0^{(2)} + a_i^{(2)} y_i + a_j^{(2)} y_j + a_{ii}^{(2)} y_i^2 + a_{ij}^{(2)} y_i y_j + a_{jj}^{(2)} y_j^2.$$

Оцінка моделей застосовується така ж як і на першому ряді, але відбирається $F_2 < F_1$ кращих моделей.

Процес побудови рядів повторюється поки значення функціоналу, який має зміст середнього квадрату помилки, зменшується. Якщо в ряду t отримаємо збільшення помилки, то процес завершується.

Нечіткий метод групового врахування аргументів

В роботі [96, 146] була розглянута лінійна інтервальна модель регресії:

$$Y = A_0 z_0 + A_1 z_1 + \dots + A_p z_p, \quad (1.24)$$

де A_i — нечіткі числа трикутного вигляду, які задаються двома параметрами:

$$A_i = (a_i, c_i),$$

де a_i — центр інтервалу, c_i — ширина інтервалу, $c_i \geq 0$. Відповідно Y — нечітке число, параметри якого (центр і ширина) визначаються наступним чином:

$$a_Y = \sum_{i=1}^p a_i z_i = \mathbf{a}^\top \mathbf{z}, \quad c_Y = \sum_{i=1}^p c_i |z_i| = \mathbf{c}^\top |\mathbf{z}|.$$

Інтервальна модель є коректною, якщо дійсне значення вихідної величини Y належить інтервалу невизначеності:

$$\begin{cases} \mathbf{a}^\top \mathbf{z} - \mathbf{c}^\top |\mathbf{z}| \leq y \\ \mathbf{a}^\top \mathbf{z} + \mathbf{c}^\top |\mathbf{z}| \geq y \end{cases}.$$

Якщо розглянути вибірку $\mathcal{D} = \{(y_t, \mathbf{z}_t) : t = 1, \dots, N\}$, тоді модель (1.24) є коректною, якщо множина $\{(a_i, c_i) : i = 1, \dots, p\}$ задовольняє обмеженням:

$$\begin{cases} \mathbf{a}^\top \mathbf{z}_t - \mathbf{c}^\top |\mathbf{z}_t| \leq y_t \\ \mathbf{a}^\top \mathbf{z}_t + \mathbf{c}^\top |\mathbf{z}_t| \geq y_t \end{cases}, \quad t = 1, \dots, N.$$

Таки чином для оціночної лінійної інтервальної моделі необхідно знайти такі значення параметрів (a_i, c_i) нечітких коефіцієнтів A_i , які задовольняють:

- Дійсні значення виходу y_t належать оціночному інтервалу для Y_t .
- Сумарна ширина оціночного інтервалу мінімальна.

Ці вимоги для квадратичного часткового опису можна звести до наступної задачі лінійного програмування:

$$c_0 N + c_1 \sum_{t=1}^N |x_{it}| + c_2 \sum_{t=1}^N |x_{jt}| + c_3 \sum_{t=1}^N |x_{it} \cdot x_{jt}| + \\ + c_4 \sum_{t=1}^N |x_{it}^2| + c_5 \sum_{t=1}^N |x_{jt}^2| \longrightarrow \min, \quad (1.25)$$

з обмеженнями:

$$\begin{cases} a_0 + a_1 x_{it} + a_2 x_{jt} + a_3 (x_{it} \cdot x_{jt}) + a_4 x_{it}^2 + a_5 x_{jt}^2 - c_0 - \\ - c_1 |x_{it}| - c_2 |x_{jt}| - c_3 |x_{it} \cdot x_{jt}| - c_4 |x_{it}^2| - c_5 |x_{jt}^2| \leq y_t \\ a_0 + a_1 x_{it} + a_2 x_{jt} + a_3 (x_{it} \cdot x_{jt}) + a_4 x_{it}^2 + a_5 x_{jt}^2 + c_0 + \\ + c_1 |x_{it}| + c_2 |x_{jt}| + c_3 |x_{it} \cdot x_{jt}| + c_4 |x_{it}^2| + c_5 |x_{jt}^2| \geq y_t \\ c_0, c_1, c_2, c_4, c_5 \geq 0, \quad t = 1, \dots, N \end{cases} \quad (1.26)$$

Але задача у формі (1.25-1.26) не є зручною для застосування стандартних алгоритмів, оскільки відсутня умова невід'ємності коефіцієнтів a_l . Таким чином для її вирішення необхідно перейти до двоїстої задачі, яка розв'язується симплекс-методом.

Алгоритм НМГВА

Задані: дані спостережень \mathcal{D} , загальний вигляд моделі і опорних функцій; зовнішній критерій оптимальності (функціонал L), рівень збіжності $\epsilon > 0$.

Результат: оптимальна модель $\mathcal{M}_{\text{опт}}$.

Алгоритм починає роботу з

розбиття загальної вибірки на навчальну $\mathcal{D}_{N_{\text{навч}}}$ та перевіірочну $\mathcal{D}_{N_{\text{перев}}}$;

ітератор числа рядів $r \leftarrow 0$;

кількість входів $F^{(0)} \leftarrow n$;

повторювати

генерація множини часткових моделей кандидатів на основі вибраної опорної функції: $\left\{ \mathcal{M}_i^{(r)} : i = 1, \dots, \binom{2}{F^{(r)}} \right\}$;

для кожної часткової моделі кандидата $\mathcal{M}_i^{(r)}$ виконати
 розв'язуємо задачу (1.25-1.26) і знаходимо значення (a_i, c_i) ;
 обчислюємо значення зовнішнього критерію $L_i^{(r)} = L(\mathcal{M}_i^{(r)})$;
 кінець перебору часткових моделей кандидатів;
 обчислення середнього критерію $L^{(r)}$ для часткових моделей;
 $r \leftarrow r + 1$;
 вибір F кращих моделей і формування входів нового ряду $F^{(r)} \leftarrow F$;
 поки не виконається критерій зупинки $|L^{(r+1)} - L^{(r)}| < \epsilon$;
 з F кращих моделей передостаннього ряду вибирається оптимальна модель
 $\mathcal{M}_{\text{опт}}$ і відновлюється її аналітичний вигляд;
 кінець роботи алгоритму.

1.4. Висновки і постановка задач роботи

Клас досліджуваних процесів обмежується стохастичними процесами, які представлені у вигляді моделі вхід-вихід. При цьому випадкова складова стохастичної залежності є адитивним гомоскедастичним шумом, а детермінована складова — нелінійний динамічний процес. Основною проблемою прогнозування таких процесів є наявність помилки наближення моделі, яка пов'язана з неточним вибором структури моделі, і помилки оцінювання моделі, яка пов'язана з вибором неоптимальних параметрів моделі. Наявність таких помилок пов'язана з дилемою зміщення-варіації моделі. Невизначеність зміщення-варіації моделі розкривається за допомогою механізму регуляризації. Найбільш ефективним механізмом регуляризації є байєсівський підхід. Про це свідчать результати таких змагань з інтелектуального аналізу даних і побудови моделей як DMC та CoIL Challenge⁵, які є одним із найефективніших засобів порівняння якості великої кількості існуючих підходів.

Показано, що структурованість моделі дозволяє легко визначити вплив одного або пари факторів входу на вихід і, в разі наявності функцій на-

⁵<http://www.data-mining-cup.com>, <http://www.liacs.nl/~putten/library/cc2000/>

лежності, інтерпретувати отриману модель на мові нечіткої логіки. Це допомагає правильно задавати початкову структуру моделі на основі наявної апріорної інформації, в тому числі і нечітко визначеної, і проводити подальші уточнення структури моделі. Для забезпечення структурованості доцільно використовувати розклад аналізу варіацій (ANOVA), при чому бажано обмежитися парною взаємодією факторів для значного спрощення структури і запобігання проблемі прокляття вимірності.

Задачі дисертаційної роботи сформулюємо так:

1. Створити нову інформаційну технологію синтезу нечітких баз знань пониженої складності на основі нейронечітких моделей.
2. За допомогою технології опорних векторів забезпечити робастність нейронечітких моделей для описання стохастичних процесів з домінуючою випадковою складовою.
3. Вивчити недоліки існуючих методів регуляризації некоректних задач відновлення залежностей та отримати новий удосконалений критерій адекватності нейронечітких моделей.
4. Розробити новий метод побудови нейронечітких моделей, який забезпечить високоякісне прогнозування стохастичних процесів. Метод повинен враховувати апріорну інформацію про структуру моделі, працювати на коротких вибірках і оцінювати похибки прогнозу.
5. На основі розроблених методів створити нову інформаційну систему обробки статистичних даних та застосувати її до розв'язку реальних задач прогнозування.

РОЗДІЛ 2

РОЗВИТОК НЕЙРОНЕЧІТКИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ АНАЛІЗУ СТОХАСТИЧНИХ ПРОЦЕСІВ

2.1. Нечітка логіка в нейронних моделях

Оскільки нечітка логіка узагальнює поняття характеристичної функції булевої логіки, то за допомогою лінгвістичного терму A , що визначається нечіткою множиною з функцією належності $\mu_A(x) \in [0; 1]$, можна представити нечітке твердження щодо x . Класична нечітка логіка оперує з логічними функціями і, або, якщо...тоді над нечіткими змінними [160]. Приклад застосування нечіткого логічного висновку можна знайти наприклад в роботі [161]. Але якщо замість звичайних операторів $\min(\cdot)$ та $\max(\cdot)$ застосовувати алгебраїчні оператори добутку та суми, тоді нечіткі моделі стають більш гнучкими, і забезпечують відповідність між нейронними мережами і нечіткою логікою, наприклад, якщо функції належності задаються у вигляді радіальних базисних функцій або В-сплайнів [45, 162]. Такі моделі також полегшують формальне сприйняття нелінійних систем. Дамо формальне визначення нейронечітких моделей.

Означення 2.1. Нейронечіткою моделлю в канонічній формі будемо називати лінійну комбінацію

$$f(\mathbf{x}) = \sum_i w_i \mu_{A^i}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n, \quad w_i \in \mathbb{R}, \quad (2.1)$$

де $\mu_{A^i}(\mathbf{x})$ — дійсні невід’ємні функції належності змінної \mathbf{x} , які задовольняють умові $\sum_i \mu_{A^i}(\mathbf{x}) = 1$; A^i — лінгвістичні терми визначені на \mathcal{X} .

Для нейронечітких моделей важливий наступний результат [52, 163]:

Теорема 2.2. Якщо для представлення входу $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$ та виходу $y \in \mathcal{Y} \subset \mathbb{R}$ нечіткої системи використовуються відповідно дійсні невід’ємні функції належності $\mu_{A^i}(\mathbf{x})$ та $\mu_{B^j}(y)$, де A^i , B^j — лінгвістичні терми визначені відповідно на \mathcal{X} та \mathcal{Y} , при цьому виконуються умови $\sum_i \mu_{A^i}(\mathbf{x}) = 1$, а також $\forall j : \int \mu_{B^j}(y) dy = \text{const}$; якщо функції нечіткої логіки визначені алгебраїчними операторами добутку та суми, тоді результат нечіткого логічного висновку Мамдані, дефазифікований методом середнього значення (центроїдним методом), відповідає виходу нейронечіткої моделі (2.1):

$$y(\mathbf{x}) = \sum_i w_i \mu_{A^i}(\mathbf{x}). \quad (2.2)$$

При цьому вагові коефіцієнти $w_i = \sum_j c_{ij} y_j^c$, де y_j^c — центр функції належності виходу μ_{B^j} , c_{ij} — ймовірність правила типу Мамдані: “якщо $\mathbf{x} \in A^i$, тоді $y \in B^j$ ”, при цьому $\sum_j c_{ij} = 1$, $c_{ij} \in [0; 1]$.

Доведення цієї теореми приведено в додатку В. Задача генерації нечітких правил на основі аналітичної нейронечіткої моделі зводиться до відшукування ймовірностей c_{ij} за заданими коефіцієнтами w_i . Для цього необхідно розв’язати систему лінійних рівнянь. Розв’язок легко отримати, якщо функції належності виходу наприклад задані у формі В-сплайнів (1-го або 2-го порядку). В цьому випадку максимум два коефіцієнти c_{ij} для заданого i ненульові.

В загальному випадку функції належності визначають не для всього вектору входу $\mathbf{x} = (x^1, \dots, x^n)^\top$, а для кожної вхідної змінної x^k окремо, оскільки змінні можуть мати різний фізичний зміст і класифікуватися по-різному. Нехай область визначення змінної x^k покривається скінченною кількістю нечітких множин A_k^i , $i = 1, \dots, m_k$ з функціями належності $\mu_{A_k^i}(x^k)$:

$$x^k \in \mathcal{X}_k \subseteq \bigcup_{i=1}^{m_k} \text{supp } \mu_{A_k^i}(x^k),$$

де $\text{supp } f$ — компактний носій функції f , тобто найменша замкнута область, зовні якої f тотожно рівна 0. А область визначення змінної виходу y покривається скінченною кількістю нечітких множин B^i , $i = 1, \dots, m_0$ з функціями належності $\mu_{B^i}(y)$:

$$y \in \mathcal{Y} \subseteq \bigcup_{i=1}^{m_0} \text{supp } \mu_{B^i}(y),$$

Таким чином повний набір нечітких правил, кожне з яких можна описати як

$$\begin{aligned} r\text{-те правило : } & \text{якщо } x^1 \in A_1^{i_1} \text{ і якщо } x^2 \in A_2^{i_2} \text{ і } \dots \\ & \text{якщо } x^n \in A_n^{i_n} \text{ тоді } y \in B^{i_0}, \end{aligned}$$

складатиметься з $R = m_0 m_1 \cdots m_n$ правил (індексу r відповідає послідовність $\{i_0, i_1, \dots, i_n\}$, $i_k = 0, \dots, m_k$). Традиційним застосуванням такого підходу є побудова нейронечіткої моделі на основі В-сплайнів, де для формування функцій належності для вектору входу використовується тензорний добуток В-сплайнів однієї змінної [45]:

$$\mu_{A^j}(\mathbf{x}) = \prod_{k=1}^n N_{i_k}^d(x^k), \quad j = 1, \dots, M = m_1 m_2 \cdots m_n,$$

де індексу j відповідає послідовність $\{i_1, \dots, i_n\}$. N_i^d — i -й В-сплайн порядку d , який рекурсивно визначається на основі заданої множини вузлів $\{a_i : i = 1, \dots, m\}$ наступним чином [164]:

$$\begin{aligned} N_i^0(x) &= \begin{cases} 1, & \text{якщо } x \in [a_i; a_{i+1}) \\ 0, & \text{в іншому випадку} \end{cases}, \\ N_i^d(x) &= \lambda_i(x) N_i^{d-1}(x) + (1 - \lambda_{i+1}(x)) N_{i+1}^{d-1}(x), \quad \lambda_i(x) = \frac{x - a_i}{a_{i+d} - a_i}. \end{aligned} \tag{2.3}$$

Теорема 2.2 показує, що нелінійна система за допомогою нейронечіткої моделі може бути представлена у вигляді лінійної в параметрах структури, що дозволяє застосовувати швидкі та прості процедури навчання. Основною проблемою при цьому є прокляття вимірності, пов'язане із експонентною складністю алгоритму навчання $\mathcal{O}(m^n)$, що зумовлено відповідною кількістю нечітких правил. Подальші дослідження в цьому підрозділі присвячені саме питанням зниження складності алгоритмів навчання.

2.2. Нейронечітка модель у формі Бернштейна

Технологія створення векторних зображень за допомогою кривих була винайдена на початку 60-х років П'єром Без'є, інженером фірми Renault, і Полем де Кастельжо (Кастельє), інженером компанії Citroen. Оскільки вони обидва працювали над створенням креслень нових автомобілів, всі їх дослідження до певного часу не розповсюджувалися. Робота Поля де Кастельжо була завершена першою але так жодного разу і не була опублікованою. П'єру Без'є пощастило більше. Його іменем і була названа нова технологія, хоча математичні вирази, відповідно до яких виконуються перетворення Без'є, були винайдені в основному паном де Кастельжо [165]. Математична теорія барицентричних координат і алгоритм Кастельжо, які широко використовуються при побудові кривих та поверхонь Без'є в області автоматизованого проектування трьохвимірних систем, знаходять своє застосування і при побудові нейронечітких моделей для нелінійних динамічних систем.

2.2.1. Алгоритм побудови нейронечіткої моделі у формі Бернштейна

Розглянемо алгоритм побудови нейронечітких моделей для нелінійних динамічних систем, запропонований К.Харрісом [58, 59]. Основа цього підходу полягає у використанні поліномів у формі Бернштейна однієї та двох змінних в якості базисних для представлення виходу нелінійної системи у вигляді наступного розкладу Габора-Колмогорова:

$$y(\mathbf{x}) = b + \sum_{k=1}^n B_k(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^n B_{pq}(x^p, x^q) + e(\mathbf{x}), \quad (2.4)$$

де вектор входу нелінійної системи $\mathbf{x} = (x^1, \dots, x^n)^\top \in \mathcal{X} \subset \mathbb{R}^n$. \mathcal{X} — простір входу системи розмірності n . Величина $e(\mathbf{x})$ відображає адитивну шумову складову системи. Для формування поліномів у формі Бернштейна однієї та двох змінних використовуються лінійні комбінації відповідних базисних

поліномів Бернштейна від барицентричних координат u та \mathbf{u} :

$$B_k(x^k) = \sum_{j=0}^d w_j \phi_j^d(u(x^k)), \quad (2.5)$$

$$B_{pq}(x^p, x^q) = \sum_{i+j+k=d} w_{ijk} \phi_{ijk}^d(\mathbf{u}(x^p, x^q)), \quad (2.6)$$

де $u \in \mathcal{U}^1 \subset \mathbb{R}^1$, $\mathbf{u} = (u, v)^\top \in \mathcal{U}^2 \subset \mathbb{R}^2$. \mathcal{U}^1 , \mathcal{U}^2 — простори барицентричних координат. Базисні поліноми Бернштейна порядку d однієї і двох змінних відповідно мають вигляд:

$$\begin{aligned} \phi_j^d(u) &= \binom{d}{j} \cdot u^j (1-u)^{d-j}, \\ \phi_{ijk}^d(\mathbf{u}) &= \binom{d}{i, j, k} \cdot u^i v^j (1-u-v)^k, \quad i+j+k=d. \end{aligned} \quad (2.7)$$

При цьому кожен поліном у формі Бернштейна відповідає виходу нейронечіткої підсистеми (2.1), а базисні поліноми Бернштейна виступають в ролі функцій належності для представлення її входів. Зазначимо, що за побудовою базисні поліноми Бернштейна є невід'ємними і задовольняють умовам:

$$\sum_{j=0}^d \phi_j^d(u) = 1, \quad \sum_{i+j+k=d} \phi_{ijk}^d(\mathbf{u}) = 1.$$

Тож їх можна використовувати в ролі функцій належності.

Таким чином нелінійна система (2.4), структура якої зображена на рис. 2.1, розглядається як суперпозиція таких нейронечітких підсистем. Барицентричні координати в свою чергу обчислюються за допомогою обернених відображень Кастельжо:

$$\Psi_k(\Lambda_k) : x^k \mapsto u(x^k) \in \text{supp } \phi(\cdot) = [0; 1] \quad \text{та}$$

$$\Psi_{pq}(\Lambda_{pq}) : (x^p, x^q) \mapsto \mathbf{u}(x^p, x^q) \in \text{supp } \phi(\cdot, \cdot) = \triangle\{u \geq 0, v \geq 0, u+v \leq 1\}.$$

Прямий алгоритм Кастельжо детально описаний в розділі 2.2.2. Відображення Ψ_k , Ψ_{pq} будуються на основі наперед заданих множин базових контрольних точок Λ_k , Λ_{pq} , використовуючи ітераційний чисельний метод зворотного розповсюдження помилки [166]. Після обчислення барицентричних

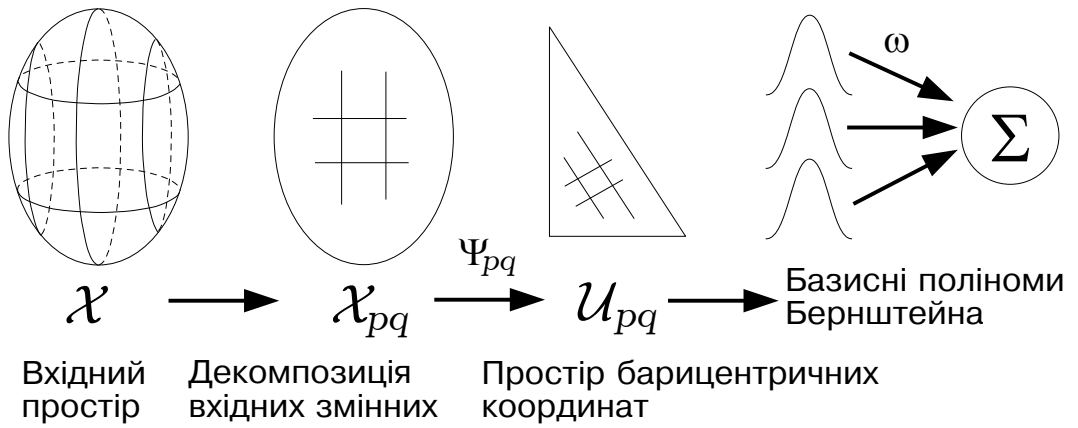


Рис. 2.1. Структура нейронечіткої моделі К.Харріса.

координат знаходяться оцінки вагових коефіцієнтів відповідних рівнянь регресії (2.4, 2.5, 2.6) двох етапним методом найменших квадратів (МНК) або ортогональним методом найменших квадратів (ОМНК) [167, 168, 169]. Тобто відбувається навчання нейронечітких підсистем.

Загальна кількість невідомих коефіцієнтів при базисних поліномах Бернштейна, які відповідають фактично нечітким правилам, становить $p = n(d + 1) + \frac{1}{4}n(n - 1)(d + 1)(d + 2)$. Тож значною перевагою підходу К.Харріса є квадратична складність алгоритму навчання $\mathcal{O}(n^2)$, порівняно з локальними нейронечіткими моделями Такагі-Сугено або з традиційними нейронечіткими моделями на основі В-сплайнів, де складність становить $\mathcal{O}(m^n)$.

Зауваження 2.3. Базисний поліном Бернштейна є частковим випадком В-сплайну.

Доведення. Дійсно, як показано в [164], якщо в (2.3) покласти

$$a_1 = \dots = a_m = 0, \quad a_{m+1} = \dots = a_{2m} = 1,$$

тоді $N_i^d(u) = \phi_i^d(u)$, $i = 1, \dots, m$, $u \in [0; 1)$. □

Будемо називати нейронечіткі моделі Харріса нейронечіткими моделями у формі Бернштейна. Недоліки підходу К.Харріса:

1. Ітеративність чисельного методу негативно впливає на швидкість його роботи, особливо при великій кількості вхідних змінних і великому розмірі вибірки.
2. Залишається відкритим питання про вибір порядку поліномів.
3. Невизначеність розміщення базових контрольних точок.
4. Необґрунтована можливість незалежного використання нечітких правил, які сформовані на основі окремих підмоделей.
5. Функції належності у формі базисних поліномів Бернштейна високих порядків не створюють достатньо розрізнених нечітких множин. Вони сильно перекриваються.

2.2.2. Класичний алгоритм Кастельжо

Класичний прямий алгоритм Кастельжо добре відомий як один з методів побудови кривих та поверхонь Без'є [164, 170]. У випадку побудови кривої Без'є задається довільна множина базових контрольних точок на площині $\{\mathbf{a}_j \in \mathbb{R}^2 : j = 1, \dots, d\}$. Для кожної барицентричної координати u , використовуючи ітеративну формулу лінійної інтерполяції двох контрольних точок:

$$\begin{aligned} \mathbf{a}_j^t(u) &= (1 - u)\mathbf{a}_j^{t-1}(u) + u\mathbf{a}_{j+1}^{t-1}(u), \quad \mathbf{a}_j^0 \equiv \mathbf{a}_j, \\ t &= 1, \dots, d; \quad j = 0, \dots, d - t; \end{aligned} \quad (2.8)$$

знаходиться точка-елемент кривої Без'є $\mathbf{a}_0^d(u)$ (рис. 2.2).

У випадку побудови поверхні Без'є задається довільна множина базових контрольних точок $\{\mathbf{r}_{ijk} \in \mathbb{R}^3 : i + j + k = d\}$. Для кожної пари барицентричних координат u і v , використовуючи ітеративну формулу лінійної інтерполяції кожних трьох контрольних точок:

$$\begin{aligned} \mathbf{r}_{ijk}^t(\mathbf{u}) &= u\mathbf{r}_{i+1jk}^{t-1}(\mathbf{u}) + v\mathbf{r}_{ij+1k}^{t-1}(\mathbf{u}) + (1 - u - v)\mathbf{r}_{ijk+1}^{t-1}(\mathbf{u}), \quad \mathbf{r}_{ijk}^0 \equiv \mathbf{r}_{ijk}, \\ t &= 1, \dots, d; \quad i + j + k = d - t; \end{aligned} \quad (2.9)$$

знаходиться точка-елемент поверхні Без'є $\mathbf{r}_{000}^d(\mathbf{u})$ (рис. 2.3).

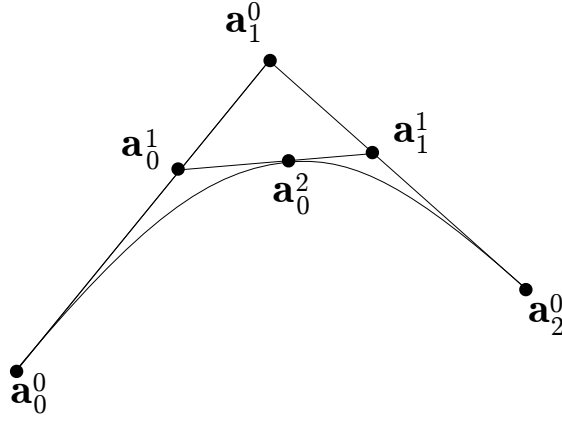


Рис. 2.2. Побудова кривої Без'є методом Кастельжо.

2.2.3. Обернений алгоритм Кастельжо

Не порушуючи загальності розглянемо обернений алгоритм Кастельжо для поверхонь Без'є. Більш простий варіант алгоритму для кривих Без'є будується аналогічно. Показано, наприклад в [170], що будь-яка точка на поверхні Без'є може бути представлена у вигляді лінійної комбінації базисних поліномів Бернштейна:

$$\mathbf{r}_{000}^d(\mathbf{u}) = \sum_{i+j+k=d} \mathbf{r}_{ijk} \phi_{ijk}^d(\mathbf{u}). \quad (2.10)$$

Цей важливий наслідок слідує з властивості рекурсивного обчислення базисних поліномів Бернштейна:

$$\phi_{ijk}^d(\mathbf{u}) = u\phi_{i-1jk}^{d-1}(\mathbf{u}) + v\phi_{ij-1k}^{d-1}(\mathbf{u}) + (1-u-v)\phi_{ijk-1}^{d-1}(\mathbf{u}).$$

Обернений алгоритм Кастельжо полягає в наступному [58, 166]. Для задачі (2.6) вектор $\mathbf{r}_{000}^d(\mathbf{u})$ розглядається як

$$\mathbf{r}_{000}^d = (x^p, x^q, B_{pq}(x^p, x^q))^{\top}.$$

У цьому випадку рівність (2.10) можна переписати у вигляді наступної системи рівнянь:

$$\begin{cases} (x^p, x^q)^{\top} = \sum_{i+j+k=d} \mathbf{b}_{ijk} \phi_{ijk}^d(\mathbf{u}) \\ B_{pq}(x^p, x^q) = \sum_{i+j+k=d} w_{ijk} \phi_{ijk}^d(\mathbf{u}) \end{cases},$$

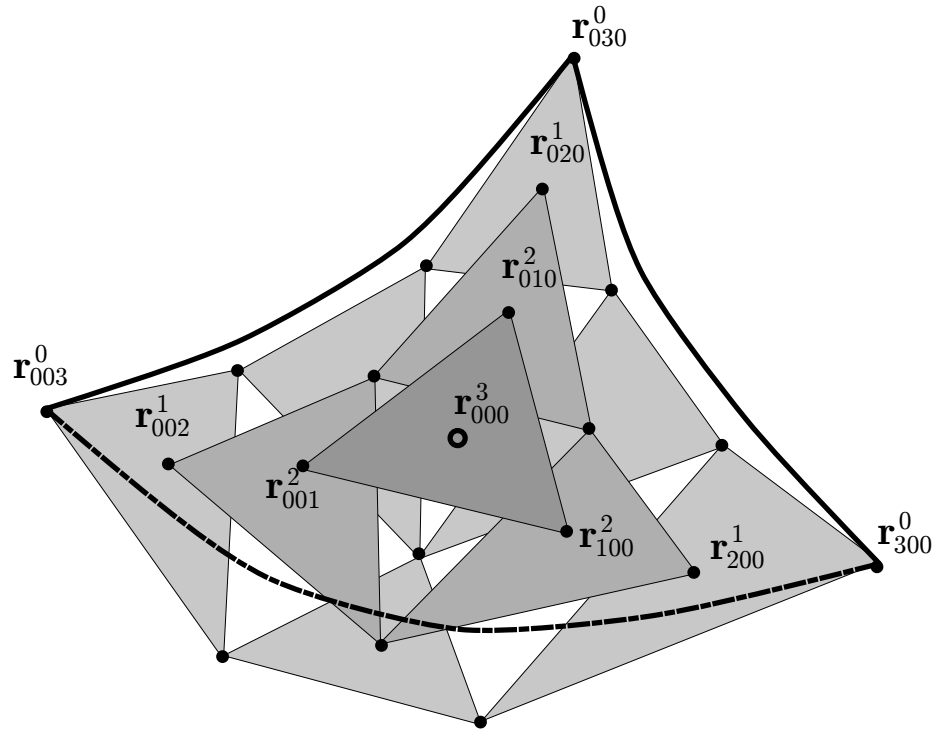


Рис. 2.3. Робота алгоритму Кастельжо для поверхні Без'є.

де \mathbf{b}_{ijk} — проекція \mathbf{r}_{ijk} на площину X_{pq} . Тоді часткова задача наближення поліному у формі Бернштейна $B_{pq}(x^p, x^q)$ поверхнею Без'є зводиться до визначення вагових коефіцієнтів w_{ijk} . Для цього необхідно побудувати відображення підпростору вхідних змінних $\mathcal{X}_{pq} = \text{span}\{\mathbf{e}^p, \mathbf{e}^q\}$ у простір барицентричних координат \mathcal{U}^2 на основі заданої множини базових контрольних точок $\Lambda_{pq} = \{\mathbf{b}_{ijk} \in \mathbb{R}^2 : i + j + k = d\}$, $\mathbf{b}_{ijk} \equiv \mathbf{b}_{ijk}^0$ (рис. 2.4). Побудова такого відображення на основі ітераційного чисельного методу зворотного розповсюдження помилки полягає в наступному:

1. Задається множина базових контрольних точок Λ_{pq} , яка формує опуклу оболонку таку, що містить всі пари (x^p, x^q) .
2. Метод застосовується для кожної пари (x^p, x^q) з вхідного підпростору. Вибираємо конкретну пару $\mathbf{x}_{pq} = (x^p, x^q)^\top$.
3. Нехай $m = 1$ — перша ітерація. Вибираємо випадково пару барицентричних координат $\mathbf{u}^{(m)} \in \Delta\{u^{(m)} \geq 0, v^{(m)} \geq 0, u^{(m)} + v^{(m)} \leq 1\}$.

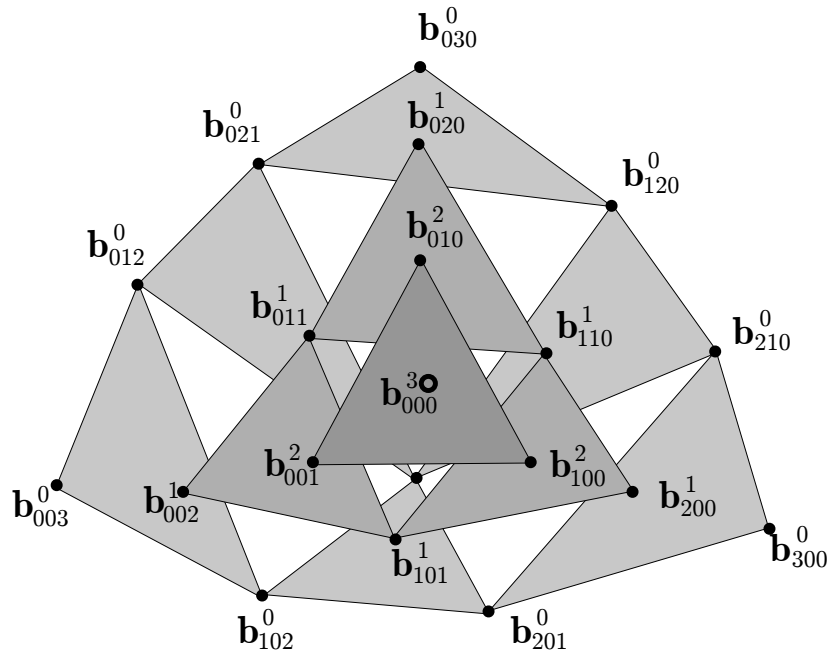


Рис. 2.4. Обчислення контрольних точок для побудови проекції поверхні Без'є на площині.

4. Для $\mathbf{u}^{(m)}$ обчислюється проекція $\hat{\mathbf{x}}_{pq}$ відповідної точки на поверхні Без'є за допомогою алгоритму Кастельжо, використовуючи аналог формули (2.9) на площині:

$$\mathbf{b}_{ijk}^t(\mathbf{u}^{(m)}) = u\mathbf{b}_{i+1jk}^{t-1}(\mathbf{u}^{(m)}) + v\mathbf{b}_{ij+1k}^{t-1}(\mathbf{u}^{(m)}) + (1-u-v)\mathbf{b}_{ijk+1}^{t-1}(\mathbf{u}^{(m)}), \\ t = 1, \dots, d; \quad i + j + k = d - t; \quad \hat{\mathbf{x}}_{pq}(\mathbf{u}^{(m)}) = \mathbf{b}_{000}^d(\mathbf{u}^{(m)}).$$

Зазначимо, що формула (2.10) не використовується, тому що не дає необхідних, як буде видно далі, проміжних значень контрольних точок.

5. За схемою зворотного розповсюдження помилки знаходиться нова точка:

$$\mathbf{x}_{pq}^*(\mathbf{u}^{(m)}) = \hat{\mathbf{x}}_{pq}(\mathbf{u}^{(m)}) + \alpha [\mathbf{x}_{pq} - \hat{\mathbf{x}}_{pq}(\mathbf{u}^{(m)})],$$

де α — рівень навчання — $0 < \alpha \ll 1$.

6. Наступна пара барицентричних координат $\mathbf{u}^{(m+1)}$ обчислюється за допомогою формули:

$$\mathbf{u}^{(m+1)} = (\mathbf{b}_{100}^{d-1} - \mathbf{b}_{001}^{d-1}, \mathbf{b}_{010}^{d-1} - \mathbf{b}_{001}^{d-1})^{-1} (\mathbf{x}_{pq}^* - \mathbf{b}_{001}^{d-1}),$$

де контрольні точки $\mathbf{b}_{001}^{d-1}, \mathbf{b}_{010}^{d-1}, \mathbf{b}_{100}^{d-1}$ отримані на кроці 4.

7. Метод продовжується поки не отримаємо пару барицентричних координат $\mathbf{u}^{(*)}$, для якої $\|\hat{\mathbf{x}}_{pq} - \mathbf{x}_{pq}\| \leq \varepsilon$, де ε — задана точність методу. Як це показано в [58], побудований таким чином метод є збіжним.

Як буде показано далі, за рахунок спеціального вибору множини базових контрольних точок, можна вивести аналітичну формулу обчислення барицентричних координат.

2.3. Швидке навчання нейронечіткої моделі у формі Бернштейна

2.3.1. Швидке обернене відображення Кастельжо

Автором запропонований новий аналітичний метод обчислення барицентричних координат в частковій задачі наближення функції двох змінних поверхнею Без'є [88, 80]. Основою методу є наступна теорема про швидке обернене відображення Кастельжо:

Теорема 2.4. Нехай для довільної множини точок $\{\mathbf{x}\} \subset \mathbb{R}^2$ задана множина базових контрольних точок $\Lambda = \{\mathbf{b}_{ijk} \in \mathbb{R}^2 : i + j + k = d\}$ порядку d така, що задовольняє умовам:

1. Опукла оболонка $\text{co}\{\mathbf{b}_{ijk}\} \supset \{\mathbf{x}\}$.
2. Множина внутрішніх точок $\text{int co}\{\mathbf{b}_{ijk}\} \neq \emptyset$.
3. Виконуються наступні рівності

$$\begin{aligned} 2\mathbf{b}_{i+1j+1k} &= \mathbf{b}_{ij+2k} + \mathbf{b}_{i+2jk}, & 2\mathbf{b}_{ij+1k+1} &= \mathbf{b}_{ij+2k} + \mathbf{b}_{ijk+2}, \\ 2\mathbf{b}_{i+1jk+1} &= \mathbf{b}_{i+2jk} + \mathbf{b}_{ijk+2}, & i + j + k + 2 &= d. \end{aligned} \quad (2.11)$$

Тоді обернене відображення Кастельжо має вигляд:

$$\Psi(\Lambda) : \mathbf{x} \mapsto \mathbf{u}(\mathbf{x}) = (\mathbf{b}_{0d0} - \mathbf{b}_{00d}, \mathbf{b}_{d00} - \mathbf{b}_{00d})^{-1} (\mathbf{x} - \mathbf{b}_{00d}), \quad (2.12)$$

при цьому $\mathbf{u} = (u, v)^\top \in \Delta\{u \geq 0, v \geq 0, u + v \leq 1\}$.

Доведення. Випишемо три суміжні кроки ітерації t ($t = 1, \dots, d-1$) прямого алгоритму Кастельжо побудови точки $\mathbf{b}_{000}^d = \mathbf{x}$ проекції поверхні Без'є по заданим барицентричним координатам $(u, v)^\top$:

$$\begin{aligned} u\mathbf{b}_{ij+2k}^{t-1} + v\mathbf{b}_{i+1j+1k}^{t-1} + (1-u-v)\mathbf{b}_{ij+1k+1}^{t-1} &= \mathbf{b}_{ij+1k}^t, \\ u\mathbf{b}_{ij+1k+1}^{t-1} + v\mathbf{b}_{i+1jk+1}^{t-1} + (1-u-v)\mathbf{b}_{ijk+2}^{t-1} &= \mathbf{b}_{ijk+1}^t, \\ u\mathbf{b}_{i+1j+1k}^{t-1} + v\mathbf{b}_{i+2jk}^{t-1} + (1-u-v)\mathbf{b}_{i+1jk+1}^{t-1} &= \mathbf{b}_{i+1jk}^t, \end{aligned} \quad (2.13)$$

та один крок наступної ітерації $t+1$, ($i+j+k = d-t-1$)

$$u\mathbf{b}_{ij+1k}^t + v\mathbf{b}_{i+1jk}^t + (1-u-v)\mathbf{b}_{ijk+1}^t = \mathbf{b}_{ijk}^{t+1}. \quad (2.14)$$

Підставивши рівності (2.11, 2.13) в (2.14), в результаті спрощень, отримаємо наступний вираз:

$$u\mathbf{b}_{ij+2k}^{t-1} + v\mathbf{b}_{i+2jk}^{t-1} + (1-u-v)\mathbf{b}_{ijk+2}^{t-1} = \mathbf{b}_{ijk}^{t+1},$$

який можна узагальнити у формі:

$$u\mathbf{b}_{0d0} + v\mathbf{b}_{d00} + (1-u-v)\mathbf{b}_{00d} = \mathbf{b}_{000}^d = \mathbf{x}. \quad (2.15)$$

За умови існування оберненої матриці (умова 2 теореми), маємо:

$$(u, v)^\top = (\mathbf{b}_{0d0} - \mathbf{b}_{00d}, \mathbf{b}_{d00} - \mathbf{b}_{00d})^{-1} \cdot (\mathbf{x} - \mathbf{b}_{00d}). \quad (2.16)$$

Умова 1 теореми забезпечує $(u, v)^\top \in \triangle\{u \geq 0, v \geq 0, u+v \leq 1\}$. □

Графічне зображення умови (2.11) приведене на рис. 2.5.

Наслідок 2.5. Для практичного застосування теореми 2.4 рівність (2.15) зручно переписати у вигляді наступної системи рівнянь:

$$\begin{cases} ub_{0d0}^p + vb_{d00}^p + wb_{00d}^p = x^p \\ ub_{0d0}^q + vb_{d00}^q + wb_{00d}^q = x^q \\ u + v + w = 1 \end{cases}, \quad \mathbf{b} = (b^p, b^q)^\top,$$

яку найпростіше розв'язати відносно u та v за допомогою правила Крамера.

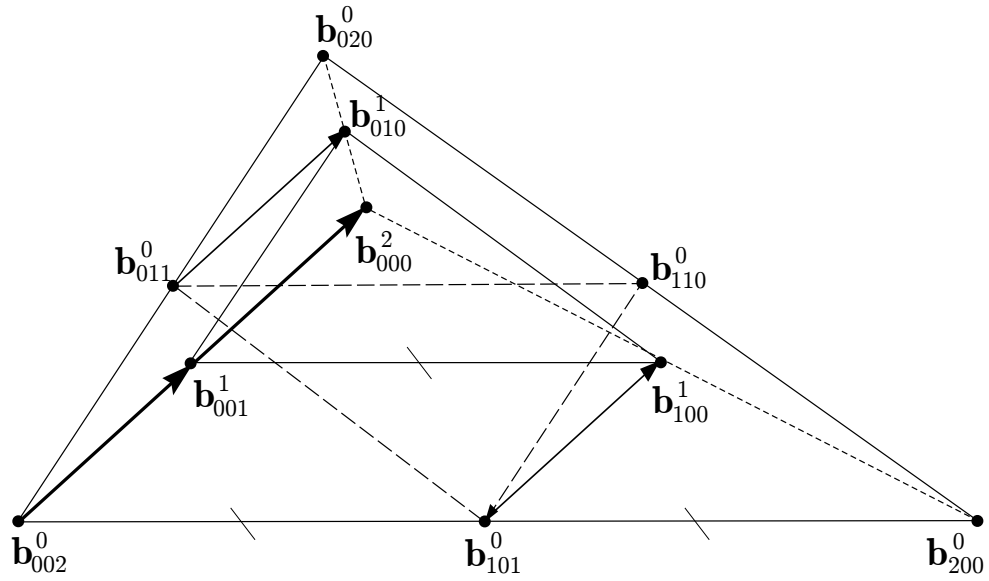


Рис. 2.5. Симетричне розташування контрольних точок проекції поверхні Без'є на площині.

Аналогічну теорему легко сформулювати для випадку оберненого відображення Кастельжо однієї змінної.

Теорема 2.6. Нехай для довільної множини точок $\{x\} \subset \mathbb{R}$ задана множина базових контрольних точок $\Lambda = \{b_j \in \mathbb{R} : j = 0, \dots, d\}$ порядку d така, що задовольняє умовам:

1. $b_0 < b_1 < \dots < b_d$, $[b_0; b_d] \supset \{x\}$.

2. Виконуються рівності:

$$2b_{j+1} = b_j + b_{j+2}, \quad j = 0, \dots, d-2. \quad (2.17)$$

Тоді обернене відображення Кастельжо має вигляд:

$$\Psi(\Lambda) : x \mapsto u(x) = (x - b_0) / (b_d - b_0), \quad (2.18)$$

при цьому $u \in [0; 1]$.

Доведення. Випишемо два суміжні кроки ітерації t ($t = 1, \dots, d-1$) прямого алгоритму Кастельжо побудови точки $b_0^d = x$ проекції кривої Без'є по заданій

барицентричній координаті u :

$$\begin{aligned} b_j^t(u) &= (1-u)b_j^{t-1}(u) + ub_{j+1}^{t-1}(u), \\ b_{j+1}^t(u) &= (1-u)b_{j+1}^{t-1}(u) + ub_{j+2}^{t-1}(u), \quad b_j^0 \equiv b_j, \\ t &= 1, \dots, d; \quad j = 0, \dots, d-t-1; \end{aligned} \quad (2.19)$$

та один крок ітерації $t+1$:

$$b_j^{t+1}(u) = (1-u)b_j^t(u) + ub_{j+1}^t(u). \quad (2.20)$$

Підставивши рівності (2.17, 2.19) в (2.20), в результаті спрощень, отримаємо наступний вираз:

$$b_j^{t+1}(u) = (1-u)b_j^{t-1}(u) + ub_{j+2}^{t-1}(u),$$

який можна узагальнити у формі:

$$b_0^d(u) = (1-u)b_0^0(u) + ub_d^0(u),$$

що і доводить (2.18). З умови 1 теореми очевидно, що $u \in [0; 1]$. \square

2.3.2. Аналітичний метод визначення барицентричних координат

З метою подолання таких недоліків підходу К. Харріса як неточність обчислення барицентричних координат, ітеративність методу зворотного розповсюдження помилки та невизначеність розміщення базових контрольних точок автором запропонований альтернативний метод обчислення барицентричних координат. Основою методу є теореми про швидке обернене відображення Кастельжо однієї змінної (теорема 2.6) та двох змінних (теорема 2.4). Метод полягає в наступному:

1. Для кожної вхідної змінної x задається множина базових контрольних точок $\{b_j : j = 0, \dots, d\}$ на відрізку $[x_{\min}; x_{\max}]$ за правилом:

$$b_j = x_{\min} + (x_{\max} - x_{\min}) \cdot j/d.$$

2. Барицентрична координата u для x знаходиться за допомогою швидкого оберненого відображення Кастельжо однієї змінної, яке визначається за формулою (2.18).
3. Для кожної пари змінних x^p, x^q задається множина базових контрольних точок $\{\mathbf{b}_{ijk} : i + j + k = d\}$ на площині \mathcal{X}_{pq} за схемою:

- Введемо поняття рівня контрольної точки (L) та номеру контрольної точки (N) на заданому рівні, поставивши у відповідність значення L та N кожній з базових контрольних точок наступним чином:

L/N	0	1	2	\dots	$d-1$	d
0	$\mathbf{b}_{d,0,0}$	—	—	\dots	—	—
1	$\mathbf{b}_{d-1,0,1}$	$\mathbf{b}_{d-1,1,0}$	—	\dots	—	—
2	$\mathbf{b}_{d-2,0,2}$	$\mathbf{b}_{d-2,1,1}$	$\mathbf{b}_{d-2,2,0}$	\dots	—	—
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
$d-1$	$\mathbf{b}_{1,0,d-1}$	$\mathbf{b}_{1,1,d-2}$	$\mathbf{b}_{1,2,d-3}$	\dots	$\mathbf{b}_{1,d-1,0}$	—
d	$\mathbf{b}_{0,0,d}$	$\mathbf{b}_{0,1,d-1}$	$\mathbf{b}_{0,2,d-2}$	\dots	$\mathbf{b}_{0,d-1,1}$	$\mathbf{b}_{0,d,0}$

- Визначимо значення проміжних допоміжних змінних:

$$\begin{aligned}
 T_{p\max} &= 2x_{p\max} - x_{p\min}, & T_{p\min} &= 2x_{p\min} - x_{p\max}, \\
 T_{q\max} &= 2.5x_{q\max} - 1.5x_{q\min}, & T_{q\min} &= 1.5x_{q\min} - 0.5x_{q\max}, \\
 \delta T_p &= (T_{p\max} - T_{p\min})/d, & \delta T_q &= (T_{q\max} - T_{q\min})/d.
 \end{aligned}$$

- Координати b_p, b_q кожної з контрольних точок визначаються за формулами:

$$\begin{aligned}
 b_p(L, N) &= T_{p\min} + (d - L) \cdot 0.5 \cdot \delta T_p + N \cdot \delta T_p, \\
 b_q(L, N) &= T_{q\min} + (d - L) \cdot \delta T_q;
 \end{aligned}$$

легко переконалися, що таким чином задані контрольні точки задовольняють всім трьом умовам теореми 2.4.

4. Барицентричні координати u, v для пари x^p, x^q знаходяться за допомогою швидкого оберненого відображення Кастельжо двох змінних, яке визначається за формулою (2.12).

2.4. Оптимальне навчання нейронечіткої моделі у формі Бернштейна

2.4.1. Регуляризація поліномів у формі Бернштейна

Зауважимо, що задачі (2.5) і (2.6) можна інтерпретувати як задачі наближення поліномів у формі Бернштейна однієї та двох змінних відповідно кривими та поверхнями Без'є [88]. Надалі, для простоти викладення, не порушуючи загальності, будемо розглядати задачу наближення поліномів у формі Бернштейна однієї змінної кривою Без'є, тобто задачу (2.5) у формі:

$$y(x) = f(x) + e(x), \quad f(x) = \sum_{j=0}^d w_j \phi_j^d(u(x|\Lambda)).$$

Залишається відкритим питання щодо визначення системи базових контрольних точок (СБКТ) Λ^* , яка б дозволила найкращим чином побудувати наближення поліноміальної функції однієї змінної кривою Без'є, тобто задовольняла певним критеріям оптимальності. Тож під регуляризацією оберненого відображення Кастельжо (або регуляризацією поліномів у формі Бернштейна) будемо розуміти визначення саме такої оптимальної СБКТ, яка забезпечує найкраще наближення поліномів у формі Бернштейна кривою Без'є за певним критерієм. Позначимо (апостеріорну) коваріаційну матрицю оцінок вагових коефіцієнтів (надалі в цьому підрозділі просто коваріаційну матрицю) як $\mathbf{\Omega}_p = \text{cov } \hat{\mathbf{w}}$, де $\mathbf{w} = (w_0, \dots, w_d)^\top$ — вектор вагових коефіцієнтів. Введемо наступні означення:

Означення 2.7. А-оптимальною СБКТ будемо називати таку, що мінімізує слід коваріаційної матриці:

$$\Lambda_{\text{A-opt}}^* : \text{tr } \mathbf{\Omega}_p \longrightarrow \min. \quad (2.21)$$

Мінімізація сліду коваріаційної матриці відповідає мінімізації середньої дисперсії оцінок коефіцієнтів, що має зміст суми квадратів головних півосей еліпсоїду розсіювання оцінок [157, 171].

Означення 2.8. D-оптимальною СБКТ будемо називати таку, що мінімізує значення визначника відповідної коваріаційної матриці:

$$\Lambda_{D\text{-opt}}^* : \det \mathbf{\Omega}_p \longrightarrow \min, \quad (2.22)$$

Такий визначник має зміст об'єму еліпсоїду розсіювання оцінок, тобто узагальненої дисперсії оцінок невідомих вагових коефіцієнтів рівняння регресії.

Означення 2.9. G-оптимальною СБКТ будемо називати таку, що мінімізує максимальну дисперсію спрогнозованих значень регресії:

$$\Lambda_{G\text{-opt}}^* : \max_{x \in \mathcal{X}} \{D[f(x)]\} \longrightarrow \min, \quad (2.23)$$

де дисперсія спрогнозованих значень визначається як

$$D[f(x)] = \mathbf{a}^\top(x) \mathbf{\Omega}_p \mathbf{a}(x), \quad \mathbf{a}(x) = (\phi_0^d(u(x|\Lambda)), \dots, \phi_d^d(u(x|\Lambda)))^\top,$$

де $u(x|\Lambda)$ — барицентричні координати визначені за допомогою оберненого відображення Кастельжо на основі СБКТ Λ для $x \in \mathcal{X} \subset \mathbb{R}$.

Означення 2.10. I-оптимальною СБКТ будемо називати таку, що мінімізує узагальнену дисперсію спрогнозованих значень регресії [172]:

$$\Lambda_{I\text{-opt}}^* : J = \int_{\mathcal{X}} D[f(x)] d\mu(x) \longrightarrow \min, \quad (2.24)$$

де μ деяка міра на просторі входу \mathcal{X} .

Припускаючи, що випадкова складова моделі (2.4) має нормальний розподіл $e(x) \sim \mathcal{N}(0, \sigma^2)$, а для отримання оцінок невідомих коефіцієнтів застосовується МНК, тоді коваріаційна матриця матиме вигляд:

$$\begin{aligned} \mathbf{\Omega}_p &= \text{cov } \hat{\mathbf{w}} = \mathbb{E} [(\hat{\mathbf{w}} - \mathbf{w})(\hat{\mathbf{w}} - \mathbf{w})^\top] = \\ &= \mathbb{E} \left((\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} - (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{f} \right) \left((\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} - (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{f} \right)^\top \\ &= (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbb{E} [(\mathbf{y} - \mathbf{f})(\mathbf{y} - \mathbf{f})^\top] \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1} = \sigma^2 (\mathbf{A}^\top \mathbf{A})^{-1}, \end{aligned}$$

де матриця $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_N)^\top = \{\phi_j^d(u(x_i|\Lambda))\}_{ij}$, $j = 0, \dots, d$; $i = 1, \dots, N$, $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)^\top$, $\mathbf{f} = (\mathbf{f}(x_1), \dots, \mathbf{f}(x_N))^\top$, N — об'єм навчальної вибірки. Надалі будемо вважати $\sigma^2 = 1$, що не обмежує загальності задач (2.21-2.24).

Застосування теорії побудови оптимальних планів в нашому випадку обмежується пасивною ідентифікацією, оскільки розглядаються системи, в яких неможливо ставити експеримент і задавати значення входів, наприклад, економічні системи. Тож мова йде про оптимальний розподіл параметрів моделі, де в ролі параметрів виступають базові контрольні точки. Як правило А-, D-оптимальність використовується для побудови моделей, для яких важлива точність і якість оцінок коефіцієнтів регресії. В роботі [173] D-оптимальність використовується для побудови робастних моделей нелінійних систем за допомогою локально регуляризованого методу найменших квадратів. В той час як G-, I-оптимальність ефективна для задач пошуку екстремуму поверхні відгуку в заданій області. Важливо визначити, які з наведених оптимізаційних задач ефективніші для визначення структури моделі, яка забезпечує найменшу похибку при прогнозуванні.

2.4.2. Застосування еволюційних методів для визначення оптимального розподілу базових контрольних точок

Відомі приклади успішного застосування генетичних алгоритмів (ГА) в задачах оптимального планування експерименту [174], для ідентифікації нелінійних залежностей [175], а також для деяких оптимізаційних задач [89, 82, 90]. Тож для розв'язку оптимізаційних задач (2.21-2.24) пропонується використовувати ГА, а саме його реалізацію з популяціями, які не перекриваються [176]. Зазначимо, що в цьому випадку задачі (2.5) і (2.6) можна віднести до класу задач навчання нейронних систем еволюційними методами.

Об'єкти ГА в нашому випадку мають наступний зміст. В якості гену виступає базова контрольна точка. Набір алелей характеризує множину можливих значень контрольних точок. Кожна особина (хромосома) складається з

набору генів, який відповідає системі контрольних точок. Мірою придатності кожної особини, яка виражає той чи інший критерій оптимальності, виступає цільова функція вигляду [81]:

$$F_{A\text{-opt}}(\Lambda) \sim -\text{tr } \mathbf{\Omega}_p, \quad F_{D\text{-opt}}(\Lambda) \sim -\det \mathbf{\Omega}_p.$$

В результаті роботи ГА на тестових моделях були отримані оптимальні розподіли базових контрольних точок для різних розподілів вектору входу [83, 84]. При цьому точність апроксимації і короткострокового прогнозування виходу системи за допомогою моделі К.Харріса зросла в декілька разів, оскільки наведені вище оптимізації зменшують також і значення середньоквадратичної помилки. Хоч швидкість роботи наведеного ГА відносно невелика, вона може бути підвищена за рахунок використання паралельної обробки інформації [177]. Окремою темою для досліджень в цьому напрямку є питання про збіжність ГА. Хоч ГА іноді знаходить гарні розв'язки для складних задач, проте теоретичних результатів щодо збіжних властивостей і умов збіжності дуже мало. Одна із спроб дослідити ГА на збіжність — створення відповідної моделі ланцюга Маркова і ймовірнісний аналіз збіжності найпридатнішої особини до глобального екстремуму [178]. Для задач побудови довгострокового прогнозу проведення таких оптимізацій як А, D, I, G недостатньо, оскільки середньоквадратична помилка, яка використовується при побудові регресії на всіх точках вибірки, є внутрішнім критерієм системи. Для однозначного вибору моделі (параметрів моделі) як правило використовується принцип зовнішнього доповнення і відповідно зовнішні критерії [94]. В нашому випадку в якості зовнішнього доповнення використаємо додаткову перевірочну вибірку для наступного критерію оптимальності.

Означення 2.11. Оптимальною за виходом (О-оптимальною) СБКТ будемо називати таку, що мінімізує функцію ризику:

$$\Lambda_{O\text{-opt}}^* : R_p(\cdot) \longrightarrow \min, \quad (2.25)$$

де функція ризику може бути визначена в різний спосіб:

$$R_p(\cdot) = \langle \mathbf{E}|\delta(x)|, \mathbf{E}|\delta(x)|^2, \max_x |\delta(x)| \rangle, \quad x \in \mathcal{D}_{N_p} \subset \mathcal{D},$$

де $\delta(x) = y(x) - f(x)$ є похибкою наближення поліному у формі Бернштейна однієї змінної кривою Без'є. \mathcal{D}_{N_p} — перевірна вибірка даних.

При $R_p(\cdot) = \mathbf{E}|\delta(x)|^2$ оптимізаційна задача (2.25) виражає відомий середньо-квадратичний критерій регулярності (1.23). Цільова функція ГА для знаходження СБКТ оптимальної за виходом має вигляд:

$$F_{\text{O-opt}}(\Lambda) \sim -R_p(\cdot).$$

Мінімальну похибку прогнозу моделі К.Харріса було отримано саме із застосуванням оптимальної за виходом СБКТ, яка забезпечувала оптимальне нечітке представлення входу [91].

2.4.3. Деякі оцінки оптимальностей

Оцінка І-оптимальності

Функціонал (2.24) легко спрощується до вигляду $J = \text{tr}(\mathbf{M}\mathbf{\Omega}_p)$, де матриця моментів вхідного простору $\mathbf{M} = \int_{\mathcal{X}} \mathbf{a}(x)\mathbf{a}^\top(x)d\mu(x)$. Покажемо це:

$$J = \text{tr}(\mathbf{J} \equiv J) = \int_{\mathcal{X}} \text{tr}(\mathbf{a}^\top(x)\mathbf{\Omega}_p\mathbf{a}(x))d\mu(x) =$$

$$\int_{\mathcal{X}} \text{tr}(\mathbf{a}(x)\mathbf{a}^\top(x)\mathbf{\Omega}_p)d\mu(x) = \text{tr}\left(\int_{\mathcal{X}} \mathbf{a}(x)\mathbf{a}^\top(x)d\mu(x) \cdot \mathbf{\Omega}_p\right),$$

оскільки для будь-яких квадратних матриць \mathbf{A} і \mathbf{B} виконується рівність $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. Цей результат полегшує обчислення показника І-оптимальності, оскільки в багатьох випадках матрицю моментів можна визначити наперед. Крім того, очевидно, що у випадку ортогональності векторів $\mathbf{a}(x)$, матриця моментів є одиничною матрицею і І-оптимальність зводиться до А-оптимальності.

Оцінка базисних поліномів Бернштейна

Твердження 2.12. Будь-який базисний поліном Бернштейна однієї змінної порядку d допускає оцінку зверху:

$$\forall u \in (0; 1), \forall j : \phi_j^d(u) \lesssim \rho(u) = \frac{1}{\sqrt{2\pi du(1-u)}}, \quad (2.26)$$

при чому $\phi_j^d(u) \approx \rho(u)$ при $u = j/d$ (рис. 2.6).

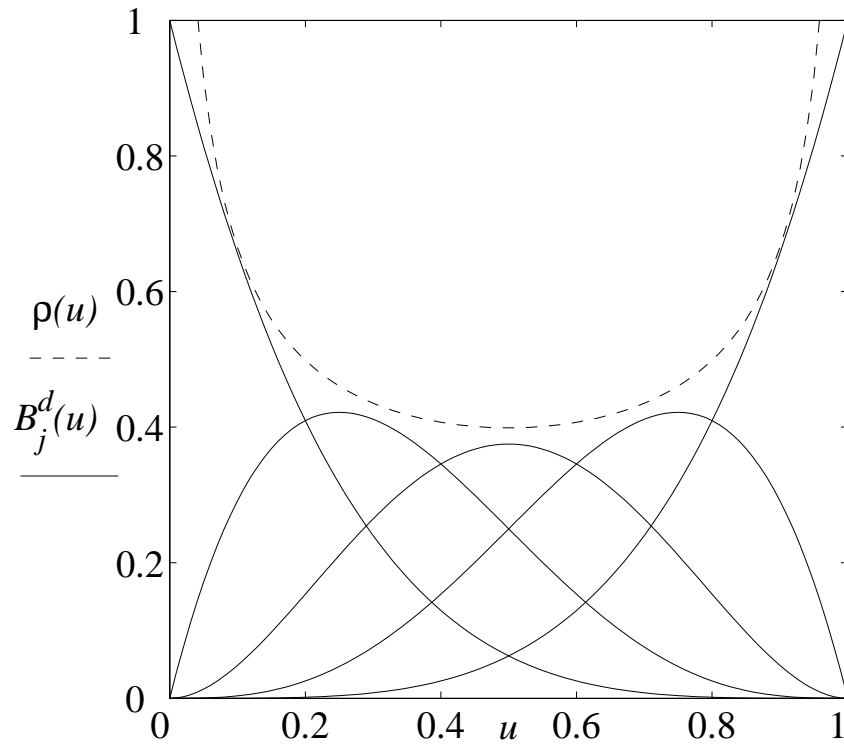


Рис. 2.6. Оцінка базисних поліномів Бернштейна 4-го порядку зверху.

Доведення. Базисні поліноми Бернштейна однієї змінної порядку d мають вигляд:

$$\phi_j^d(u) = \binom{d}{j} \cdot u^j (1-u)^{d-j},$$

і до них, як і для всяких біноміальних розподілів, можна застосувати нормальне наближення. Для довільного симетричного біноміального розподілу справедливі наступні оцінки його членів, які перенумеровані в залежності від

їх відстані до максимальної ймовірності [179]:

$$a_k = \binom{d}{m+k} \cdot u^{m+k}(1-u)^{d-m-k} \sim hr(hk),$$

$$k = -m, \dots, 0, \dots, d-m, \quad j = m+k,$$

де $m = du + \delta$, $u - 1 < \delta \leq u$ — індекс максимальної ймовірності розподілу, $h = \frac{1}{\sqrt{du(1-u)}}$, $r(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ — щільність нормального розподілу. Точність такого наближення детально розглядається в роботі [179], де наведена оцінка логарифмічної похибки наближення ς_k :

$$|\varsigma_k| < \frac{k^3}{(du(1-u))^2} + \frac{2k}{du(1-u)}.$$

Очевидно що

$$\forall u \in (0; 1), \forall j : \phi_j^d(u) = a_k \leq a_0 \sim hr(h \cdot 0) = \frac{1}{\sqrt{du(1-u)}} \cdot \frac{1}{\sqrt{2\pi}} = \rho(u),$$

що доводить оцінку зверху для базисних поліномів Бернштейна (2.26). Крім того, оцінка максимальної ймовірності $a_0 \sim \rho(u)$ є одночасно і оцінкою максимуму кожного базисного поліному Бернштейна при $u \rightarrow j/d$:

$$\forall u \in [0; 1], \forall j : \phi_j^d(u) \leq \phi_j^d\left(u_j = \frac{j}{d}\right) \sim \rho(u_j) = \frac{1}{\sqrt{2\pi du_j(1-u_j)}},$$

при чому $\forall k : \phi_j^d(u_j) \geq \phi_k^d(u_j)$. □

Оцінка сліду інформаційної матриці Фішера

Твердження 2.13. Слід інформаційної матриці Фішера від системи базисних поліномів Бернштейна допускає наближення (рис. 2.7):

$$\text{tr}(\mathbf{A}^\top \mathbf{A}) \approx \sum_{i=1}^N \frac{1}{2\sqrt{\pi du_i(1-u_i)}}. \quad (2.27)$$

Доведення. Враховуючи вище наведені оцінки для базисних поліномів Бернштейна, справедливо наступне:

$$\sum_{j=0}^d (\phi_j^d(u_i))^2 \sim \sum_{k=-m}^{d-m} (hr(kh))^2 \stackrel{\langle k \rightarrow \xi \rangle}{\approx} \frac{1}{2\pi} h^2 \int_{-\infty}^{+\infty} e^{-\xi^2 h^2} d\xi \stackrel{\langle x = \sqrt{2}\xi h \rangle}{=}$$

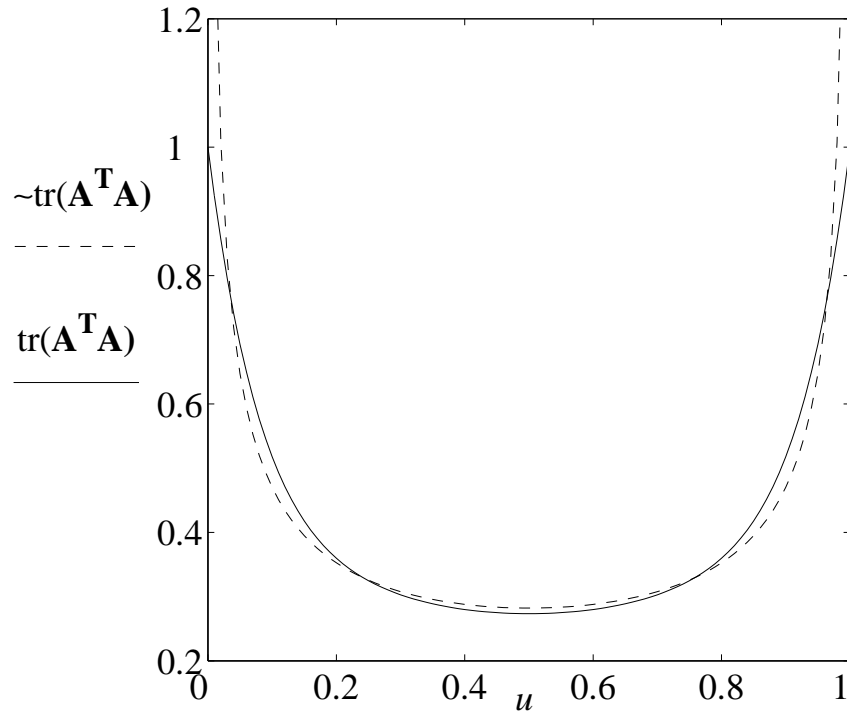


Рис. 2.7. Оцінка сліду інформаційної матриці Фішера.

$$= \frac{1}{2\pi} \frac{h}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \frac{h}{2\sqrt{\pi}} = \frac{1}{2\sqrt{\pi du_i(1-u_i)}}.$$

Відповідно для сліду інформаційної матриці Фішера (надалі просто інформаційної матриці) отримаємо:

$$\text{tr}(\mathbf{A}^\top \mathbf{A}) = \sum_{j=0}^d \sum_{i=1}^N (\phi_j^d(u(x_i|\Lambda)))^2 \approx \sum_{i=1}^N \frac{1}{2\sqrt{\pi du_i(1-u_i)}}.$$

□

З твердження 2.13 можна зробити деякі висновки:

- ▷ Очевидно, що чим ближчі барицентричні координати до центру (до 0.5), тим менше значення сліду інформаційної матриці.
- ▷ При рівномірному розподілі барицентричної координати на $[0; 1]$ можна перейти від граничних сум до визначеного інтегралу, який легко обчислити через тригонометричну заміну:

$$\text{tr}(\mathbf{A}^\top \mathbf{A}) \sim \frac{N}{2\sqrt{\pi d}} \int_0^1 \frac{du}{\sqrt{u(1-u)}} = \frac{N}{2\sqrt{\pi d}} \cdot \pi = \frac{N}{2} \sqrt{\frac{\pi}{d}}. \quad (2.28)$$

Оцінка A,D-оптимальності на основі власних чисел

Зазначимо, що коваріаційна матриця — це симетрична, позитивно визначена матриця, власні числа якої дійсні та невід’ємні, а їх сума рівна сліду матриці. За допомогою нерівностей між середнім арифметичним та середнім геометричним можна отримати оцінку знизу для сліду коваріаційної матриці:

$$\begin{aligned} \text{tr } \mathbf{\Omega}_p = \text{tr}(\mathbf{A}^\top \mathbf{A})^{-1} &= \sum_{i=0}^d \frac{1}{\lambda_i} \geq (d+1) \sqrt[d+1]{\prod_{i=0}^d \frac{1}{\lambda_i}} \geq \\ &\geq (d+1)^2 \left[\sum_{i=0}^d \lambda_i \right]^{-1} = (d+1)^2 [\text{tr}(\mathbf{A}^\top \mathbf{A})]^{-1}, \end{aligned} \quad (2.29)$$

де λ_i^{-1} — власні числа коваріаційної матриці, λ_i — власні числа інформаційної матриці. Аналогічно визначник коваріаційної матриці, який дорівнює добутку власних чисел, також допускає оцінку знизу:

$$\begin{aligned} \det \mathbf{\Omega}_p = \det(\mathbf{A}^\top \mathbf{A})^{-1} &= \prod_{i=0}^d \frac{1}{\lambda_i} \geq (d+1)^{d+1} \left[\sum_{i=0}^d \lambda_i \right]^{-d-1} = \\ &= (d+1)^{d+1} [\text{tr}(\mathbf{A}^\top \mathbf{A})]^{-d-1} \approx (d+1)^{d+1} \left[\sum_{i=1}^N \frac{1}{2\sqrt{\pi d u_i(1-u_i)}} \right]^{-d-1}. \end{aligned} \quad (2.30)$$

Очевидно, що ті розподіли барицентричних координат, для яких в нерівностях (2.29, 2.30) мають місце рівності, тобто A,D-оптимальні розподіли, співпадають і можуть визначатися умовою [171]:

$$\mathbf{A}^\top \mathbf{A} \approx (d+1)^{-1} \sum_{i=1}^N \frac{1}{2\sqrt{\pi d u_i(1-u_i)}} \mathbf{I}_{d+1}.$$

Але легко показати, що така матриця не існує. Це слідує з того, що базисні поліноми Бернштейна, з яких складається матриця \mathbf{A} , додатні. Тож можна зробити висновок про те, що оптимальність у нашому випадку це паритет двох критеріїв:

1. Кожна барицентрична координата повинна знаходитись якомога ближче до центру області визначення.

2. Барицентричні координати повинні розташовуватися якомога далі одна від одної, з метою покращення обумовленості інформаційної матриці.

Зауважимо, що наведені вище оцінки знизу для A , D -оптимальностей не слід використовувати як критерії мінімізації самих оптимальностей, тобто мінімізація сліду чи визначника коваріаційної матриці не еквівалентна максимізації сліду інформаційної матриці. Зазначимо також, що для неперервних планів справедливою є теорема еквівалентності [157], згідно якої неперервний D -оптимальний план є також і G -оптимальним.

2.5. Збалансовані нейронечіткі моделі

Головною проблемою використання класичних нечітких баз знань є їх експонентна складність, яка виражена у великій кількості нечітких правил. Це викликано необхідністю створення нового правила для кожної нової комбінації належностей факторів до нечітких множин. Більш детально це розглянуто в підрозділі 2.1. Це робить нечіткі бази знань непрозорими для експертів, оскільки практично неможливо застосовувати їх напряду для отримання корисної інформації. Більш того, для здійснення нечіткого логічного висновку необхідно знати значення всіх факторів, що входять до правила. Тобто неможливо зробити висновок за відсутності знань щодо хоча б одного з факторів.

З метою підвищення прозорості і зменшення кількості нечітких правил доцільно провести декомпозицію нечіткої бази знань. Фактично це означає, що доцільно провести декомпозицію нейронечіткої моделі на менші підмоделі, які описують залежність вихідного фактору від окремого вхідного фактору або максимум пари вхідних факторів. Таким чином складність нечіткої бази знань може бути значно зменшена — з експонентної до квадратичної. При цьому повинна існувати можливість застосовувати нечіткий логічний висновок окремо для кожної нейронечіткої підмоделі та комбінувати результати цих висновків. Саме цього намагаються досягти узагальнені нейронечіткі моделі

К.Харріса (нейронечіткі моделі у формі Бернштейна), які використовують поліноми у формі Бернштейна однієї та двох змінних в якості підмоделей. Проте залишається необґрунтованим незалежне використання таких підмоделей. В цьому підрозділі ми сформуємо спеціальний клас так званих збалансованих нейронечітких моделей, для яких незалежне використання підмоделей буде нами обґрунтовано. І покажемо, яким чином нейронечіткі моделі у формі Бернштейна можуть бути зведені до збалансованих.

Означення 2.14. Нейронечіткі моделі вигляду

$$f(\mathbf{x}) = b + \sum_{k=1}^n g_k(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^n g_{pq}(x^p, x^q),$$

де

$$g_k(x^k) = \sum_i w_i^k \mu_{A_k^i}(x^k), \quad g_{pq}(x^p, x^q) = \sum_j w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q),$$

будемо називати збалансованими, якщо математичне сподівання функцій g_k , g_{pq} за умови багатовимірного рівномірного розподілу випадкової величини \mathbf{x} на своїй області визначення дорівнює нулю. При цьому нейронечіткі підмоделі визначаються як $f_k = b + g_k$ та $f_{pq} = b + g_{pq}$. Величина b називається зміщенням нейронечіткої збалансованої моделі і характеризує середнє значення виходу.

Таким чином, за відсутності знань щодо значення змінної x^k логічно припустити, що змінна x^k з однаковою ймовірністю може приймати будь-яке значення зі своєї області визначення. При цьому вплив цієї змінної на вихід моделі після розкриття невизначеності за допомогою математичного сподівання буде нульовим. Таким чином збалансованість дозволяє генерувати нечіткі правила на основі окремої підмоделі незалежно від інших підмоделей. Означення збалансованості коректне і не суперечить означенню 2.1, оскільки нейронечіткі підмоделі можуть бути легко приведені до канонічної форми:

$$f_k(x^k) = b + \sum_i w_i^k \mu_{A_k^i}(x^k) = \sum_i (w_i^k + b) \mu_{A_k^i}(x^k),$$

$$f_{pq}(x^p, x^q) = b + \sum_j w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q) = \sum_j (w_j^{pq} + b) \mu_{A_{pq}^j}(x^p, x^q),$$

в силу властивостей $\sum_i \mu_{A_k^i}(x^k) = 1$, $\sum_j \mu_{A_{pq}^j}(x^p, x^q) = 1$.

Лема 2.15 (достатня умова збалансованості нейронечіткої моделі). Нехай випадкова величина \mathbf{x} рівномірно розподілена на своїй області визначення. Тоді якщо

$$\forall i : \mathbb{E} [\mu_{A_k^i}(x^k)] = \text{const}_k, \quad \forall j : \mathbb{E} [\mu_{A_{pq}^j}(x^p, x^q)] = \text{const}_{pq}, \quad (2.31)$$

а також сума коефіцієнтів при функціях належності рівна нулю:

$$\sum_i w_i^k = 0, \quad \sum_j w_j^{pq} = 0,$$

тоді відповідна нейронечітка модель є збалансованою.

Доведення леми про достатню умову збалансованості. Як легко бачити, математичне сподівання функцій $g_k(x^k)$, $g_{pq}(x^p, x^q)$ рівне нулю:

$$\mathbb{E} [g_k(x^k)] = \sum_i w_i^k \mathbb{E} [\mu_{A_k^i}(x^k)] = 0 \cdot \text{const}_k = 0,$$

$$\mathbb{E} [g_{pq}(x^p, x^q)] = \sum_j w_j^{pq} \mathbb{E} [\mu_{A_{pq}^j}(x^p, x^q)] = 0 \cdot \text{const}_{pq} = 0.$$

□

Теорема 2.16. Якщо виконується умова (2.31), але сума коефіцієнтів при функціях належності не рівна нулю:

$$\sum_i w_i^k = \Lambda_k \neq 0, \quad \sum_j w_j^{pq} = \Lambda_{pq} \neq 0,$$

то нейронечітку модель можна звести до збалансованої.

Доведення. Для кожної підмоделі виконаємо наступні перетворення:

$$g_k(x^k) = \sum_{i=1}^{L_k} w_i^k \mu_{A_k^i}(x^k) = \sum_{i=1}^{L_k} \left(w_i^k - \frac{\Lambda_k}{L_k} \right) \mu_{A_k^i}(x^k) + \frac{\Lambda_k}{L_k},$$

$$g_{pq}(x^p, x^q) = \sum_{j=1}^{L_{pq}} w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q) = \sum_{j=1}^{L_{pq}} \left(w_j^{pq} - \frac{\Lambda_{pq}}{L_{pq}} \right) \mu_{A_{pq}^j}(x^p, x^q) + \frac{\Lambda_{pq}}{L_{pq}}.$$

Перевизначимо зміщення та коефіцієнти при функціях належності

$$\tilde{w}_i^k \longleftarrow w_i^k - \frac{\Lambda_k}{L_k}, \quad \tilde{w}_j^{pq} \longleftarrow w_j^{pq} - \frac{\Lambda_{pq}}{L_{pq}}, \quad \tilde{b} \longleftarrow b + \sum_{k=1}^n \frac{\Lambda_k}{L_k} + \sum_{p>q}^n \frac{\Lambda_{pq}}{L_{pq}}.$$

Легко перевірити, що новостворена нейронечітка модель вигляду

$$\tilde{f}(\mathbf{x}) = \tilde{b} + \sum_{k=1}^n \tilde{g}_k(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^n \tilde{g}_{pq}(x^p, x^q),$$

де

$$\tilde{g}_k(x^k) = \sum_i \tilde{w}_i^k \mu_{A_k^i}(x^k), \quad \tilde{g}_{pq}(x^p, x^q) = \sum_j \tilde{w}_j^{pq} \mu_{A_{pq}^j}(x^p, x^q),$$

задовольняє обом умовам леми 2.15 і є збалансованою. При цьому $\tilde{f}(\mathbf{x}) = f(\mathbf{x})$. \square

Теорему 2.16 можна застосувати для нейронечітких моделей у формі Бернштейна і отримати збалансовані нейронечіткі моделі у формі Бернштейна, оскільки справедливі наступні твердження, які задовольняють умову (2.31). При цьому обернене відображення Кастельжо повинно бути побудоване так, щоб рівномірний розподіл змінної \mathbf{x} на своїй області визначення індукував рівномірний розподіл відповідних барицентричних координат.

Твердження 2.17. Математичне сподівання будь-якого базисного поліному Бернштейна $\phi_j^d(u)$ порядку d за умови рівномірного розподілу випадкової величини u на $[0; 1]$ рівне $1/(d+1)$.

Доведення. Доведемо, що математичне сподівання всіх базисних поліномів Бернштейна порядку d однакове. Для цього достатньо показати, що $\mathbb{E}[\phi_j^d(u)] = \mathbb{E}[\phi_{j+1}^d(u)]$.

$$\mathbb{E}[\phi_j^d(u)] = \int_0^1 \phi_j^d(u) P(u) du = \int_0^1 \binom{d}{j} u^j (1-u)^{d-j} P(u) du.$$

Оскільки випадкова величина u розподілена рівномірно на $[0; 1]$, то щільність розподілу $P(u) = 1$. Обчислюючи інтеграл за частинами, маємо

$$\begin{aligned} \mathbb{E} [\phi_j^d(u)] &= \int_0^1 \binom{d}{j} u^j (1-u)^{d-j} du = \int_0^1 \binom{d}{j} \frac{1}{j+1} (1-u)^{d-j} du^{j+1} = \\ &= \binom{d}{j} \frac{1}{j+1} (1-u)^{d-j} u^{j+1} \Big|_0^1 - \int_0^1 \binom{d}{j} \frac{1}{j+1} u^{j+1} d(1-u)^{d-j} = \\ &= 0 - 0 + \int_0^1 \binom{d}{j} \frac{d-j}{j+1} u^{j+1} (1-u)^{d-j-1} du = \\ &= \int_0^1 \binom{d}{j+1} u^{j+1} (1-u)^{d-j-1} du = \mathbb{E} [\phi_{j+1}^d(u)], \end{aligned}$$

Це також означає, що $\mathbb{E} [\phi_j^d(u)] = \mathbb{E} [\phi_d^d(u)]$.

$$\mathbb{E} [\phi_d^d(u)] = \int_0^1 \binom{d}{d} u^d (1-u)^{d-d} du = \frac{u^{d+1}}{d+1} \Big|_0^1 = \frac{1}{d+1}.$$

Цей результат можна отримати і з інших міркувань. А саме, якщо загальна кількість базисних поліномів рівна $d+1$, а площі під базисними поліномами Бернштейна однакові і в сумі дорівнюють 1 (це слідує з властивості базисних поліномів Бернштейна), то кожна площа рівна $1/(d+1)$. \square

Аналогічно можна довести що:

Твердження 2.18. Математичне сподівання будь-якого базисного поліному Бернштейна $\phi_{ijk}^d(\mathbf{u})$ порядку d за умови рівномірного розподілу випадкової величини \mathbf{u} на $\triangle\{u \geq 0, v \geq 0, u+v \leq 1\}$ рівне $2/((d+1)(d+2))$.

2.6. Висновки до розділу

Метод побудови нейронечітких моделей у формі Бернштейна, в якому базисні поліноми Бернштейна використовуються в якості функцій належності, доцільно використовувати для синтезу нечітких баз знань квадратичної складності. Серед недоліків базового методу слід відзначити невизначеність

вибору базових контрольних точок та негативний вплив ітеративності чисельного методу знаходження барицентричних координат на швидкодію алгоритму. Розроблений автором прискорений метод навчання нейронечітких моделей у формі Бернштейна, де для визначення барицентричних координат використовується швидко обернене відображення Кастельжо, позбавлений цих недоліків. При цьому швидкість навчання нейронечітких моделей у формі Бернштейна на тестових вибірках зросла в 3–4 рази. Для обґрунтування незалежного використання нечітких правил, які сформовані на основі окремих підмоделей, автором введені збалансовані нейронечіткі моделі. Тож при формуванні нечітких баз знань пониженої складності слід спиратися на збалансовані нейронечіткі моделі у формі Бернштейна.

Теорія побудови оптимальних планів дозволяє забезпечити робастність нейронечітких моделей у формі Бернштейна. Так, А-оптимальна система базових контрольних точок мінімізує слід матриці коваріації оцінок вагових коефіцієнтів нейронечіткої моделі. Відповідно для визначення барицентричних координат використовується оптимальне обернене відображення Кастельжо. Для визначення оптимального розміщення контрольних точок можна використовувати еволюційні методи, зокрема генетичний алгоритм. Проте використання такого класу алгоритмів пов'язано з значними обчислювальними витратами. І в наступному розділі буде розглянуто інший спосіб забезпечення робастності оцінок параметрів регресії, а саме — регресія опорних векторів.

РОЗДІЛ 3

ТЕХНОЛОГІЯ ОПОРНИХ ВЕКТОРІВ. ПОБУДОВА БАЙЄСІВСЬКОЇ РЕГРЕСІЇ ОПОРНИХ ВЕКТОРІВ У ФОРМІ БЕРНШТЕЙНА

3.1. Регресія опорних векторів. Робастність

Розглянемо популярний в останні роки підхід, який започаткований В.Вапником [17, 23] і відомий як механізм опорних векторів (МОВ). Це універсальний алгоритм побудови моделей, що навчаються, заснований на теорії статистичного навчання [180], який реалізує принцип мінімізації структурного ризику. В основу покладені ідеї методу потенціальних функцій М.Айзермана [181].

Механізм опорних векторів може бути застосований для задач побудови регресії, якщо відповідним чином визначити критерій оптимізації, тобто функцію ризику, а саме — застосувати міру відстані.

Припустимо, що справжній процес описується невідомою функцією $y(\mathbf{x})$ вектору вхідних змінних $\mathbf{x} \in \mathcal{X}$, де розмірність вхідного простору $\dim \mathcal{X} = n$. Позначимо через $\mathcal{M}(\mathbf{x}, \mathbf{w})$ сімейство (ансамбль, клас) функцій f параметризоване вектором параметрів \mathbf{w} . Припустимо також, що невідома залежність $y(\mathbf{x})$ належить сімейству $\mathcal{M}(\mathbf{x}, \mathbf{w})$. Позначимо через $\hat{\mathbf{w}}$ значення \mathbf{w} , яке мінімізує деяку міру похибки між $y(\mathbf{x})$ та $\mathcal{M}(\mathbf{x}, \hat{\mathbf{w}})$. Метою ідентифікації процесу є знаходження оцінки невідомого вектору параметрів $\hat{\mathbf{w}}$ за даними навчальної вибірки $\mathcal{D} = \{(y_j, \mathbf{x}_j) : j = 1, \dots, N\}$ за умови, що шум адитивний:

$$y_j = f(\mathbf{x}_j) + \delta_j,$$

де δ_j — незалежні і однаково розподілені випадкові величини.

3.1.1. Лінійна регресія опорних векторів

Розглянемо клас лінійних функцій:

$$\mathcal{M}(\mathbf{x}, \mathbf{w}) = \langle \mathbf{w}, \mathbf{x} \rangle_{\mathcal{X}} + b, \quad \mathbf{w} \in \mathcal{X}, \quad b \in \mathbb{R},$$

де $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ означає скалярний добуток в просторі \mathcal{X} . Класична лінійна регресія опорних векторів має за мету знайти таке значення вектору параметрів \mathbf{w} , що відповідна функція $f(\mathbf{x}) = \mathcal{M}(\mathbf{x}, \mathbf{w})$ має не більше ніж ϵ -відхилення від реальних значень виходу y_j і при цьому є якомога гладкою [25]. Одним із поширених способів забезпечити найбільшу гладкість є мінімізація норми вектору параметрів [182, 24, 22], тобто $\|\mathbf{w}\|_{\mathcal{X}}^2 = \langle \mathbf{w}, \mathbf{w} \rangle_{\mathcal{X}}$. Тоді ідентифікація процесу зводиться до наступної задачі випуклої оптимізації:

$$\frac{1}{2} \|\mathbf{w}\|_{\mathcal{X}}^2 \longrightarrow \min,$$

з обмеженнями:

$$\begin{cases} y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} - b \leq \epsilon \\ -y_j + \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} + b \leq \epsilon \end{cases}.$$

Однак задача випуклої оптимізації в такому вигляді не завжди розв'язна, оскільки система рівнянь обмежень може бути несумісною. Тож набагато зручніше розглядати наступну (м'яку) постановку:

$$\frac{1}{2} \|\mathbf{w}\|_{\mathcal{X}}^2 + \beta \sum_{j=1}^N (\xi_j + \xi_j^*) \longrightarrow \min, \quad (3.1)$$

з обмеженнями:

$$\begin{cases} y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} - b \leq \epsilon + \xi_j \\ -y_j + \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} + b \leq \epsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0 \end{cases}. \quad (3.2)$$

де ξ_j, ξ_j^* — штрафні змінні (змінні нежорсткості) за порушення того чи іншого обмеження. Константа $\beta > 0$ грає роль балансу між гладкістю функції f і

ступінню виконання обмежень, тобто точністю апроксимації. Фактично β має зміст коефіцієнту регуляризації.

В контексті статистичної теорії навчання задачу ідентифікації процесу можна сформулювати як задачу мінімізації регуляризованого ризику представленого як [183]:

$$R_{reg}(f) = R_{emp}(f) + \lambda \Theta(f), \quad \lambda \geq 0$$

де емпіричний ризик:

$$R_{emp}(f) = \frac{1}{N} \sum_{j=1}^N \mathcal{C}(y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} - b),$$

а величина $\Theta(f)$ — характеристика складності моделі. Функція \mathcal{C} називається функцією втрат. Якщо в якості складності моделі взяти гладкість її функції представлення, а в якості функції втрат взяти запропоновану Вапником [17] ϵ -нечутливу функцію:

$$\mathcal{C}_{\epsilon}(\delta) = |\delta|_{\epsilon} = \begin{cases} 0, & \text{для } |\delta| \leq \epsilon \\ |\delta| - \epsilon, & \text{для } |\delta| > \epsilon \end{cases},$$

то мінімізація структурного ризику

$$R_{reg}(f) = \frac{1}{N} \sum_{j=1}^N |y_j - \langle \mathbf{w}, \mathbf{x}_j \rangle_{\mathcal{X}} - b|_{\epsilon} + \frac{\lambda}{2} \|\mathbf{w}\|_{\mathcal{X}}^2$$

відповідає випуклій оптимізації (3.1-3.2) за умови, якщо

$$\beta = \frac{1}{N\lambda}.$$

3.1.2. Нелінійна регресія опорних векторів

Якщо ж припустити, що справжня залежність нелінійна і може бути описана лінійною в параметрах моделлю шляхом попереднього відображення простору вхідних змінних \mathcal{X} розмірності n в деякий багатовимірний характеристичний простір¹ (англ. feature space) \mathcal{F} розмірності q (як правило $n \ll q$),

¹В термінах методу потенціальних функцій М.Айзермана цей простір називають спрямляючим [181]

то задача нелінійної регресії опорних векторів у вхідному просторі зводиться до лінійної в характеристичному:

$$\mathcal{M}(\mathbf{x}, \mathbf{w}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle_{\mathcal{F}} + b, \quad \mathbf{w} \in \mathcal{F}, \quad b \in \mathbb{R}, \quad \Phi : \mathcal{X} \rightarrow \mathcal{F}. \quad (3.3)$$

Оптимізаційна задача при цьому приймає вигляд:

$$\frac{1}{2} \|\mathbf{w}\|_{\mathcal{F}}^2 + \beta \sum_{j=1}^N (\xi_j + \xi_j^*) \longrightarrow \min, \quad (3.4)$$

з обмеженнями:

$$\begin{cases} y_j - \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} - b \leq \epsilon + \xi_j \\ -y_j + \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} + b \leq \epsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0 \end{cases}. \quad (3.5)$$

Графічне зображення побудови регресії з використанням ϵ -нечутливої смуги представлено на рис. 3.1. Такий вигляд функції втрат забезпечує робастність моделей РОВ, яка полягає в невисокій чутливості до зміни значень входів за межами ϵ -смуги і в абсолютній нечутливості в межах ϵ -смуги.

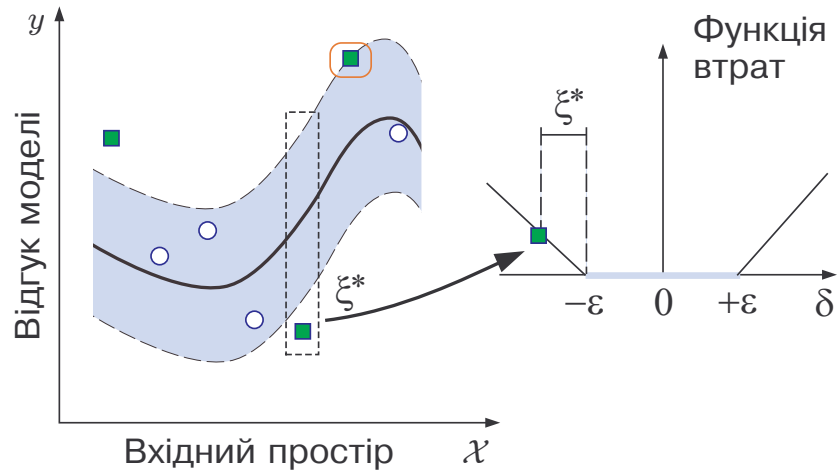


Рис. 3.1. ϵ -нечутлива смуга регресії опорних векторів.

Очевидно, що задача (3.4-3.5) є задачею знаходження умовного екстремуму і для її розв'язання зручно використати множники Лагранжа $\alpha_j, \gamma_j, \alpha_j^*, \gamma_j^*$

і записати функцію Лагранжа у вигляді:

$$L = \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle_{\mathcal{F}} + \beta \sum_{j=1}^N (\xi_j + \xi_j^*) + \sum_{j=1}^N \alpha_j (y_j - \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} - b - \epsilon - \xi_j) + \\ + \sum_{j=1}^N \alpha_j^* (-y_j + \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} + b - \epsilon - \xi_j^*) - \sum_{j=1}^N (\gamma_j \xi_j + \gamma_j^* \xi_j^*). \quad (3.6)$$

При чому $\alpha_j, \gamma_j, \alpha_j^*, \gamma_j^* \geq 0$. Оптимальність досягається в сідловій точці за умови рівності нулю частинних похідних:

$$\begin{cases} \partial_b L = \sum_{j=1}^N (\alpha_j^* - \alpha_j) = 0, \\ \partial_{\mathbf{w}} L = \mathbf{w} - \sum_{j=1}^N (\alpha_j - \alpha_j^*) \Phi(\mathbf{x}_j) = \mathbf{0}, \\ \partial_{\xi_j} L = \beta - \alpha_j - \gamma_j = 0 \\ \partial_{\xi_j^*} L = \beta - \alpha_j^* - \gamma_j^* = 0. \end{cases} \quad (3.7)$$

Підставляючи рівності (3.7) в лагранжиан (3.6) отримаємо цільову функцію двоїстої задачі максимізації у наступному вигляді:

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} - \\ - \epsilon \sum_{j=1}^N (\alpha_j + \alpha_j^*) + \sum_{j=1}^N y_j (\alpha_j - \alpha_j^*), \quad (3.8)$$

з обмеженнями:

$$\sum_{j=1}^N (\alpha_j - \alpha_j^*) = 0, \quad \alpha_j, \alpha_j^* \in [0; \beta].$$

Перехід до двоїстої задачі значно спрощує розв'язання задачі оптимізації особливо у випадках, коли розмірність характеристичного простору q більша за розмір вибірки N . Крім цього за умовами доповнюючої нежорсткості теореми Каруш-Куна-Таккера отримуємо:

$$\begin{cases} \alpha_j (y_j - \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} - b - \epsilon - \xi_j) = 0, \\ \alpha_j^* (-y_j + \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} + b - \epsilon - \xi_j^*) = 0, \\ (\beta - \alpha_j) \xi_j = 0, \\ (\beta - \alpha_j^*) \xi_j^* = 0. \end{cases}$$

Звідси можна зробити очевидні висновки:

1. Точки які лежать строго вище ϵ -нечутливої смуги мають $\alpha = \beta, \alpha^* = 0$.
2. Точки які лежать строго нижче ϵ -нечутливої смуги мають $\alpha = 0, \alpha^* = \beta$.
3. Точки які лежать строго в ϵ -нечутливій смузі мають $\alpha = 0, \alpha^* = 0$.
4. Не існує такої точки для якої обидва множники Лагранжа одночасно були б ненульові. Це означає, що завжди має місце рівність:

$$\alpha_j \alpha_j^* = 0, \quad j = 1, \dots, N.$$

Зміщення моделі b визначається з рівняння:

$$y_j - \langle \mathbf{w}, \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}} - b = \epsilon, \quad \text{при} \quad \alpha_j \in (0; \beta), \quad \xi_j = 0.$$

Звідси визначаються опорні вектори як такі, що мають один ненульовий множник Лагранжа, тобто або $\alpha_j > 0$ або $\alpha_j^* > 0$. Множина опорних векторів задається як:

$$SV = \{\mathbf{x}_i : |y_i - f(\mathbf{x}_i)| \geq \epsilon\}.$$

Оскільки з рівнянь (3.7) ми також отримуємо вираз для вектору параметрів:

$$\mathbf{w} = \sum_{j=1}^N (\alpha_j - \alpha_j^*) \Phi(\mathbf{x}_j), \quad (3.9)$$

то функція (3.3) запишеться у вигляді так званого розкладу опорних векторів (неопорні або ж пасивні вектори мають $\alpha_j = 0$ і $\alpha_j^* = 0$, що обнуляє відповідний доданок):

$$f(\mathbf{x}) = \sum_{j=1}^N (\alpha_j - \alpha_j^*) \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}) \rangle_{\mathcal{F}} + b. \quad (3.10)$$

Таким чином вектор параметрів \mathbf{w} а з ним і апроксимуюча функція $f(\mathbf{x})$ може бути повністю описані як лінійні комбінації елементів розрідженої множини навчальної вибірки (опорних векторів) представлені в характеристичному просторі. Це означає, що $f(\cdot) \in \text{span}\{\langle \Phi(\mathbf{x}_j), \Phi(\cdot) \rangle_{\mathcal{F}}\}_{j=1}^N$.

3.1.3. Зведення РОВ до задачі квадратичного програмування

Позначимо шуканий вектор множників Лагранжа \mathbf{m} такий, що:

$$m_j = \alpha_j, \quad m_{j+N} = \alpha_j^*, \quad j = 1, \dots, N.$$

Введемо матрицю \mathbf{G} таку, що

$$\mathbf{G} = \begin{pmatrix} \mathbf{D} & -\mathbf{D} \\ -\mathbf{D} & \mathbf{D} \end{pmatrix},$$

де матриця $\mathbf{D} = \{d_{ij}\}$, $d_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{F}}$. Тоді відповідна до (3.8) задача квадратичного програмування в матричному вигляді перепишеться:

$$\mathbf{m}_{opt} = \arg \min \frac{1}{2} \mathbf{m}^\top \mathbf{G} \mathbf{m} + \mathbf{c}^\top \mathbf{m}, \quad (3.11)$$

де вектор $\mathbf{c} = (\epsilon - y_1, \dots, \epsilon - y_N, \epsilon + y_1, \dots, \epsilon + y_N)^\top$. З обмеженнями:

$$\begin{cases} \mathbf{a}^\top \mathbf{m} = 0, & \mathbf{a} = (\underbrace{1, \dots, 1}_N, \underbrace{-1, \dots, -1}_N)^\top, \\ 0 \leq m_j \leq \beta, & j = 1, \dots, 2N. \end{cases}$$

Для вектору параметрів відповідно маємо:

$$\mathbf{w} = (\mathbf{Q}, -\mathbf{Q}) \mathbf{m}, \quad \mathbf{Q} = (\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_N)).$$

Задача (3.11) є класичною задачею опуклого квадратичного програмування поліноміальної складності [184], яку можна розв'язати методом Вульфа, методом активних обмежень [185, 186], методом L-BFGS (limited memory Broyden-Fletcher-Goldfarb-Shanno), який відноситься до класу квазі-ньютонівських методів. Широкий спектр методів розв'язання цієї задачі приведено в [114].

Розклад опорних векторів (3.10) в матричному вигляді²:

$$f(\mathbf{x}) = \Phi(\mathbf{x})^\top \mathbf{w} + b = \Phi(\mathbf{x})^\top \left(\sum_{\mathbf{x}_i \in SV^+} m_i \Phi(\mathbf{x}_i) - \sum_{\mathbf{x}_j \in SV^-} m_{j+N} \Phi(\mathbf{x}_j) \right) + b.$$

²В узагальненій теорії машини опорних векторів, де характеристичний простір це гільбертів простір з відтворюючими ядрами і скалярним добутком $\langle \cdot, \cdot \rangle_{\mathcal{F}}$, модель має вигляд: $f(\mathbf{x}) = \langle \mathbf{w}, \Phi(\mathbf{x}) \rangle_{\mathcal{F}} + b$. Надалі вважаємо, що скалярний добуток задано як: $\langle \mathbf{x}, \mathbf{z} \rangle_{\mathcal{F}} = \sum_i x_i \bar{z}_i$.

де множина опорних векторів $SV = SV^+ \cup SV^-$, $\emptyset = SV^+ \cap SV^-$:

$$SV^+ = \{\mathbf{x}_i : y_i - f(\mathbf{x}_i) \geq \epsilon\}, \quad SV^- = \{\mathbf{x}_i : y_i - f(\mathbf{x}_i) \leq -\epsilon\}.$$

SV^+ — опорні вектори, що знаходяться вище смуги нечутливості або на її верхній межі, SV^- — опорні вектори, що знаходяться нижче смуги нечутливості або на її нижній межі. Таким чином з одного боку кількість параметрів моделі це розмірність характеристичного простору \mathcal{F} плюс зміщення моделі b , тобто $q + 1$, а з іншого боку за рахунок розкладу опорних векторів як представлення розрідженої вхідної вибірки в характеристичному просторі: $N_{sv} + 1$, де N_{sv} — кількість опорних векторів. Як правило число $N_{eff} = N_{sv} + 1$ набагато менше ніж $q + 1$ і його можна інтерпретувати як кількість ефективних параметрів моделі.

3.2. Розвиток і застосування механізму опорних векторів

Якщо в якості характеристичного простору регресії опорних векторів взяти гільбертовий простір з відтворюючими ядрами (ГПВЯ), то алгоритмічна складність обчислень значно зменшується, що показано зокрема в роботі [187]. Зазначимо, що В-сплайни непарного порядку є допустимими відтворюючими ядрами [22], а значить, згідно із зауваженням 2.3, і базисні поліноми Бернштейна непарного порядку допустимі. Проте надалі в роботі характеристичний простір не обмежується до ГПВЯ.

Для автоматичного вибору параметру нечутливості ϵ в роботі [27] запропонований алгоритм ν -регресії опорних векторів (ν -РОВ). На основі ν -РОВ М.Лоу і Дж.Квок [38] започаткували теорію ймовірнісного підтвердження значень гіперпараметрів β та ϵ , яка буде розвинена автором в підрозділі 3.3.2.

Дж.Плат розробив алгоритм послідовної мінімальної оптимізації [188], який за рахунок розбиття задачі квадратичного програмування на мінімально можливі підзадачі, які легко розв'язуються аналітично, дозволяє навчати МОВ набагато швидше.

При ідентифікації часових рядів з сильними шумами з метою зменшення зміщення моделей опорних векторів в роботі [189] запропонований ітеративний алгоритм зменшення кореляції залишків. Інший підхід для сильно зашумлених процесів описаний в [190]. В роботі [191] механізм опорних векторів застосований до прихованих марковських моделей, заснованих на теорії прихованих дискримінантів марковських ланцюгів. В [192] приведено порівняння механізму опорних векторів з класифікаторами на основі радіальних базисних функцій.

Цікаве практичне застосування можна знайти в роботі [193], де нелінійна РОВ застосовується в області дослідження магнітно-резонансної інтраскопії для відтворення функціонування мозку в залежності від рівня кисню в крові без хірургічного втручання. В роботі [194] метод опорних векторів застосовується з метою побудови прогнозів на фондовій біржі, а в [195] пропонується метод прогнозування футбольних ігор на базі машини розділяючої гіперплощини як іноді ще називають машину опорних векторів.

3.3. Індуктивний метод побудови байєсівської моделі регресії опорних векторів у формі Бернштейна. Алгоритм ПРИАМ

3.3.1. Байєсівська регресія опорних векторів

Для знаходження і обґрунтування вибору параметрів β та ϵ регресії опорних векторів часто застосовують так званий байєсівський підхід [39, 40, 196]. Байєсівська регресія опорних векторів (БРОВ) припускає, що ϵ -нечутлива функція втрат відповідає наступній моделі шумів [38]:

$$P(\delta_i|f, \mathcal{M}) = \frac{\beta}{2(1 + \epsilon\beta)} \exp(-\beta|\delta_i|_\epsilon). \quad (3.12)$$

Дисперсія шуму визначається гіперпараметром β . Таким чином функція правдоподібності має вигляд:

$$P(\mathcal{D}|f, \mathcal{M}) = \left(\frac{\beta}{2(1 + \epsilon\beta)} \right)^N \cdot \exp \left(-\beta \sum_{i=1}^N |\delta_i|_\epsilon \right).$$

Апріорний розподіл вектору параметрів визначимо як багатовимірний гаусівський розподіл з нульовим середнім і одиничною дисперсією:

$$P(\mathbf{w}|\mathcal{M}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{I}) = \sqrt{\left(\frac{1}{2\pi}\right)^q} \cdot \exp(-R_{\mathbf{w}}(\mathbf{w})), \quad R_{\mathbf{w}}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^\top \mathbf{w}.$$

Найбільш ймовірне значення вектору параметрів \mathbf{w}_{mp} і зміщення b_{mp} , які визначають функцію f_{mp} , знаходяться з умови максимуму апостеріорного умовного розподілу ймовірності $P(f|\mathcal{D}, \mathcal{M})$, який визначається як:

$$P(f|\mathcal{D}, \mathcal{M}) = \frac{P(\mathcal{D}|f, \mathcal{M})P(\mathbf{w}|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})}.$$

В логарифмічному масштабі:

$$\begin{aligned} f_{mp} &= \arg \min_{\mathbf{w}} [-\log P(\mathcal{D}|f, \mathcal{M}) - \log P(\mathbf{w}|\mathcal{M})] = \\ &= \arg \min_{\mathbf{w}} -N \ln \frac{\beta}{2(1 + \epsilon\beta)} + \beta \sum_{i=1}^N |y_i - f(\mathbf{x}_i)|_\epsilon - \frac{q}{2} \ln \frac{1}{2\pi} + \frac{1}{2}\mathbf{w}^\top \mathbf{w}. \end{aligned} \quad (3.13)$$

Слід зазначити, що в цьому байєсівському висновку значення гіперпараметрів β і ϵ (ширини смуги нечутливості) вважаються рівноймовірними. Очевидно, що така задача зводиться знову ж таки до (3.4-3.5). Таким чином регуляризований ризик складає:

$$R_{reg}(\mathbf{w}) = \beta N R_{emp}(\mathbf{w}) + R_{\mathbf{w}}(\mathbf{w}) = \beta \sum_{i=1}^N |y_i - f(\mathbf{x}_i)|_\epsilon + \frac{1}{2}\mathbf{w}^\top \mathbf{w}.$$

Для знаходження оптимальних значень гіперпараметрів β , ϵ можна застосовувати ітеративну процедуру Маккея описану в 1.2.3. Відповідно, після знаходження \mathbf{w}_{mp} і b_{mp} , мінімізується зворотній логарифм підтвердження гіперпараметрів:

$$-\ln P(\mathcal{D}|\mathcal{M}) \approx -N \ln \frac{\beta}{2(1 + \epsilon\beta)} + \frac{1}{2} \ln \det \mathbf{H}_{mp} + R_{reg}(\mathbf{w}_{mp}), \quad (3.14)$$

де \mathbf{H}_{mp} — гессіан регуляризованого ризику R_{reg} в точці \mathbf{w}_{mp} . Оскільки в матричному вигляді представлення

$$f(\mathbf{x}) = \Phi(\mathbf{x})^\top \mathbf{w} + b,$$

для вибірки об'ємом N легко перетворюється на лінійну в параметрах модель $\mathbf{f} = \mathbf{A}\mathbf{w} + b\mathbf{I}_N$, де $\mathbf{A} = (\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_N))^\top$ то матриця Гессе регуляризованого ризику по відношенню до вектора параметрів \mathbf{w} визначається як

$$\mathbf{H} = \nabla_{\mathbf{w}}^2 R_{reg}(\mathbf{w}) = \mathbf{I}_q,$$

в усіх точках крім множини критичних (маргінальних) точок:

$$CP = \{\mathbf{w} \in \mathcal{F} : \exists i : |y_i - f(\mathbf{x}_i)| = \epsilon\}.$$

Іншими словами, маючи найбільш ймовірний вектор параметрів \mathbf{w}_{mp} , всі точки, які лежать на верхній та нижній межах ϵ -смуги є критичними для обчислення гессіану. Це слідує з того, що $\nabla_{\mathbf{w}}^2 R_{emp}(\mathbf{w}_{mp}) \equiv \mathbf{0}$ всюди крім точок, що задовольняють $|y_i - f(\mathbf{x}_i)| = \epsilon$, в цих точках функція ризику не-диференційовна. Для того, щоб уникнути цієї проблеми в [38] запропоновано наблизити ϵ -нечутливу функцію втрат наступною гладкою функцією:

$$\mathcal{C}_\eta(u) = \zeta_\eta(u - \epsilon) + \zeta_\eta(-u - \epsilon),$$

де $\zeta_\eta(u) = \frac{1}{\eta} \log(1 + \exp(\eta u))$, $\eta > 0$. В [40] вводиться м'яка функція втрат вигляду:

$$\mathcal{C}_\mu(u) = \begin{cases} -u - \epsilon, & \text{якщо } u \in (-\infty; -(1 + \mu)\epsilon) \\ \frac{(u + (1 - \mu)\epsilon)^2}{4\mu\epsilon}, & \text{якщо } u \in [-(1 + \mu)\epsilon; -(1 - \mu)\epsilon] \\ 0, & \text{якщо } u \in (-(1 - \mu)\epsilon; (1 - \mu)\epsilon) \\ \frac{(u - (1 - \mu)\epsilon)^2}{4\mu\epsilon}, & \text{якщо } u \in [(1 - \mu)\epsilon; (1 + \mu)\epsilon] \\ u - \epsilon, & \text{якщо } u \in ((1 + \mu)\epsilon; +\infty) \end{cases},$$

де параметр $\mu \in (0; 1]$. При $\mu \rightarrow 0$ м'яка функція втрат збігається до ϵ -нечутливої функції. Але такі підходи призводять до складних інтегралів, які не обчислюються в аналітичному вигляді, і призводять до порушення робастності моделей РОВ.

3.3.2. Критерій байєсівського підтвердження адекватності моделі

З метою спрощення аналітичних обчислень і забезпечення робастності моделей РОВ автором запропоновано застосувати локально згладжену в нескінченно малому околі критичних точок функцію втрат [87, 86].

Лема 3.1. Для як завгодно малих околів $W_\eta(-\epsilon) = (-\epsilon - \eta; -\epsilon + \eta)$, $W_\eta(\epsilon) = (\epsilon - \eta; \epsilon + \eta)$ точок $-\epsilon$, ϵ існує така C^2 гладка функція \mathcal{S}_ϵ , яка рівна ϵ -нечутливій функції втрат Вапника \mathcal{C}_ϵ за межами цих околів.

Доведення. Достатньо розглянути тільки невід'ємну дійсну піввісь (оскільки функція втрат симетрична). Визначимо всюди гладку першу похідну $\mathcal{S}'_\epsilon(u)$, у вигляді:

$$\mathcal{S}'_\epsilon(u) = \begin{cases} 0, & u \in [0; \epsilon - \eta] \\ l(u), & u \in (\epsilon - \eta; \epsilon + \eta) \\ 1, & u \in [\epsilon + \eta; \infty) \end{cases}, \quad (3.15)$$

де $l(u)$ — деяка C^1 гладка функція в околі $W_\eta(\epsilon)$, тобто на інтервалі $(\epsilon - \eta; \epsilon + \eta)$, яка задовольняє умовам:

$$\int_{\epsilon - \eta}^{\epsilon + \eta} l(u) du = \eta, \quad \lim_{u \rightarrow \epsilon - \eta} l(u) = 0, \quad \lim_{u \rightarrow \epsilon + \eta} l(u) = 1, \quad l'_+(\epsilon - \eta) = l'_-(\epsilon + \eta) = 0,$$

де l'_+ , l'_- — похідні справа та зліва відповідно. Очевидно, що таким чином побудована перша похідна $\mathcal{S}'_\epsilon(u)$ забезпечує C^2 гладку функцію втрат \mathcal{S}_ϵ , яка рівна \mathcal{C}_ϵ за межами околу $W_\eta(\epsilon)$ на додатній півосі. Тож для доведення посилення теореми знайдемо функцію $l(u)$ в класі логістичних функцій, тобто функцій вигляду:

$$l(\pi(u)) = \frac{e^{\pi(u)}}{1 + e^{\pi(u)}}. \quad (3.16)$$

Очевидно, що для такого класу функцій виконуються умови:

$$\lim_{\pi \rightarrow -\infty} l(\pi) = 0, \quad \lim_{\pi \rightarrow +\infty} l(\pi) = 1.$$

Застосуємо дробово-лінійне перетворення вигляду

$$\pi(u) = \frac{u - \epsilon}{\eta^2 - (u - \epsilon)^2}$$

для того, щоб конформно відобразити інтервал $(\epsilon - \eta; \epsilon + \eta)$ на дійсну вісь $(-\infty; +\infty)$. Легко переконатися, що відображення $\pi(u)$ переводить точки $\epsilon - \eta$, ϵ , $\epsilon + \eta$ відповідно в точки $-\infty$, 0 , $+\infty$. Це дає змогу записати логістичну функцію (3.16) від змінної u у вигляді:

$$l(\pi(u)) = \frac{e^{\pi(u)}}{1 + e^{\pi(u)}} = \frac{1}{1 + e^{-\pi(u)}} = \left(1 + \exp\left(\frac{u - \epsilon}{(\eta - \epsilon)^2 - \eta^2}\right)\right)^{-1}. \quad (3.17)$$

Зображення першої похідної $S'_\epsilon(u)$ з логістичною функцією (3.17) в η -околі точки ϵ приведено на рис. 3.2. В силу антисиметричності цієї функції відносно

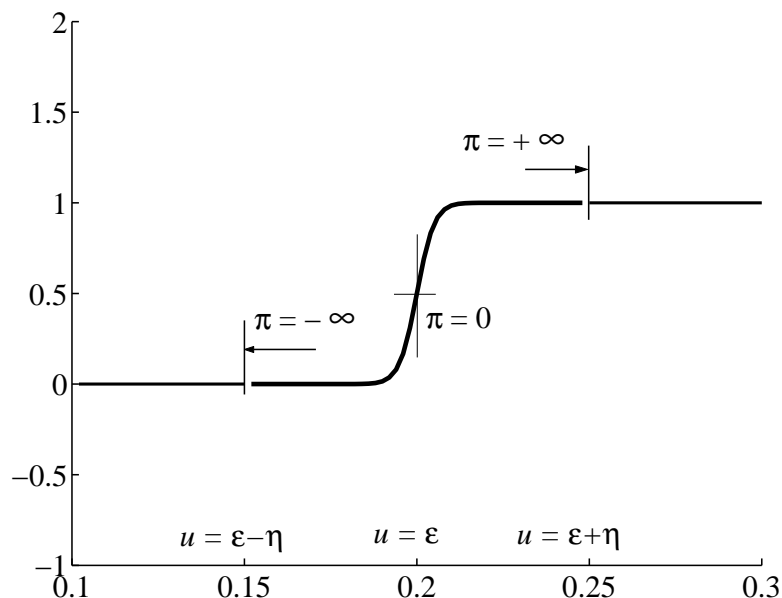


Рис. 3.2. Графік першої похідної для $\epsilon = 0.2$, $\eta = 0.05$.

точки $(\epsilon; 0.5)$, оцінивши площу під графіком в η -околі, легко бачити, що

$$\int_{\epsilon-\eta}^{\epsilon+\eta} l(u) du = \int_{\epsilon}^{\epsilon+\eta} 1 \cdot du = \eta.$$

Залишилось перевірити значення похідних на кінцях η -околу. Похідна логістичної функції:

$$l'(u) = \frac{e^{-\pi(u)}}{(1 + e^{-\pi(u)})^2} \cdot \pi'_u(u), \quad \pi'_u(u) = \frac{\eta^2 + (u - \epsilon)^2}{(\eta^2 - (u - \epsilon)^2)^2}.$$

Значення похідних на кінцях відповідно становлять:

$$l'_+(\epsilon - \eta) = \lim_{u \downarrow \epsilon - \eta} l'(u) = 0, \quad l'_-(\epsilon + \eta) = \lim_{u \uparrow \epsilon + \eta} l'(u) = 0.$$

Таким чином, якщо вибрати першу похідну $S'_\epsilon(u)$ з логістичною функцією $l(u)$ вигляду (3.17), тоді забезпечується існування такої функції втрат $\mathcal{S}_\epsilon(u)$, яка співпадає з $\mathcal{C}_\epsilon(u)$ поза околom W_η і має вигляд:

$$\mathcal{S}_\epsilon(u) = \int_{-\infty}^u S'_\epsilon(t) dt.$$

□

Оскільки друга похідна функції втрат \mathcal{S}_ϵ'' в точці ϵ рівна:

$$\mathcal{S}_\epsilon''(\epsilon) = \frac{e^0}{(1+e^0)^2} \cdot \frac{\eta^2}{\eta^4} = \frac{1}{4\eta^2},$$

то для достатньо малих значень η (відповідні η -околи покривають тільки маргінальні елементи вибірки) гессіан для локально згладженої функції втрат рівний:

$$\mathbf{H} = \mathbf{I}_q + \frac{\beta}{4\eta^2} \mathbf{I}_M,$$

де $\mathbf{I}_M = \text{diag}\{\mathbf{d}\}$ — діагональна матриця з $d_{ii} = 1$ для m маргінальних елементів вибірки $|y_i - f(\mathbf{x}_i)| = \epsilon$ і $d_{ii} = 0$ для решти елементів. Тоді (3.14) приймає вигляд:

$$-\ln P(\mathcal{D}|\mathcal{M}) \approx -N \ln \frac{\beta}{2(1+\epsilon\beta)} + \frac{m}{2} \ln \left(1 + \frac{\beta}{4\eta^2} \right) + R_{reg}(\mathbf{w}_{mp}).$$

Покажемо яким чином можна ще більше спростити зворотній логарифм підтвердження гіперпараметрів.

Теорема 3.2. Для будь-якої ϵ -нечутливої смуги регресії опорних векторів існує як завгодно мале відхилення $\Delta\epsilon$ таке, що для ϵ_1 -нечутливої смуги ($\epsilon_1 = \epsilon - \Delta\epsilon$) гессіан регуляризованого ризику $\mathbf{H}_{mp} \equiv \mathbf{I}_q$ в усіх точках і при цьому вектори зберігають властивість опорності.

Доведення. Позначимо множину похибок наближення регресії опорних векторів як $\{\delta_i, \delta_i = |y_i - f(\mathbf{x}_i)|\}_{i=1}^N$. Якщо $\nexists i : \delta_i = \epsilon$, тоді покладемо $\Delta\epsilon = 0$. Якщо існує хоча б одне $i : \delta_i = \epsilon$, тоді виберемо таку підмножину похибок

$\{\delta_j\}_j$, для яких $\delta_j < \epsilon$. Очевидно, що існує $\delta_{max} = \max\{\delta_j : \delta_j < \epsilon\} < \epsilon$. Виберемо як завгодно мале значення $\Delta\epsilon$ з інтервалу $(0; \epsilon - \delta_{max})$.

Таким чином вибране відхилення $\Delta\epsilon$ забезпечує те, що не існує точок вибірки, які лежать на верхній чи нижній межах ϵ_1 -нечутливої смуги (рис. 3.3). Оскільки вектори які лежать на верхній чи нижній межах ϵ -

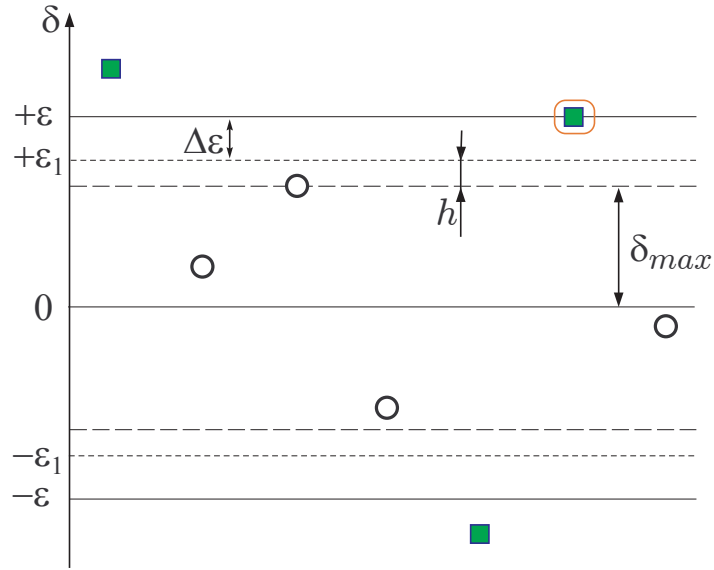


Рис. 3.3. Вибір відхилення $\Delta\epsilon$ для заданої вибірки. Опорні вектори зображені зафарбованими квадратами.

нечутливої смуги є опорними, то при звуженні смуги на $\Delta\epsilon$ вони не втрачають властивість опорності, і жодний неопорний вектор не стає опорним. Очевидно, що

$$\min_i |\delta_i - \epsilon_1| = h > 0.$$

Виберемо η таке, що $0 < \eta < h$, тоді за лемою 3.1 існує функція \mathcal{S}_{ϵ_1} , яка співпадає з ϵ_1 -нечутливою функцією втрат \mathcal{C}_{ϵ_1} за межами околів $W_\eta(\epsilon_1)$. Оскільки не існує точок вибірки, які лежать в околах $W_\eta(\epsilon_1)$, а за межами цих околів друга похідна функції втрат дорівнює нулю, то можна зробити висновок про те, що $\nabla_{\mathbf{w}}^2 R_{emp}(\mathbf{w}) \equiv \mathbf{0}$. Це фактично доводить, що гессіан регуляризованого ризику можна вважати в усіх точках рівним

$$\mathbf{H} = \mathbf{I}_q,$$

оскільки для обчислювальних методів можна знехтувати як завгодно малим зменшенням ширини смуги нечутливості (достатньо вибрати величину $\Delta\epsilon$ менше ніж точність операцій технічного обчислювального засобу, тобто менше ніж дозволяє розрядність мантиси). \square

Враховуючи вище доведену теорему, наближення (3.14) можна переписати у вигляді:

$$-\ln P(\mathcal{D}|\mathcal{M}) \approx \text{КБП}(\mathcal{M}, \epsilon, \beta) = -N \ln \frac{\beta}{2(1 + \epsilon\beta)} + R_{reg}(\mathbf{w}_{mp}), \quad (3.18)$$

Цей критерій будемо називати критерієм байєсівського підтвердження (КБП) для РОВ.

Взявши похідну в (3.18) відносно гіперпараметру β і прирівнявши її нулю, отримаємо:

$$-\frac{N}{\beta} + \frac{N\epsilon}{1 + \epsilon\beta} + NR_{emp}(\mathbf{w}_{mp}) = 0.$$

Екстремальність за гіперпараметром β дає першу умову оптимальності побудованої регресії опорних векторів:

$$-\frac{1}{\beta(1 + \epsilon\beta)} + R_{emp}(\mathbf{w}_{mp}) = 0. \quad (3.19)$$

Другу умову оптимальності можна отримати з умови екстремальності (3.18) за гіперпараметром ϵ .

Твердження 3.3. Швидкість зменшення емпіричного ризику при збільшенні ширини смуги нечутливості прямо пропорційна кількості опорних векторів.

Доведення. Спочатку знайдемо похідну функції втрат Валника по ϵ :

$$(|\delta|_\epsilon)'_\epsilon = \lim_{\Delta\epsilon \rightarrow 0} \frac{|\delta|_{\epsilon+\Delta\epsilon} - |\delta|_\epsilon}{\Delta\epsilon} = -\frac{|\delta|_\epsilon}{|\delta| - \epsilon}.$$

Тоді

$$(R_{emp})'_\epsilon = \left(\frac{1}{N} \sum_{j=1}^N |\delta_j|_\epsilon \right)'_\epsilon = -\frac{N_{sv}}{N}.$$

де N_{sv} — кількість опорних векторів. \square

Враховуючи результат леми 3.3, взявши похідну в (3.18) відносно гіперпараметру ϵ і прирівнявши її нулю, отримаємо другу умову оптимальності:

$$\frac{N\beta}{1 + \epsilon\beta} - \beta N_{sv} = 0 \Leftrightarrow N_{sv} = \frac{N}{1 + \epsilon\beta}. \quad (3.20)$$

Як буде показано далі, ця умова має змістовну інтерпретацію, а саме — вона відображає апіорну ймовірність опорності вектору.

Умови (3.19-3.20) оптимальності РОВ приводять в загальному випадку до системи нелінійних алгебраїчних рівнянь:

$$\begin{cases} -1 + \beta(1 + \epsilon\beta) \cdot R_{emp}(\epsilon, \beta) = 0, \\ -N + (1 + \epsilon\beta) \cdot N_{sv}(\epsilon, \beta) = 0. \end{cases} \quad (3.21)$$

Цю систему можна розв'язувати методами простої ітерації, Ньютона, Левенберга-Марквардта³. В даному дисертаційному дослідженні пропонується використовувати рефлексивний метод Ньютона [197]. Також для знаходження оптимальних значень гіперпараметрів можна застосовувати методи нелінійної оптимізації безпосередньо до (3.18). Ефективність того чи іншого підходу залежить від його реалізації. Загальна схема алгоритму оптимізації за гіперпараметрами має наступний вигляд.

Задані: дані спостережень \mathcal{D} , модель \mathcal{M} , початкові значення гіперпараметрів $\epsilon^{(0)}$, $\beta^{(0)}$, мінімізатор \mathfrak{S} .

Результат: оптимальні значення гіперпараметрів.

Початок

ітератор $t \leftarrow 0$;

повторювати

обчислення МАП оцінки (3.13) з гіперпараметрами $\epsilon^{(t)}$, $\beta^{(t)}$;

обчислення КБП(\mathcal{M} , $\epsilon^{(t)}$, $\beta^{(t)}$) за формулою (3.18);

обчислення нових значень $\epsilon^{(t+1)}$, $\beta^{(t+1)}$, використовуючи \mathfrak{S} метод;

$t \leftarrow t + 1$;

поки не досягли необхідного рівня збіжності для методу \mathfrak{S} .

³http://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm

Інтуїтивний метод адаптації параметрів регресії опорних векторів

Оскільки (3.19) є фактично умовою оптимальності регресії опорних векторів за гіперпараметрами ϵ та β , то процес адаптації зводиться до спроб задовольнити цю умову.

Якщо припустити, що модель незміщена і справжній шум дійсно описується (3.12), тоді ймовірність того, що конкретному \mathbf{x}_i буде відповідати значення виходу y_i за межами ϵ -смуги нечутливості, тобто вектор \mathbf{x}_i є опорним, рівна:

$$p_{sv} = \int_{\mathbb{R} \setminus (-\epsilon, \epsilon)} P(y_i | \mathbf{x}_i, f, \epsilon, \beta, \mathcal{M}) d\delta = 1 - \int_{-\epsilon}^{\epsilon} \frac{\beta}{2(1 + \epsilon\beta)} \exp(-\beta \cdot 0) d\delta = \frac{1}{1 + \epsilon\beta}.$$

де відхилення $\delta = y_i - f(\mathbf{x}_i)$.

Припустимо, що на основі наперед заданих значень гіперпараметрів ϵ_0 , β_0 побудована регресія з N_{sv} опорними векторами і емпіричним ризиком R_{emp} . Очевидно, що апіорна ймовірність опорності вектору (очікувана ймовірність) рівна:

$$P(H) = \frac{1}{1 + \epsilon_0\beta_0}.$$

При цьому апостеріорна ймовірність (реально отримана) рівна:

$$P(H | \epsilon = \epsilon_0, \beta = \beta_0) = \frac{N_{sv}}{N}.$$

Інтуїтивно зрозуміло що, якщо апостеріорна ймовірність більша (занадто багато опорних векторів), то ширина смуги нечутливості є менша за потрібну, і тому її вигідно збільшити. Це правило можна записати як:

$$\epsilon_1 = \epsilon_0 \cdot \frac{P(H | \epsilon = \epsilon_0, \beta = \beta_0)}{P(H)} = \epsilon_0 \cdot \frac{N_{sv}}{N} (1 + \epsilon_0\beta_0),$$

де ϵ_1 — нове рекомендоване значення. Це означає що, якщо ми отримали вдвічі більше опорних векторів ніж розраховували, то необхідно збільшити ширину смуги нечутливості також вдвічі. Нове значення β_1 можна отримати з (3.19) за правилом:

$$\beta_1 = \frac{1}{1 + \epsilon_0\beta_0} \cdot \frac{1}{R_{emp}}.$$

Практичні дослідження підтвердили ефективність такого інтуїтивного висновку, хоча строго він не доведений.

Аналіз критерію байєсівського підтвердження

При виведенні КБП був використаний метод Лапласа. Зміст цього методу полягає у наближенні апостеріорного розподілу функцій $P(f|\mathcal{D}, \mathcal{M})$ нормальним розподілом, суміщуючи моди цих розподілів. В цьому випадку можна аналітично обчислити інтеграл підтвердження. Проте, як показано в роботі [34], метод Лапласа систематично переоцінює підтвердження і не може бути використаний для точного його підрахунку. Для точного підрахунку застосовують метод розповсюдження сподівань [126], який наближає апостеріорний розподіл функцій нормальним розподілом, суміщуючи перші два моменти цих розподілів. Також для отримання асимптотично точної оцінки інтегралу застосовують метод Монте-Карло [125]. Тож основною перевагою методу Лапласа в порівнянні з вище згаданими методами є його швидкість.

Більш глибокий аналіз методу Лапласа показує, що цей метод систематично переоцінює підтвердження для всіх моделей і може бути використаний для швидкого порівняння рівня адекватності моделей [85]. На рис. 3.4 наведені оцінки залежності $-\ln P(\mathcal{D}|\mathcal{M})$ від різних моделей (модель з гіперпараметром), які отримані наближеннями Монте-Карло та Лапласа. Таким чином КБП можна використовувати як оцінку адекватності моделей при їх порівнянні. Наприклад, для порівняння адекватності моделей \mathcal{M}_1 та \mathcal{M}_2 (близьких в просторі \mathcal{H}) достатньо обчислити КБП і вибрати модель з меншим значенням КБП, і відповідно з більшим значенням (зміщеної) оцінки підтвердження. Це дає можливість застосовувати методи нелінійної оптимізації для пошуку оптимальної моделі \mathcal{M}^0 .

Для загальності опишемо коротко наближення підтвердження за мето-

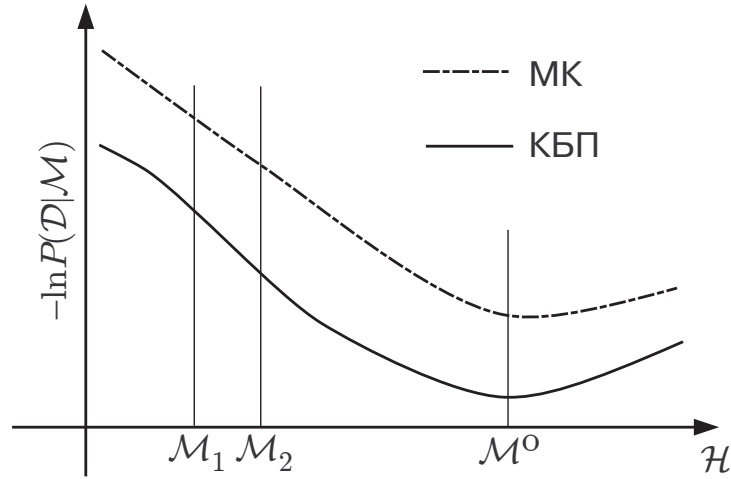


Рис. 3.4. Порівняння оцінок підтвердження методом Монте-Карло (МК) та методом Лапласа (КБП).

дом Монте-Карло. Як добре відомо має місце наступна оцінка:

$$\int h(\phi)P(\phi)d\phi \approx \frac{1}{T} \sum_{i=1}^T h(\phi_i),$$

де значення ϕ_i згенеровані відповідно до розподілу $P(\phi)$. Тож для оцінки підтвердження можна використовувати наближення:

$$P(\mathcal{D}|\mathcal{M}) = \int_{\mathcal{F}} P(\mathcal{D}|\mathbf{w}, \mathcal{M})P(\mathbf{w}|\mathcal{M})d\mathbf{w} \approx \frac{1}{T} \sum_{i=1}^T P(\mathcal{D}|\mathbf{w}_i, \mathcal{M}),$$

де \mathbf{w}_i згенеровані відповідно до апіорного розподілу $P(\mathbf{w}|\mathcal{M})$, який в нашому випадку $\mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{I})$. Звідси отримаємо оцінку зворотного логарифму підтвердження у вигляді:

$$-\ln P(\mathcal{D}|\mathcal{M}) = -N \ln \frac{\beta}{2(1 + \epsilon\beta)} - \ln \left(\frac{1}{T} \sum_{i=1}^T \exp(-\beta N R_{emp}(\mathbf{w}_i)) \right).$$

Достатньо точна оцінка може бути отримана тільки при відносно великих значеннях $T = 10^5$. Для зменшення кількості вибірових значень і побудови ефективної траєкторії дискретизації застосовують зокрема модель Гіббса методу Монте-Карло для марковських ланцюгів. Застосування такого підходу можна знайти зокрема в роботі [198].

3.3.3. Схема алгоритму ПРИАМ. Характеристичний простір поліноміальних функцій Без'є-Бернштейна.

Для забезпечення структурованості моделей в якості характеристичного простору доцільно застосувати простір поліноміальних функцій Без'є-Бернштейна. Введено поняття конфігурації характеристичного простору поліноміальних функцій Без'є-Бернштейна у вигляді верхньотрикутної матриці \mathbf{C} розмірності $n \times n$, де кожен елемент $c_{i \geq j} \geq 0$ відображає ступінь впливу фактору $x_{i=j}$ або пари факторів x_i, x_j на вихідну змінну y і визначає порядок базисних поліномів Бернштейна для відповідної поліноміальної функції. Характеристичне відображення у цьому випадку матиме вигляд: $\Phi(\mathbf{C}) : \mathbf{x} \mapsto (\dots, \phi_j^{c_k}(x^k), \dots, \phi_{irt}^{c_{pq}}(x^p, x^q), \dots)^\top$. Це відображення індукує відповідний простір моделей у формі Бернштейна, який узагальнює розклад (2.4):

$$\mathcal{M}(\mathbf{w}, b, \mathbf{x}) = \mathcal{H}(\mathbf{C}, \mathbf{w}, b, \mathbf{x}) = b + \sum_{k=1}^n B_k^{c_k}(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^n B_{pq}^{c_{pq}}(x^p, x^q), \quad (3.22)$$

Для загальності міркувань покладемо $B^0(\cdot) \equiv 0$, що фактично означає відсутність впливу відповідного фактору або пари факторів на вихід.

Основним результатом третього розділу є індуктивний метод побудови байєсівської моделі РОВ у формі Бернштейна, який визначає порядок перебору моделей в просторі \mathcal{H} . Схема алгоритму ПРИАМ⁴, який реалізує даний метод має вигляд [92, 199]:

Поліноміальної Регресії опорних векторів Індуктивний Алгоритм Моделювання

Задані: дані спостережень \mathcal{D} , рівень збіжності $\mu > 0$, початкова модель $\mathcal{M}^{(0)} = \mathcal{H}(\mathbf{C}^{(0)})$, яка відповідає апіорним сподіванням.

Результат: субоптимальна модель \mathcal{M}_{opt} .

Алгоритм починає роботу з

⁴ПРИАМ — Поліноміальної Регресії опорних векторів Індуктивний Алгоритм Моделювання. Шаблон індуктивності взято з алгоритму ASMOD [55, 56]. ASMOD — це адаптивний сплайновий алгоритм моделювання, який має за основу принцип мінімізації структурного ризику.

нормування вхідного та вихідного просторів;

ітератор $t \leftarrow 0$;

повторювати

$$\mathcal{M}_{\text{opt}} \leftarrow \mathcal{M}^{(t)};$$

генерація множини моделей кандидатів: $\left\{ \mathcal{M}_{ij}^{(t+1)} = \mathcal{H} \left(\mathbf{C}_{ij}^{(t+1)} \right) \right\}_{ij}$,

де $\mathbf{C}_{ij}^{(t+1)} = \mathbf{C}^{(t)} \pm \mathbf{1}_{ij}$, $1 \leq i \leq j \leq n$, $\mathbf{1}_{ij}$ — нульова матриця з одиницею в рядку i в стовпчику j , що відповідає мінімальній зміні конфігурації простору;

для кожної моделі кандидата $\mathcal{M}_{ij}^{(t+1)}$ виконати

обчислення КБП $\left(\mathcal{M}_{ij}^{(t+1)} \right)$, мінімізація (3.18) за гіперпараметрами;

кінець перебору моделей кандидатів;

вибір моделі з найменшим значенням КБП:

$$\mathcal{M}^{(t+1)} = \arg \left(\text{КБП}^{(t+1)} = \min \text{КБП} \left(\mathcal{M}_{ij}^{(t+1)} \right) \right);$$

критерій зупинки $\leftarrow \text{КБП}^{(t+1)} + \mu > \text{КБП}^{(t)}$;

$$t \leftarrow t + 1;$$

поки не виконається критерій зупинки;

кінець роботи алгоритму.

Визначимо розмірність задачі навчання через розмірність простору входу n і об'єм вибірки даних спостережень N . Тоді складність алгоритму ПРИАМ складається з витрат на перебір моделей кандидатів і витрат на розв'язок задачі квадратичного програмування методом активних обмежень і становить $\mathcal{O}(n^2 N^3)$. Збіжність ПРИАМ забезпечується глобальною і квадратичною збіжністю рефлексивного методу Ньютона для мінімізації КБП за гіперпараметрами а також, як легко показати в силу оцінки знизу $\text{КБП} > -N \ln(\beta_{\max}/2)$ для $\beta_{\max} < \infty$, лінійною збіжністю пошуку субоптимальної моделі.

Збалансованість нейронечіткої моделі РОВ у формі Бернштейна

Відзначимо ще один важливий результат.

Твердження 3.4. Нехай за допомогою регресії опорних векторів отримана нейронечітка модель у формі Бернштейна

$$f(\mathbf{x}) = b + \sum_{k=1}^n \sum_{j=0}^{c_k} w_j^k \phi_j^{c_k}(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^n \sum_{i+r+t=c_{pq}} w_{irt}^{pq} \phi_{irt}^{c_{pq}}(x^p, x^q).$$

Тоді сума коефіцієнтів при базисних поліномах Бернштейна $\phi_j^{c_k}$ і $\phi_{irt}^{c_{pq}}$ дорівнює нулю:

$$\sum_{j=0}^{c_k} w_j^k = 0, \quad \sum_{i+r+t=c_{pq}} w_{irt}^{pq} = 0.$$

Доведення твердження 3.4. Коефіцієнти регресії опорних векторів відповідно до (3.9) дорівнюють

$$w_j^k = \sum_{e=1}^N (\alpha_e - \alpha_e^*) \phi_j^{c_k}(x_e^k), \quad w_{irt}^{pq} = \sum_{e=1}^N (\alpha_e - \alpha_e^*) \phi_{irt}^{c_{pq}}(x_e^p, x_e^q).$$

Тож сума коефіцієнтів рівна:

$$\sum_{j=0}^{c_k} w_j^k = \sum_{j=0}^{c_k} \sum_{e=1}^N (\alpha_e - \alpha_e^*) \phi_j^{c_k}(x_e^k) = \sum_{e=1}^N (\alpha_e - \alpha_e^*) \underbrace{\sum_{j=0}^{c_k} \phi_j^{c_k}(x_e^k)}_{=1} = 0$$

і аналогічно

$$\sum_{i+r+t=c_{pq}} w_{irt}^{pq} = \sum_{e=1}^N (\alpha_e - \alpha_e^*) \underbrace{\sum_{i+r+t=c_{pq}} \phi_{irt}^{c_{pq}}(x_e^p, x_e^q)}_{=1} = 0$$

в силу обмежень двоїстої задачі максимізації (3.8): $\sum_{e=1}^N (\alpha_e - \alpha_e^*) = 0$. \square

Це означає, що модель отримана за допомогою ПРИАМ автоматично є збалансованою і може бути відразу ж використана для генерації нечітких правил за підмоделями.

3.3.4. Апостеріорний прогнозний розподіл. Довірчі інтервали

В підрозділі 3.3.1. ми визначили апріорний розподіл вектору параметрів у вигляді багатовимірною нормального розподілу $\mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{I})$. Тим самим

неявно був визначений апіорний розподіл моделей $P(f|\mathcal{M}) = \mathcal{N}(\mathbf{f}|\mathbf{b}, \mathbf{K})$, оскільки для кожного \mathbf{w} з розподілу існує відповідна функція $f(\mathbf{x}) = \mathcal{M}(\mathbf{x}, \mathbf{w})$. В цьому випадку прийнято говорити про перехід від параметрично-просторового погляду на гаусові процеси до функціонально-просторового. Елементи матриці коваріації \mathbf{K} , яка побудована на вибірці вхідних векторів $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$:

$$\begin{aligned} K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) &= \text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = \mathbf{E}[(f_i - \mathbf{E}f_i)(f_j - \mathbf{E}f_j)] = \\ &= \Phi(\mathbf{x}_i)^\top \mathbf{E}[(\mathbf{w} - \mathbf{E}\mathbf{w})(\mathbf{w} - \mathbf{E}\mathbf{w})^\top] \Phi(\mathbf{x}_j) = \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}_j). \end{aligned}$$

Апостеріорний розподіл $P(f|\mathcal{D}, \mathcal{M})$ можна використати для обчислення апостеріорного прогнозного розподілу $f_* = f(\mathbf{x}_*)$ для будь-якого вектору входу \mathbf{x}_* . Для цього необхідно обчислити наступний інтеграл по апостеріорній невизначеності f :

$$P(f_*|\mathcal{D}, \mathcal{M}, \mathbf{x}_*) = \int_{\mathcal{M}} P(f_*|f, \mathbf{X}, \mathbf{x}_*) P(f|\mathcal{D}, \mathcal{M}) df.$$

Оскільки в якості апіорної інформації про модель використано гаусівський процес, то апіорний сумісний розподіл має вигляд багатовимірного нормального розподілу:

$$P(f, f_*|\mathbf{X}, \mathbf{x}_*) = \mathcal{N}\left(\begin{bmatrix} b \\ b_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} \end{bmatrix}\right),$$

де $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$, $\mathbf{k}_* = (k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_N, \mathbf{x}_*))^\top$. Звідси можна отримати умовний розподіл $f_*|f$, який виражає залежність f_* від f індуковану апіорним гаусівським процесом [35, 40]:

$$P(f_*|f, \mathbf{X}, \mathbf{x}_*) = \mathcal{N}(b_* + \mathbf{k}_*^\top \mathbf{K}^{-1}(f - b), k_{**} - \mathbf{k}_*^\top \mathbf{K}^{-1} \mathbf{k}_*).$$

Якщо розподіл шуму $P(\delta|f, \mathcal{M})$ має вигляд нормального $\mathcal{N}(\delta|0, \sigma_N^2)$, то загальна дисперсія апостеріорного прогнозного розподілу становить:

$$\sigma^2 = k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma_N^2 \mathbf{I})^{-1} \mathbf{k}_*.$$

Хоча ми використовуємо розподіл шуму відмінний від нормального, цю формулу можна використовувати як наближення дисперсії справжнього апостеріорного прогнозного розподілу. На рис. 3.5 показано, що шум байєсівської регресії опорних векторів достатньо близький до нормального.

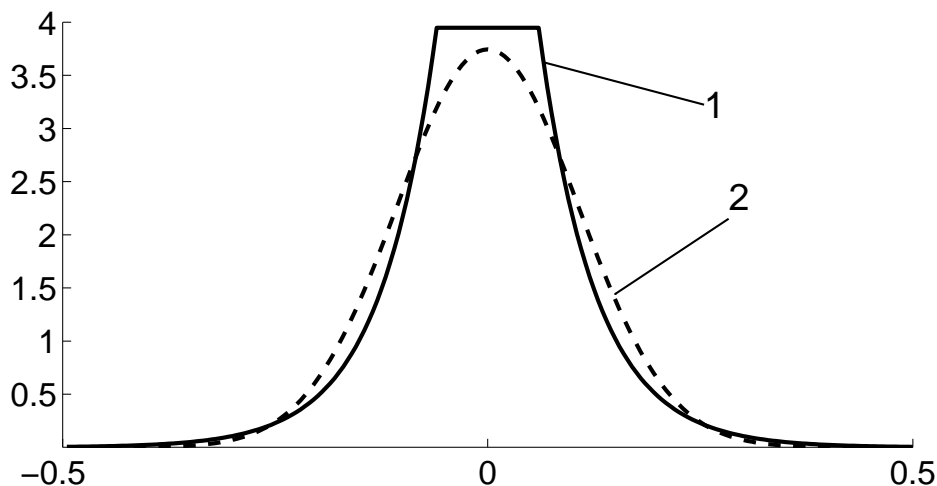


Рис. 3.5. Порівняння моделі шуму БРОВ (1) з гіперпараметрами $\beta = 15$, $\epsilon = 0.06$ та моделі нормального шуму (2) з нульовим середнім та дисперсією $\sigma_N^2 = 0.0114$.

Визначимо дисперсію шуму σ_N^2 відповідно до моделі шуму (3.12):

$$\sigma_N^2 = \int_{-\infty}^{+\infty} \delta^2 \cdot \frac{\beta}{2(1 + \epsilon\beta)} \exp(-\beta|\delta|_\epsilon) d\delta = \frac{2}{\beta^2} + \frac{\epsilon^2(\epsilon\beta + 3)}{3(\epsilon\beta + 1)}. \quad (3.23)$$

Як показано в роботі [200] для обчислення коваріацій \mathbf{K} , \mathbf{k}_* можна використовувати не повну вибірку вхідних векторів \mathbf{X} , а тільки ту її частину, яка відповідає маргінальним опорним векторам:

$$\mathbf{X}_M = \{\mathbf{x}_i : |y_i - f(\mathbf{x}_i)| = \epsilon, \quad 0 < |\alpha_i - \alpha_i^*| < \beta\}.$$

Таким чином справедлива наступна оцінка дисперсії апостеріорного прогнозного розподілу:

$$\sigma^2 \approx k_{**} - \mathbf{k}_*^\top(\mathbf{X}_M)(\mathbf{K}(\mathbf{X}_M) + \sigma_N^2 \mathbf{I})^{-1} \mathbf{k}_*(\mathbf{X}_M).$$

де σ_N^2 визначається формулою (3.23). Довірчий інтервал, який відповідає рівню довіри 95%, як правило приблизно оцінюють як $\pm 2\sigma$.

3.4. Висновки до розділу

Метод опорних векторів успішно застосовується для задач побудови нелінійної регресії. Головними перевагами регресії опорних векторів (РОВ) є її робастність, незалежність від розмірності вхідного простору, розв'язання задачі квадратичного програмування замість обернення погано обумовлених матриць в МНК і здатність робити прогноз на коротких вибірках даних. Актуальним в РОВ залишається питання автоматичного вибору гіперпараметрів. Ефективним методом визначення оптимальних значень гіперпараметрів є байєсівський висновок. Проте недиференційовність ϵ -нечутливої функції втрат Вапника ускладнює байєсівське підтвердження гіперпараметрів.

Запроваджена автором локально згладжена ϵ -нечутлива функція втрат забезпечує існування гессіану байєсівської моделі регресії опорних векторів при підтвердженні гіперпараметрів. Метод Лапласа оцінки підтвердження з локально згладженою функцією втрат генерує критерій байєсівського підтвердження для оцінки адекватності моделі РОВ. Для визначення гіперпараметрів байєсівської моделі РОВ доцільно використовувати методи нелінійної оптимізації.

Розроблений автором індуктивний метод побудови байєсівської регресії опорних векторів в характеристичному просторі поліноміальних функцій Без'є-Бернштейна індукує збалансовані нейронечіткі моделі в формі Бернштейна. Ці моделі легко можна інтерпретувати за допомогою нечітких правил. При цьому за рахунок використання поліномів у формі Бернштейна однієї та двох змінних алгоритмічна складність нечіткої бази знань (виражена в кількості правил) є квадратичною, а не експонентною як в нейронечітких мережах на основі В-сплайнів. Існує можливість оцінки довірчих інтервалів. Метод також дозволяє задавати апіорну інформацію про структуру моделі.

РОЗДІЛ 4

РОЗРОБКА ТА ЗАСТОСУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ПРОГНОЗУВАННЯ СТОХАСТИЧНИХ ПРОЦЕСІВ

4.1. Розробка інформаційної системи обробки даних спостережень

Як правило інформаційні системи розглядають на трьох рівнях [137]: концептуальному, логічному та фізичному. На концептуальному рівні визначають змістовний аспект інформаційних систем, на логічному відображають їх формалізоване представлення, а на фізичному проводять програмно-апаратну реалізацію.

4.1.1. Концептуальний рівень інформаційної системи

Виділимо такі основні структури інформаційної системи як процеси та процедури, тоді на концептуальному рівні систему можна представити у вигляді схеми на рис. 4.1. Технологія обробки інформації починається з формування інформаційного ресурсу, тобто пошуку та збору інформації. При цьому використовується мережа Інтернет та зв'язки з організаціями. Інформація передається поштою, як правило електронною, або завантажується безпосередньо з Інтернету (через FTP). Для забезпечення достовірності інформації під час її підготовки до вводу застосовується візуальний метод контролю. Після введення інформації в ЕОМ дані обробляються і перетворюються у форму зручну для візуального відображення. Процедура процесу накопичення полягає в збереженні даних на сервері. При цьому дані

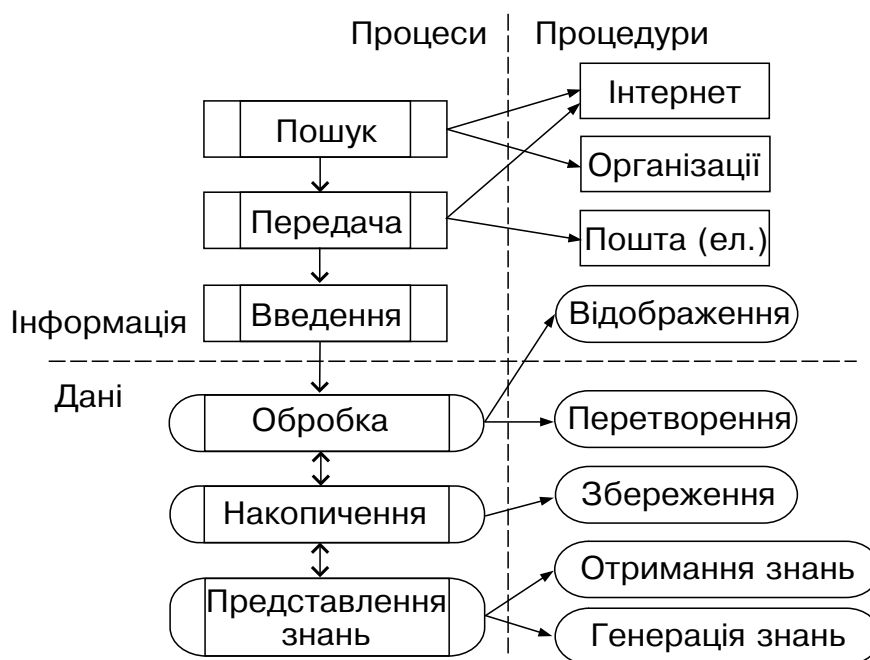


Рис. 4.1. Концептуальна модель інформаційної системи. Блоки у вигляді прямокутників відображають ручні процеси або процедури. Овальна форма блоків відповідає автоматичним операціям.

зберігаються в структурах, які забезпечують ефективне розміщення даних в пам'яті та їх швидкий пошук. Процес представлення знань складається з процедур отримання формалізованих знань і генерації нових знань на основі отриманих.

Схема інформаційних потоків системи на концептуальному рівні представлена на рис. 4.2. На схемі визначені такі елементи зовнішнього середовища як "користувач", "хостинг провайдер" та "дослідник". Показана їх взаємодія з інформаційними сховищами Б1 та Б2 через інформаційні процеси. Так хостинг провайдер підтримує сервер баз даних та забезпечує web-хостинг сайту. Дослідник має можливість поповнювати базу алгоритмів новими вдосконаленими версіями. Користувач може як завантажити та обробити вже існуючі вибірки даних так і поповнити базу вибірок новою вибіркою в спеціально визначеному форматі.

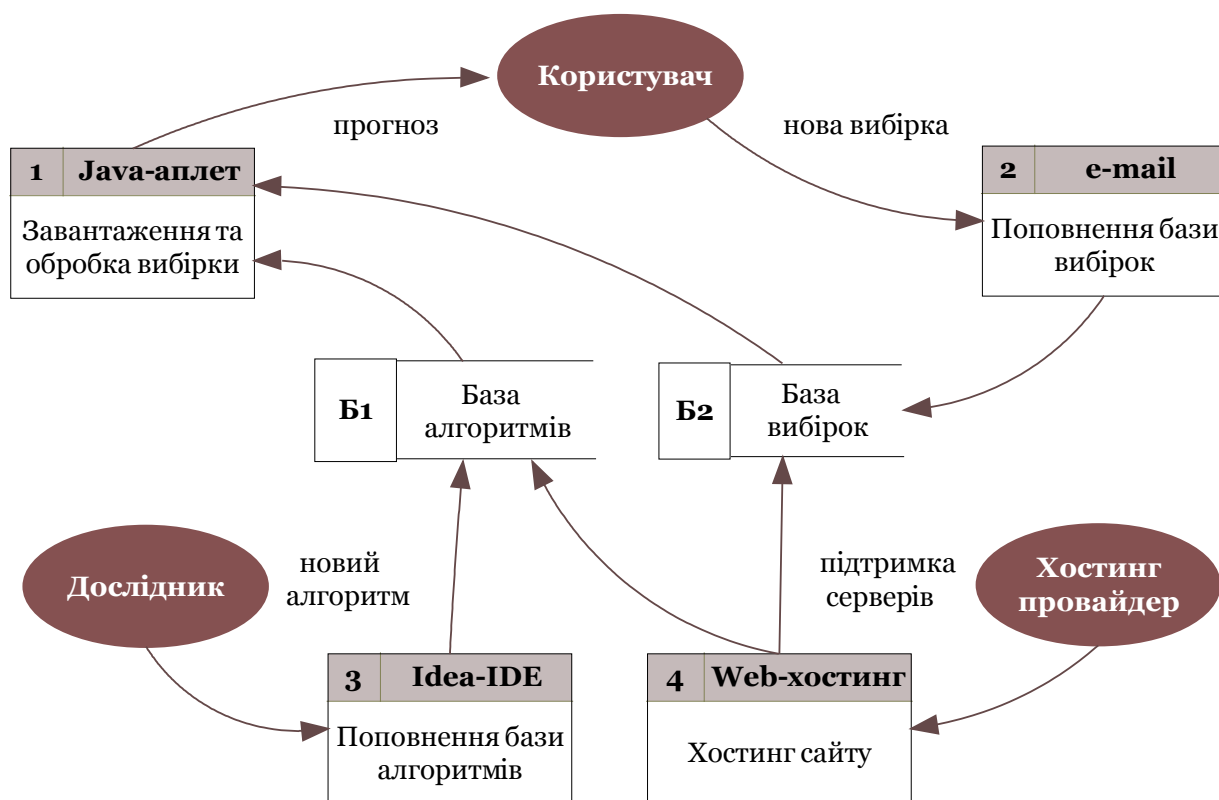


Рис. 4.2. Інформаційні потоки системи.

4.1.2. Логічний рівень інформаційної системи

Логічний рівень є комплексом взаємодіючих моделей, які формалізують інформаційні процеси. При обробці даних формуються три основні інформаційні процеси: обробка, накопичення даних і представлення знань (рис. 4.3). Відповідно модель обробки даних включає в себе алгоритми перетворення даних (сортування, пошук, створення, форматування структур), алгоритми відображення (текстова інформація, таблиці, діаграми, графіки) і допоміжні алгоритми системного керування ресурсами ЕОМ (пам'яттю, процесором, зовнішніми пристроями такими як принтер). А саме — таблична інформація включає представлення повного розв'язку у вигляді коефіцієнтів Лагранжа розкладу опорних векторів, а також послідовні ітерації байєсівського висновку з показниками КБП, значень гіперпараметрів, емпіричного та регуляризованого ризиків, кількості опорних векторів, показників СКП та САП

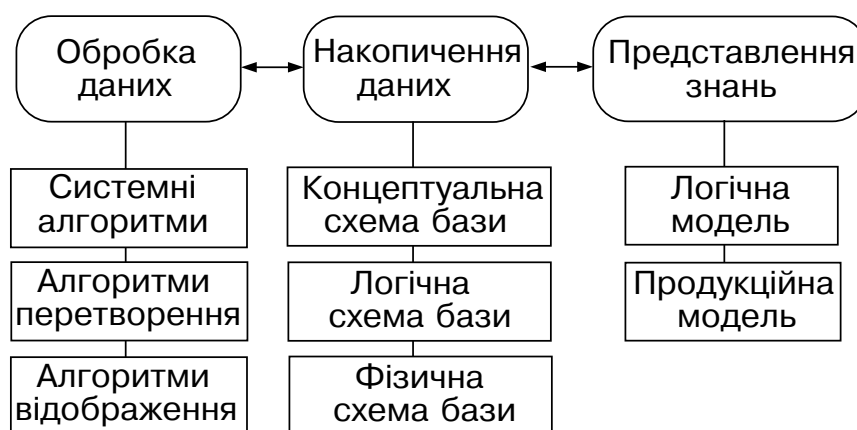


Рис. 4.3. Схема взаємодії моделей інформаційної системи.

в нормованому просторі. Графічна інформація включає результат ідентифікації ПРИАМ з довірчим 95% інтервалом, порівняльний аналіз ПРИАМ з алгоритмами МГВА, НМГВА, ANFIS та PHM¹ за залишками навчальної та прогнозовної вибірок, контурні лінії залежностей КБП та СКП від гіперпараметрів. Для еволюційного методу побудови робастних нейронечітких моделей у формі Бернштейна виводиться графічне зображення розміщення контрольних точок. Алгоритми обробки даних зберігаються в пакетах. На рис. 4.4 представлена діаграма основних пакетів інформаційної системи, які поділяються на три групи: базові пакети та пакети двох Java-апплетів. В-В.NF пакет реалізує алгоритм моделювання нейронечітких мереж на основі поліноміальних функцій Без'є-Бернштейна (розділ 2). PRIAMUS аплет реалізує алгоритм ПРИАМ (розділ 3).

Програми та дані формують завдання, які надходять із системи вводу або накопичення. Завдання утворюють вхідну чергу, з якої вони вибираються користувачем. Модель обробки даних використовує пакетний режим обробки завдань. Завдання також можуть оброблятися паралельно, оскільки реалізовані механізми багатопоточності. Схема взаємодії користувача з інформа-

¹Для оцінки якості роботи ПРИАМ за результатами прогнозування проводиться порівняльний аналіз з роботою МГВА в пакеті "NeuroShell 2" (<http://www.wardsystems.com>), НМГВА в пакеті "GMDH Modeler 0.9.37" (<mailto:igor.zayets@gmail.com>) та ANFIS і PHM в пакеті "NeuroSolutions 5" (<http://www.neurosolutions.com>).

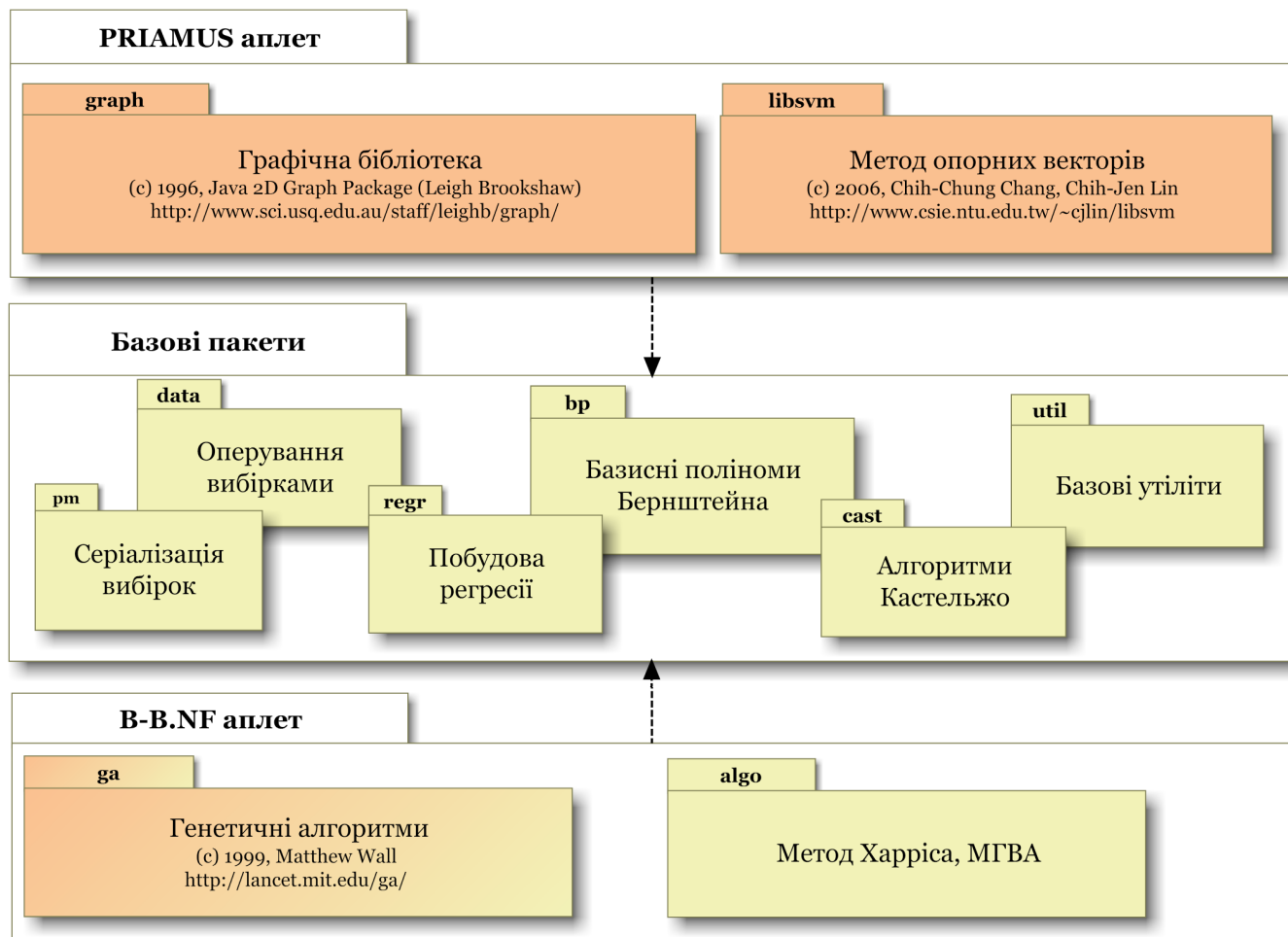


Рис. 4.4. Діаграма пакетів (бібліотек) системи.

ційною системою представлена на рис. 4.5. Відповідно користувач отримує список завдань (вибірок) та алгоритмів, вибирає необхідний алгоритм та завдання і завантажує паралельну сесію. В цій сесії користувач має можливість налаштувати параметри завдання і алгоритму, побудувати модель та отримати детальні результати прогнозу на основі цієї моделі.

Модель накопичення даних формалізує інформаційну базу, яка на сучасному рівні як правило представляється базою даних. Концептуальна схема інформаційної бази визначає яка і в якому об'ємі інформація може зберігатись: це насамперед дані спостережень за стохастичними процесами, а також супроводжуючі їх документи. Логічна схема інформаційної бази використо-

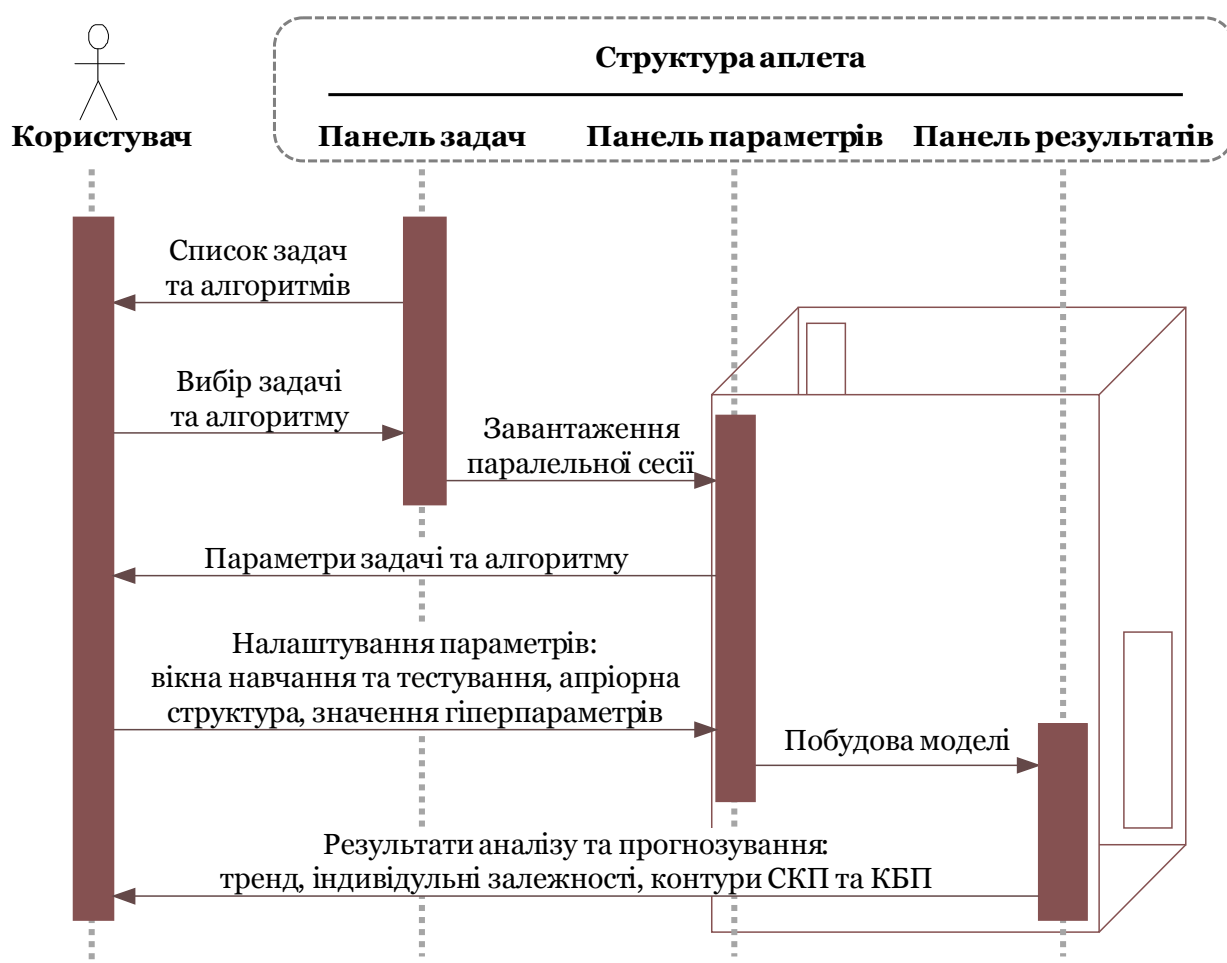


Рис. 4.5. Діаграма взаємодії користувача з аплетом.

вує реляційний підхід, де взаємозв'язки між елементами даних представлені у вигляді двовимірних таблиць, кожен стовпчик яких містить дані спостережень за окремим фактором. Фізична схема визначає як правило систему керування базою даних (СУБД). Теоретично для реляційної бази можна використовувати такі потужні всесвітньо відомі СУБД як Oracle, або MS SQL, проте оскільки ми проектуємо лише прототип наразі використовується проста але надійна СУБД `mysql`. Також існує механізм завантаження зовнішніх файлів з даними спостережень у спеціально визначеному форматі.

В якості моделей для представлення знань використовуються логічна та продукційна моделі. Логічна модель є системою індуктивного висновку, де правила виводу породжуються на основі скінченного числа прецедентів (даних

спостережень). Результатом функціонування логічної моделі є продукційна модель, де правила задаються у вигляді: якщо умова то дія, якщо причина то наслідок, якщо ситуація то рішення.

4.1.3. Фізичний рівень інформаційної системи. Програмна реалізація

На фізичному рівні інформаційна система складається з декількох підсистем. Це підсистеми, які реалізують на фізичному рівні інформаційні процеси обробки, накопичення даних та представлення знань. На рис. 4.6 представлена схема розміщення носіїв цих підсистем. З метою зменшення ви-

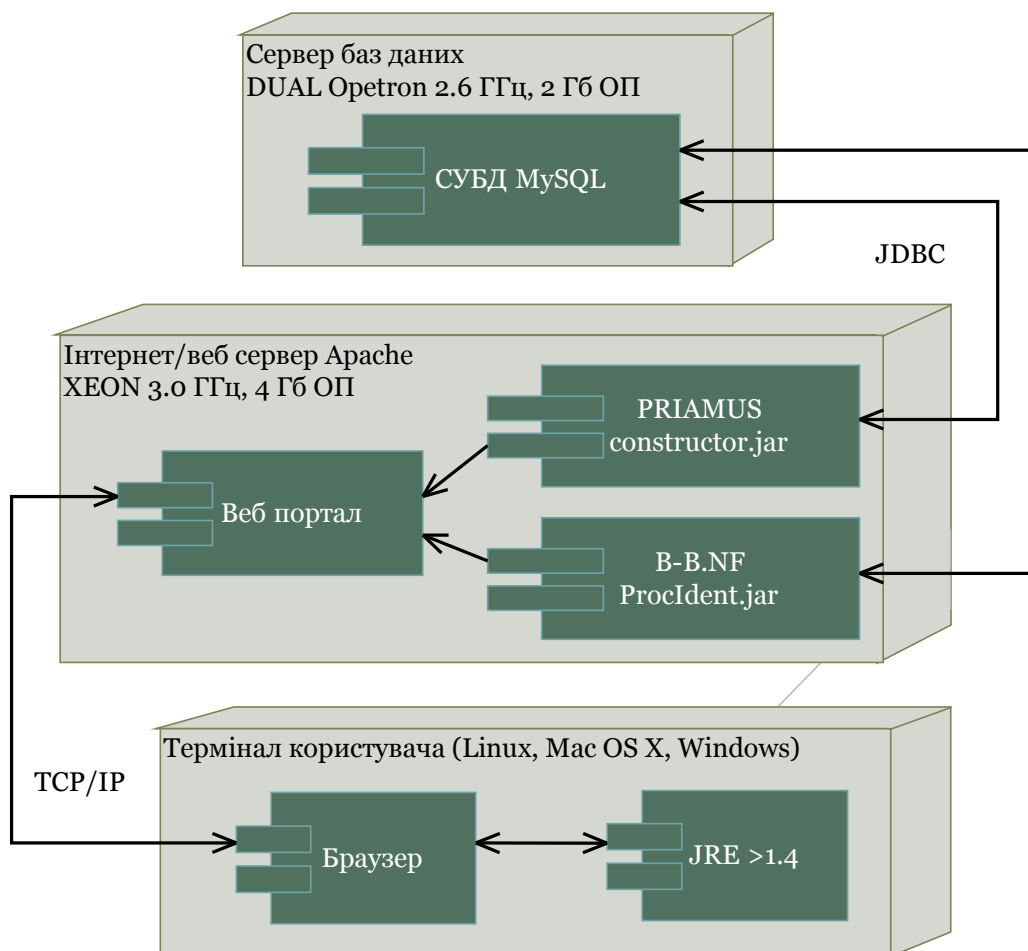


Рис. 4.6. Діаграма топологій інформаційної системи.

трат на створення і експлуатацію інформаційної системи максимально ви-

користовуються типові технічні засоби та програмне забезпечення. Тож для виконання функцій підсистем використовуються ЕОМ середнього класу — сервери. Сервери надані компанією FREEhost.com.ua. Сервер баз даних має наступні характеристики: DUAL Opetron 2.6 ГГц, 2 Гб оперативної пам'яті, вінчестерний SCSI масив для розміщення баз користувачів і масив для збереження бекапів баз даних кожні 14 днів. На сервері розміщується СУБД MySQL. Хостинг-сервер побудовано на базі процесору XEON 3.0 ГГц з 4 Гб оперативної пам'яті і дисковою системою SCSIx2. Дзеркальний SCSI рейд забезпечує високу надійність і безпеку даних. Поштовий сервер має конфігурацію: Xeon Dual Core 2.0, 2 Гб оперативної пам'яті, дискова система SATA 500 Гб x 4. На сервері розміщений сайт (веб портал), який підключає Java архіви: constructor.jar для аплету PRIAMUS і ProcIdent.jar для аплету В-В.NF. Аплети взаємодіють з базою вибірок через технологію JDBC. Таким чином будь-який користувач на своєму терміналі, на якому повинен бути браузер з підтримкою JRE версії не нижче 1.4, має можливість через TCP/IP завантажити аплет, вибірку даних та побудувати прогнозуючу модель.

Програмну реалізацію запропонованих методів та алгоритмів виконано в середовищах MATLAB 6.1, JDK 1.5.0.4.

Загальна схема інформаційної системи зображена на рис. 4.7. Показані можливі варіанти підключення користувачів: використовуючи мобільні комп'ютери, використовуючи PDA (personal digital assistant), через безпроводне з'єднання, через мобільні телефони.

4.2. Апробація PRIAM на еталонних моделях

В цьому підрозділі ми розглянемо роботу PRIAM на відомих штучних та реальних еталонних моделях нелінійної та лінійної регресій. Моделі зібрані з декількох джерел. Проект StRD² містить еталонні моделі, які сертифіковані за кількістю ступенів свободи (кількістю параметрів), показниками суми

²Statistical Reference Datasets: <http://www.itl.nist.gov/div898/strd/>

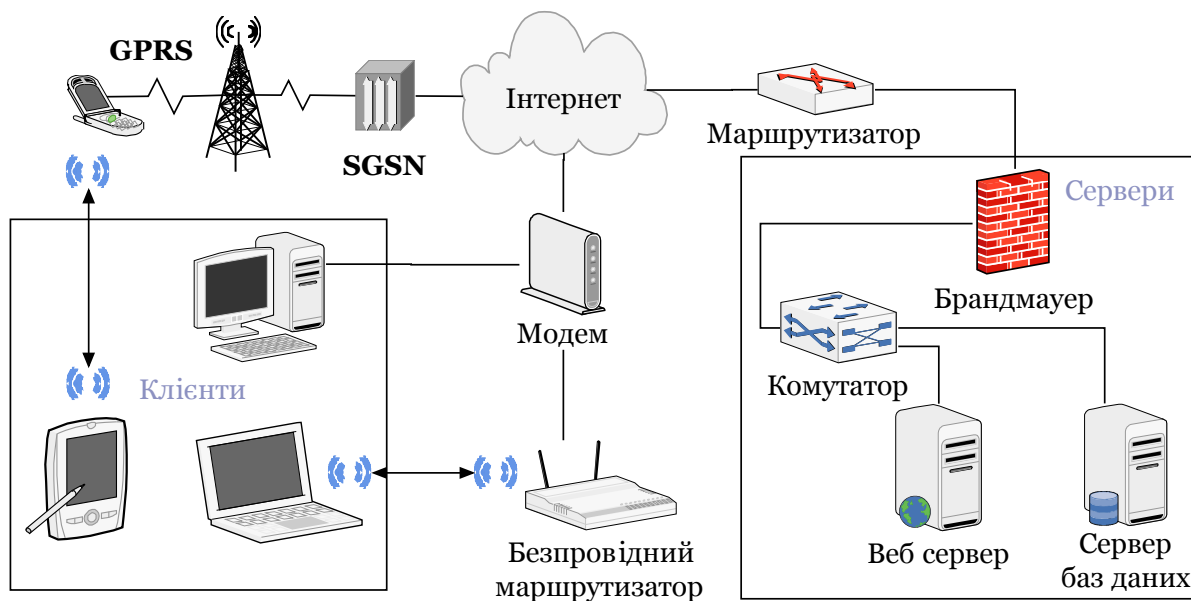


Рис. 4.7. Схема мережі інформаційної системи.

квадратів залишків та F-статистики. Метою цього проекту є покращення якості статистичного програмного забезпечення, надаючи еталонні вибірки даних з сертифікованими результатами. Другим джерелом є база даних для тестування алгоритмів машинного навчання [201]. Також були використані дані з проекту Delve ³.

4.2.1. Модель Лонглі (Longley)

Розглянемо модель, описану Дж.Лонглі [202] як типовий приклад мультиколінеарної регресії (лінійна модель):

$$y(\mathbf{x}) = b_0 + b_1x^1 + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5 + b_6x^6.$$

Змінні мають зміст наступних макроекономічних показників США з 1947 по 1962 рр.: y — число людей, які мають роботу; x^1 — індекс цін (неявний дефлятор валового національного продукту), 1954 рік — 100%; x^2 — валовий національний продукт; x^3 — число безробітних; x^4 — число військовослужбов-

³Data for Evaluating Learning in Valid Experiments: <http://www.cs.toronto.edu/~delve/>

ців; x^5 — кількість дорослого населення (старші 14 років); x^6 — рік. Рис. 4.8⁴ відображає результат ідентифікації ПРИАМ на основі 11 точок. Останні п'ять

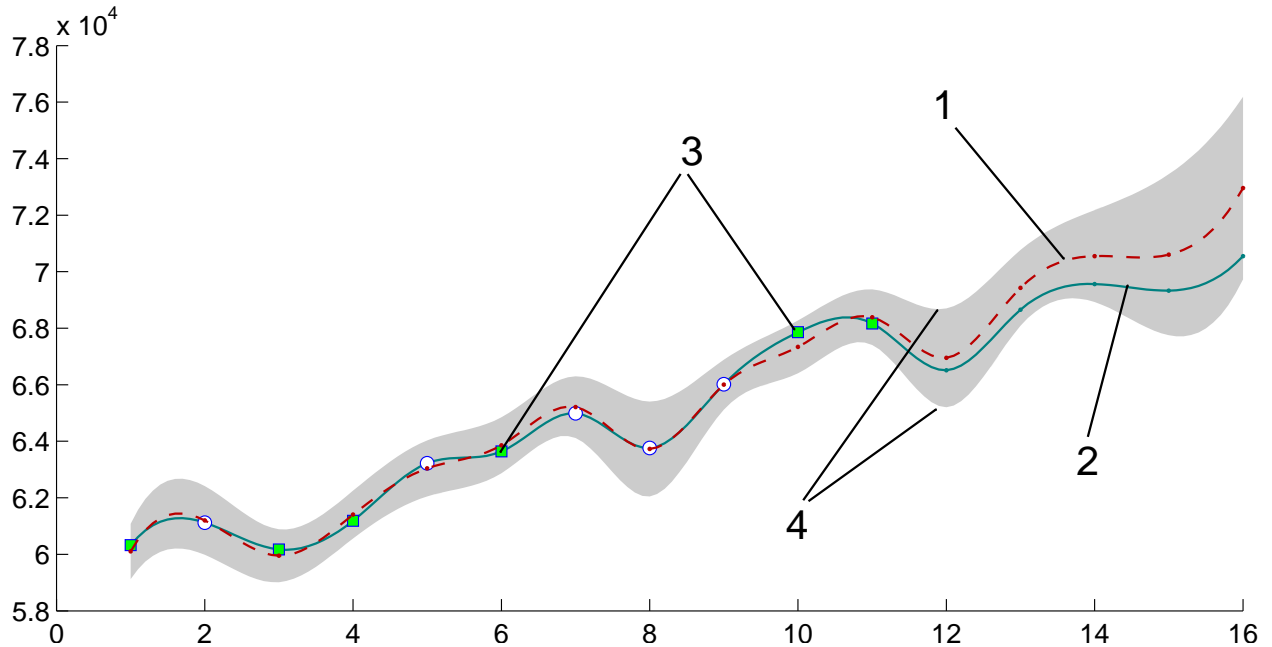


Рис. 4.8. ПРИАМ для моделі Лонглі: 1 — прогноз, 2 — реальні дані, 3 — опорні вектори, 4 — 95% довірчий інтервал.

точок показують прогноз. Модель описується шістьма опорними векторами. Повний розв'язок представлено в таблиці А.1. Апріорна і оптимальна конфігурації характеристичного простору моделі Лонглі мають відповідно вигляд діагональних матриць:

$$\mathbf{C}^{(0)} = \text{diag}\{1, 1, 1, 1, 1, 1\}, \quad \mathbf{C}^{(\cdot)} = \text{diag}\{0, 1, 1, 1, 1, 1\}.$$

Також в таблиці А.1 наведений ітеративний процес пошуку оптимальних значень гіперпараметрів на основі байєсівського підтвердження. Зокрема показані значення критерію байєсівського підтвердження (КБП), ширини смуги нечутливості $\tilde{\epsilon}$ на нормованих даних, оберненого коефіцієнту регуляризації β , кількості опорних векторів N_{sv} , емпіричного та регуляризованого ризику на навчальній вибірці, середньо квадратичної помилки і середньої абсолютної

⁴Тут і далі на всіх рисунках крива моделі РОВ проведена через точки вибірки за допомогою кубічної сплайн-інтерполяції. Це не є точним зображенням відгуку але коректно, оскільки РОВ будується як гладка апроксимація, що проходить через задані точки.

помилки на нормованій прогнозній вибірці. Розклад опорних векторів:

$$f_{\text{opt}}(\mathbf{x}) = 0.56 + 11.2k(\mathbf{x}_1, \mathbf{x}) + 20.4k(\mathbf{x}_3, \mathbf{x}) - 38.5k(\mathbf{x}_4, \mathbf{x}) - \\ - 12k(\mathbf{x}_6, \mathbf{x}) + 48.8k(\mathbf{x}_{10}, \mathbf{x}) - 29.8k(\mathbf{x}_{11}, \mathbf{x}).$$

Представлення моделі у формі Бернштейна:

$$f_{\text{opt}}(\mathbf{x}) = 0.56 + 0.003\phi_0^1(x^2) - 0.003\phi_1^1(x^2) + 0.22\phi_0^1(x^3) - 0.22\phi_1^1(x^3) + \\ + 0.09\phi_0^1(x^4) - 0.09\phi_1^1(x^4) - 0.38\phi_0^1(x^5) + 0.38\phi_1^1(x^5) - 0.42\phi_0^1(x^6) + 0.42\phi_1^1(x^6).$$

4.2.2. Модель Філіппеллі (Filippelli)

Розглянемо іншу модель, яка є результатом досліджень NIST⁵ і належить А.Філіппеллі (поліном 10-го порядку):

$$y(t) = b_0 + b_1x(t) + b_2x^2(t) + b_3x^3(t) + \dots + b_{10}x^{10}(t) + e(t).$$

Рис. 4.9 відображає результат ідентифікації ПРІАМ на основі 33 точок. Остан-

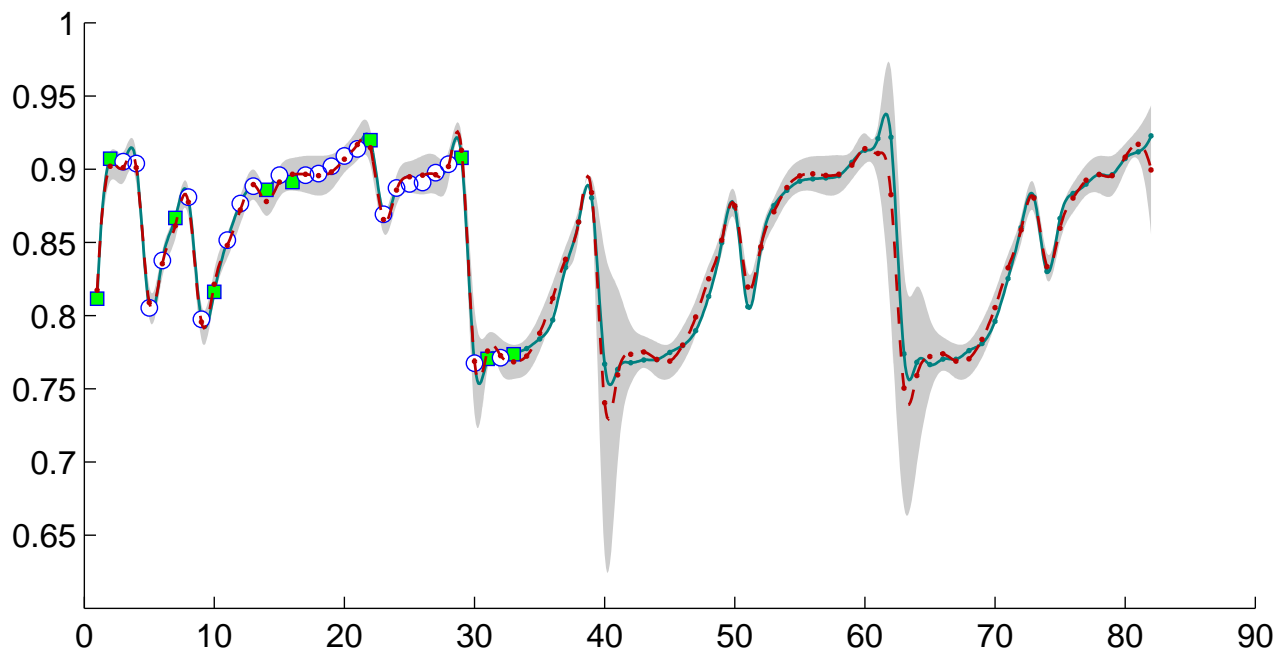


Рис. 4.9. ПРІАМ для моделі Філіппеллі.

ні 49 точок показують прогноз. Модель описується 10 опорними векторами.

⁵National Institute of Standards and Technology — www.nist.gov

Повний розв'язок представлено в таблиці А.2. Наведений ітеративний процес пошуку оптимальних значень гіперпараметрів на основі байєсівського підтвердження. Априорна і оптимальна конфігурації характеристичного простору моделі Філіппеллі мають відповідно вигляд:

$$\mathbf{C}^{(0)} = (1), \quad \mathbf{C}^{(\cdot)} = (9).$$

Розклад опорних векторів:

$$\begin{aligned} f_{\text{opt}}(x) = & 0.55 - 50k(x_1, x) + 16.6k(x_2, x) + 26.1k(x_7, x) + \\ & - 4.7k(x_{10}, x) + 50k(x_{14}, x) - 43.1k(x_{16}, x) + \\ & + 0.5k(x_{22}, x) - 6.9k(x_{29}, x) - 6k(x_{31}, x) + 17.8k(x_{33}, x). \end{aligned}$$

Представлення моделі у формі Бернштейна:

$$\begin{aligned} f_{\text{opt}}(x) = & 0.55 - 0.7\phi_0^9(x) + 0.1\phi_1^9(x) - 1.3\phi_2^9(x) - 0.9\phi_3^9(x) + \\ & + 1.2\phi_4^9(x) + 0.8\phi_5^9(x) - 0.4\phi_6^9(x) + 0.2\phi_7^9(x) + 0.8\phi_8^9(x) + 0.2\phi_9^9(x). \end{aligned}$$

4.2.3. Модель Фрідмана (Friedman)

Розглянемо так звану модель Фрідмана [53] з десятьма рівномірно розподіленими на $[0; 1]$ вхідними змінними, тільки перші п'ять з яких входять в реальну модель:

$$y(\mathbf{x}) = 10 \sin(\pi x^1 x^2) + 20 \left(x^3 - \frac{1}{2} \right)^2 + 10x^4 + 5x^5 + \mathcal{N}(0, 1.0).$$

Для навчальної вибірки згенеровані 500 точок, для прогнозування згенеровані інші 500 точок. Оптимальні значення гіперпараметрів $\beta = 7.6$, $\epsilon = 0.02$.

Представлення моделі у формі Бернштейна:

$$\begin{aligned} f_{\text{opt}}(\mathbf{x}) = & 0.51 - 0.13\phi_0^1(x^1) + 0.13\phi_1^1(x^1) - 0.14\phi_0^1(x^2) + 0.14\phi_1^1(x^2) + \\ & + 0.01\phi_0^1(x^3) - 0.01\phi_1^1(x^3) - 0.18\phi_0^1(x^4) + 0.18\phi_1^1(x^4) - 0.09\phi_0^1(x^5) + 0.09\phi_1^1(x^5). \end{aligned}$$

Як можна бачити, модель добре відображає рівнозначний вплив факторів x^1 , x^2 . Також практичну незначність фактору x^3 . І вдвічі більший ефект від x^4 , ніж від x^5 .

Спрощена модель Фрідмана

Також іноді аналізують спрощену модель Фрідмана, яка має вигляд:

$$y(\mathbf{x}) = 10 \sin(\pi x^1 x^2) + \mathcal{N}(0, 1.0).$$

Рис. 4.10 відображає результат ідентифікації ПРИАМ на основі 10 точок і результат прогнозування останніх 20 точках. Модель описується 10 опорними

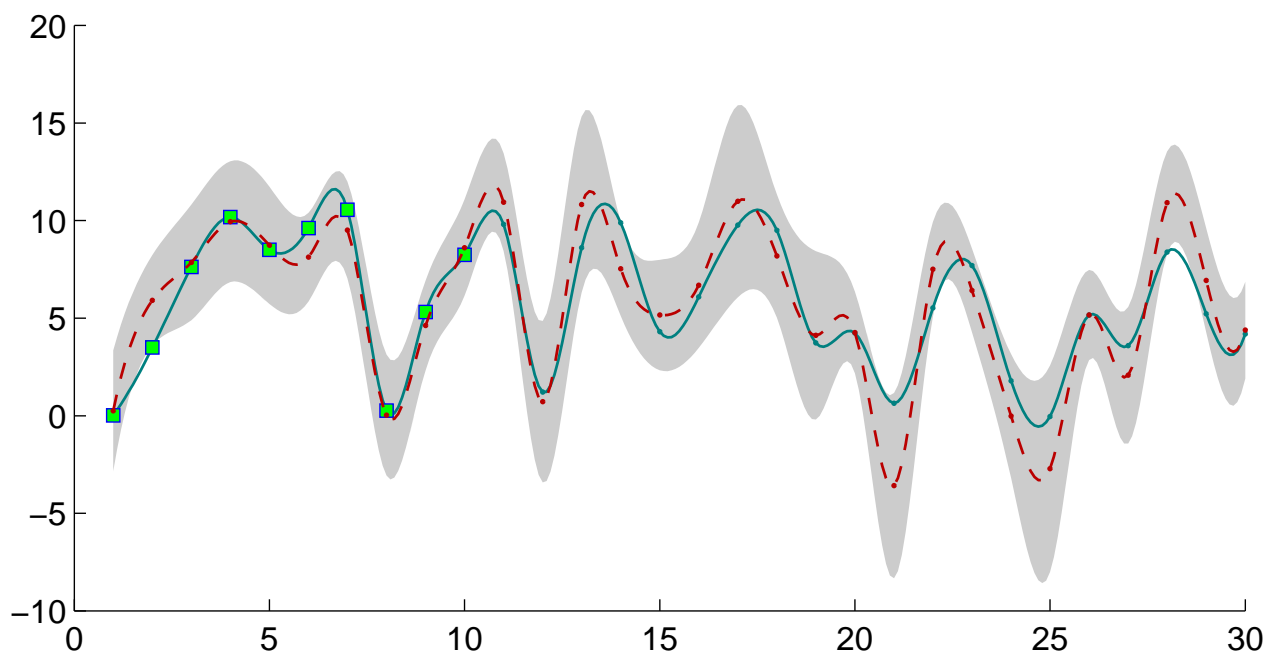


Рис. 4.10. ПРИАМ для спрощеної моделі Фрідмана.

векторами. Повний розв'язок представлено в таблиці А.3. Показані шість послідовних ітерацій байєсівської індукції. Початкова і кінцева конфігурації характеристичного простору відповідно мають вигляд:

$$\mathbf{C}^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{C}^{(\cdot)} = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}.$$

Розклад опорних векторів:

$$f_{\text{opt}}(\mathbf{x}) = 0.49 - 7.2k(\mathbf{x}_1, \mathbf{x}) - 9.7k(\mathbf{x}_2, \mathbf{x}) - 4.2k(\mathbf{x}_3, \mathbf{x}) + 1.8k(\mathbf{x}_4, \mathbf{x}) - \\ - 5k(\mathbf{x}_5, \mathbf{x}) + 9.7k(\mathbf{x}_6, \mathbf{x}) + 9.7k(\mathbf{x}_7, \mathbf{x}) + 4.9k(\mathbf{x}_8, \mathbf{x}) + 9.7k(\mathbf{x}_9, \mathbf{x}) - 9.7k(\mathbf{x}_{10}, \mathbf{x}).$$

Представлення моделі у формі Бернштейна:

$$\begin{aligned} f_{\text{opt}}(\mathbf{x}) = & 0.49 - 0.6\phi_0^2(x^1) + 0.4\phi_1^2(x^1) + 0.2\phi_2^2(x^1) - 0.02\phi_{002}^2(x^1, x^2) - \\ & - 0.29\phi_{011}^2(x^1, x^2) + 0.09\phi_{020}^2(x^1, x^2) - 0.07\phi_{101}^2(x^1, x^2) + 0.42\phi_{110}^2(x^1, x^2) - \\ & - 0.13\phi_{200}^2(x^1, x^2) - 0.6\phi_0^2(x^2) + 0.3\phi_1^2(x^2) + 0.3\phi_2^2(x^2). \end{aligned}$$

4.2.4. Модель AMPG

Розглянемо еталонну задачу моделювання витрат палива автомобілем. Це так звана задача AMPG (automobile miles per gallon) [201]. Вибірка даних містить інформацію про споживання палива для різних автомобілів з відомими характеристиками: об'ємом двигуна (x^1), кінськими силами (x^2), вагою (x^3) та прискоренням (x^4). В експерименті брало участь 392 автомобілі, серед них 300 для навчання і 92 для оцінювання якості узагальнення побудованої моделі. Оптимальні значення гіперпараметрів $\beta = 20$, $\epsilon = 0.2$. Оскільки кількість опорних векторів досить велика, наводимо представлення моделі тільки у формі Бернштейна:

$$\begin{aligned} f_{\text{opt}}(\mathbf{x}) = & 0.27 + 0.07\phi_0^1(x^1) - 0.07\phi_1^1(x^1) + 0.18\phi_0^1(x^2) - 0.18\phi_1^1(x^2) + \\ & + 0.14\phi_0^1(x^3) - 0.14\phi_1^1(x^3) + 0.03\phi_0^1(x^4) - 0.03\phi_1^1(x^4). \end{aligned}$$

Цей результат можна порівняти з методом SUPANOVA С. Гуна [57] на цьому ж модельному прикладі.

4.3. Апробація ПРИАМ на спрощеній моделі макроекономіки України

В цьому підрозділі ми спробуємо застосувати індуктивний метод побудови байєсівської моделі регресії опорних векторів у формі Бернштейна для аналізу і прогнозування поведінки реальних економічних систем. Слід зазначити, що широкий спектр сучасного програмного забезпечення і комп'ютерних технологій з економічного аналізу представлено в монографії

Г.Н.Соколової [203]. Розглянемо декілька рівнянь із спрощеної моделі економіки України запропоновані в роботах [204, 205, 206], які є результатом економічних досліджень EERC в Україні⁶, а також в роботах [207, 208, 209]. Більш складні економічні моделі, які можна було б використати для аналізу, представлені в [210, 211, 212]. Рівняння динаміки індексу споживчих цін (ІСЦ):

$$CPI_t = F1(M2_t, M2v_t, ER_t, RGDP_t).$$

Рівняння динаміки реального споживання господарств (РСГ):

$$RCON_t = F2(RDI_t, R_t).$$

Відповідно: CPI_t — індекс споживчих цін (1990=1), $M2_t$ — грошовий агрегат $M2$ включаючи валютні депозити, $M2v_t$ — швидкість обігу грошей, ER_t — офіційний курс обміну Hrn/USD, $RGDP_t$ — реальний ВВП за даними UEPLAC (в млрд. крб. 1990). $RCON_t$ — реальне споживання (в млрд. крб. 1990), R_t — ринкова відсоткова ставка, RDI_t — реальний дохід після виплати податків (в млрд. крб. 1990). Дані взято за 1998-1999 рр. з офіційних джерел: UEPLAC та Державного комітету статистики⁷.

4.3.1. Моделювання індексу споживчих цін

Рис. 4.11 відображає результат ідентифікації ПРИАМ на основі 14 точок. Останні 10 точок показують прогноз. Модель описується восьма опорними векторами. Повний розв'язок представлено в таблиці А.4. Наведений ітеративний процес пошуку оптимальних значень гіперпараметрів на основі байєсівського підтвердження. Апріорна і оптимальна конфігурації характеристичного простору моделі ІСЦ співпадають і мають вигляд:

$$\mathbf{C}^{(0)} = \text{diag}\{1, 1, 1, 1\}, \quad \mathbf{C}^{(\cdot)} = \text{diag}\{1, 1, 1, 1\}.$$

⁶Economics Education and Research Consortium, <http://www.eerc.kiev.ua>

⁷Ukrainian-European Policy and Legal Advice Centre: <http://www.ueplac.kiev.ua>; сайт ДержКомСтату: <http://www.ukrstat.gov.ua>.

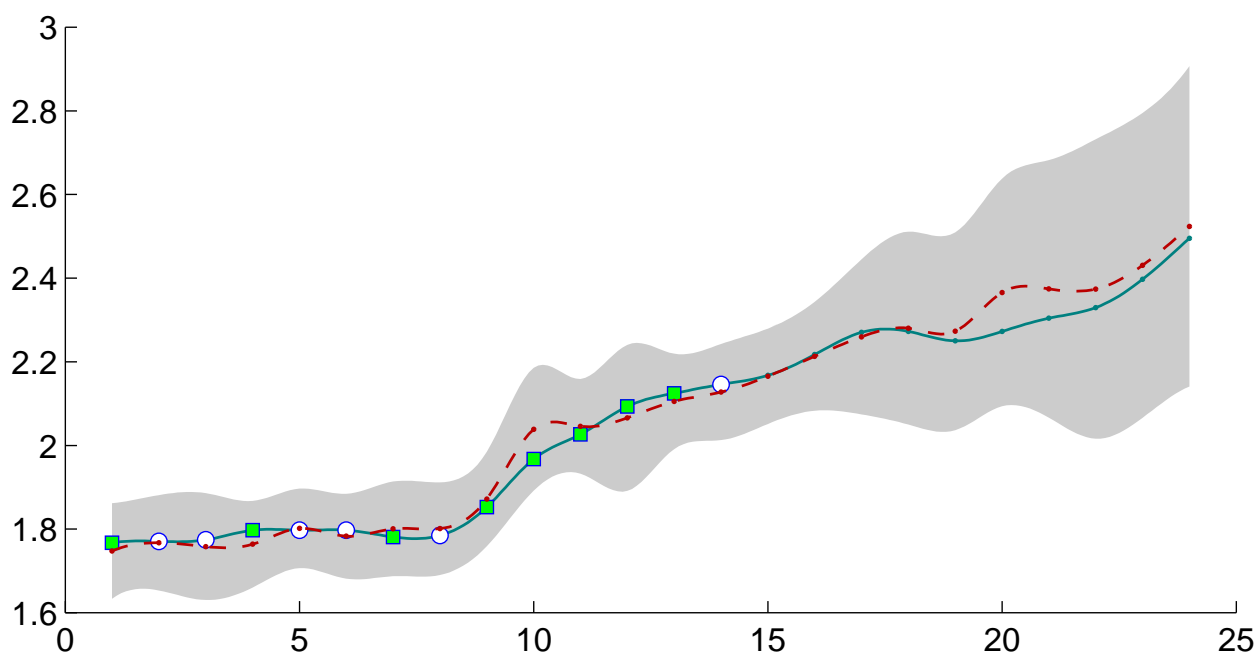


Рис. 4.11. ПРИАМ для індексу споживчих цін.

Розклад опорних векторів:

$$f_{\text{opt}}(\mathbf{x}) = 0.56 + 4.1k(\mathbf{x}_1, \mathbf{x}) + 27.7k(\mathbf{x}_4, \mathbf{x}) - 23k(\mathbf{x}_7, \mathbf{x}) - \\ - 11.6k(\mathbf{x}_9, \mathbf{x}) - 27.7k(\mathbf{x}_{10}, \mathbf{x}) - 13.4k(\mathbf{x}_{11}, \mathbf{x}) + 27.7k(\mathbf{x}_{12}, \mathbf{x}) + 16.2k(\mathbf{x}_{13}, \mathbf{x}).$$

Представлення моделі у формі Бернштейна:

$$f_{\text{opt}}(\mathbf{x}) = 0.56 - 0.37\phi_0^1(x^1) + 0.37\phi_1^1(x^1) - 0.15\phi_0^1(x^2) + 0.15\phi_1^1(x^2) - \\ - 0.21\phi_0^1(x^3) + 0.21\phi_1^1(x^3) + 0.18\phi_0^1(x^4) - 0.18\phi_1^1(x^4).$$

4.3.2. Моделювання реального споживання

Рис. 4.12 відображає результат ідентифікації ПРИАМ на основі 14 точок. Останні 10 точок показують прогноз. Модель описується п'ятьма опорними векторами. Повний розв'язок представлено в таблиці А.5. Наведений ітеративний процес пошуку оптимальних значень гіперпараметрів на основі байєсівського підтвердження. Апріорна і оптимальна конфігурації характеристичного простору моделі РСГ мають відповідно вигляд:

$$\mathbf{C}^{(0)} = \text{diag}\{1, 1\}, \quad \mathbf{C}^{(\cdot)} = \text{diag}\{1, 2\}.$$

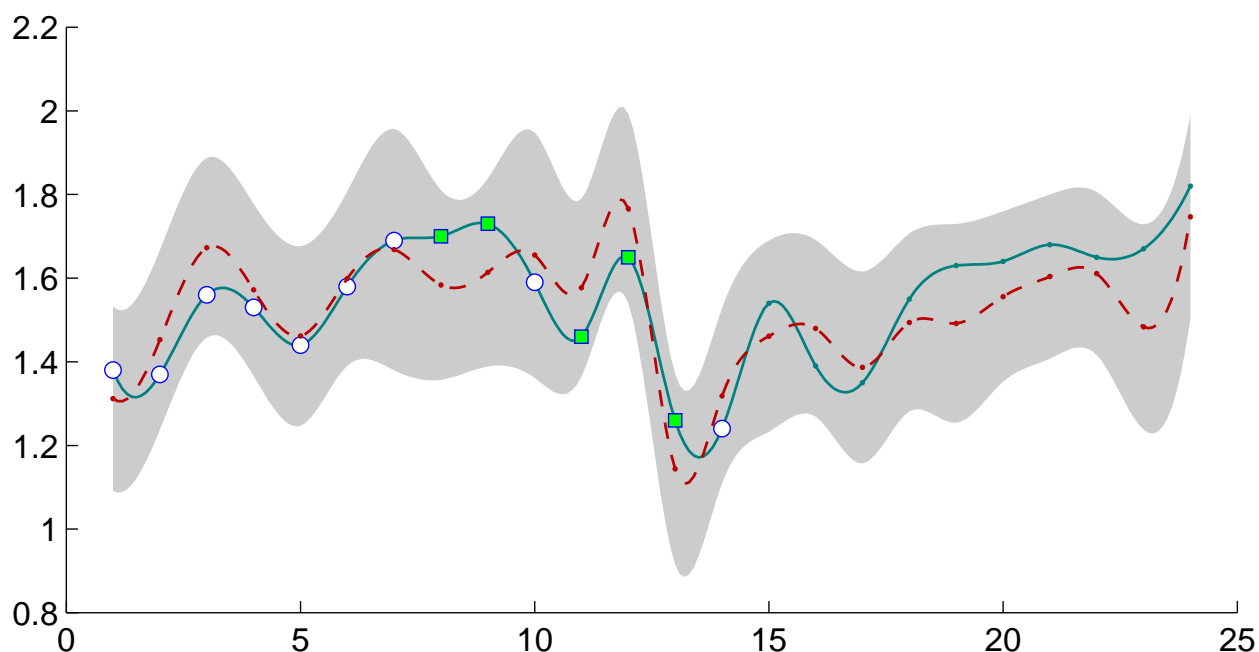


Рис. 4.12. ПРИАМ для реального споживання.

Розклад опорних векторів:

$$f_{\text{opt}}(\mathbf{x}) = 0.43 + 6.6k(\mathbf{x}_8, \mathbf{x}) + 25.5k(\mathbf{x}_9, \mathbf{x}) - \\ - 29.8k(\mathbf{x}_{11}, \mathbf{x}) - 2.6k(\mathbf{x}_{12}, \mathbf{x}) + 0.3k(\mathbf{x}_{13}, \mathbf{x}).$$

Представлення моделі у формі Бернштейна:

$$f_{\text{opt}}(\mathbf{x}) = 0.43 + 0.06\phi_0^1(x^1) - 0.06\phi_1^1(x^1) - \\ - 0.58\phi_0^2(x^2) + 0.13\phi_1^2(x^2) + 0.45\phi_2^2(x^2). \quad (4.1)$$

4.4. Апробація ПРИАМ на гідрометеорологічній моделі вітрових хвиль

Однією з основних задач Севастопольської гідрометеорологічної обсерваторії (СГМО) є гідрометеорологічне обслуговування споживачів інформацією про стан погоди та моря по закріпленій за обсерваторією зоні відповідальності, в тому числі по акваторії ряду морських портів, розташованих на південному узбережжі Криму. Для здійснення безпечного мореплавства в районі морських портів важливе значення має добовий прогноз висоти

хвилі. Основним фактором, який визначає розповсюдження та характеристики хвиль на морі є вітер. Для отримання прогнозу параметрів вітру використовуються прогностичні карти розподілу вітру на основі чисельних прогнозів полів атмосферного тиску на рівні моря. Чисельна схема прогнозу полів хвилювання, яку використовує Севастопольська гідрометеорологічна обсерваторія, ґрунтується на теоретичному методі Шулейкіна, основою якого є теорія поля вітрових хвиль. Цей метод успішно використовується для прогнозування елементів хвиль глибокого відкритого моря. Однак для побудови прогнозів на міліні або біля берега не існує ефективних теоретично обґрунтованих процедур, які б враховували особливості місцевості. Тож на практиці все залежить від досвіду прогнозистів на місцях. В цьому підрозділі ПРИ-АМ створює прогноуючі моделі, які можна використовувати для підтримки прийняття рішень прогнозистами з метою уточнення прогнозу параметрів хвиль.

Для побудови моделі використані дані щоденних спостережень за параметрами вітрових хвиль (висота, період, довжина, напрям) та показниками вітру (напрямок, швидкість, тривалість дії) з 1960 по 2006 рік. Щоденне спостереження складається з трьох термінів. Попередня обробка даних спостережень відбирає спостереження, які мають однаковий напрям вітру на протязі двох або більше термінів, оскільки суттєвий вплив вітру на хвилі можливий лише за умови тривалої дії вітру в постійному напрямі. Для описання напрямку дії вітру змінна спостереження напрямку (румб “С”, “СВ”, “В”, ...) конвертується в дві незалежні змінні: синус кута напрямку від напрямку на північ (“С”) за годинниковою стрілкою (x^1), та косинус кута напрямку (x^2). В модель також входить швидкість вітру (x^3). Навчальна вибірка за 2006 рік складається зі 100 точок. Для прогнозу відібрані 66 точок.

Оптимальні значення гіперпараметрів $\beta = 10$, $\epsilon = 0.02$. Представлення

моделі у формі Бернштейна:

$$f_{\text{opt}}(\mathbf{x}) = 0.47 + 0.09\phi_0^1(x^1) - 0.09\phi_1^1(x^1) - 0.04\phi_0^1(x^2) + 0.04\phi_1^1(x^2) - \\ - 0.35\phi_0^1(x^3) + 0.35\phi_1^1(x^3).$$

Тепер можна визначити який напрям вітру має найбільший вплив на хвилювання моря в Севастопольській бухті. Для цього треба розв'язати наступну задачу максимізації:

$$\max \leftarrow 0.09\phi_0^1(x^1) - 0.09\phi_1^1(x^1) - 0.04\phi_0^1(x^2) + 0.04\phi_1^1(x^2) = \\ = 0.09(1 - \sin \alpha) - 0.09 \sin \alpha - 0.04(1 - \cos \alpha) + 0.04 \cos \alpha.$$

Легко переконатися, що $\alpha_{\max} = 294^\circ$, або ж приблизно “ЗСЗ” (захід-північ-захід). Це пояснюється тим, що вихід у відкрите море з бухти має північно-західний напрям.

Той самий висновок можна отримати на основі аналізу нечіткої бази знань. Для генерації нечітких правил випишемо нейронечіткі підмоделі в канонічній формі:

$$f_1(x^1) = 0.56\phi_0^1(x^1) + 0.38\phi_1^1(x^1), \\ f_2(x^2) = 0.43\phi_0^1(x^2) + 0.51\phi_1^1(x^2), \\ f_3(x^3) = 0.12\phi_0^1(x^3) + 0.82\phi_1^1(x^3).$$

Функції належності для змінних входу ($x^1 \equiv \sin \alpha$, $x^2 \equiv \cos \alpha$, $x^3 \equiv$ швидкість) визначаються через базисні поліноми Бернштейна першого порядку (ϕ_0^1 , ϕ_1^1). Вони відповідають нечітким змінним: румб “З” та “В”; румб “Ю” та “С”; слабкий та сильний вітер.

Визначимо нечіткі множини для змінної виходу (висота хвиль) через п'ять В-сплайнів першого порядку N_0^1 , N_1^1 , N_2^1 , N_3^1 , N_4^1 (2.3) визначені на вузлах $\{-0.25; 0; 0.25; 0.5; 0.75; 1; 1.25\}$, які відповідають нечітким змінним: штиль, слабке хвилювання, помірне хвилювання, сильне хвилювання, шторм з центрами в точках $\{0; 0.25; 0.5; 0.75; 1\}$. Вони відповідають висоті хвиль від 0.1 метра до 2.4 метра.

Ймовірності правил, що індуковані нейронечіткою моделлю f_1 , можна знайти розв'язавши систему:

$$\begin{cases} 0.56 = 0 \cdot c_{11}^1 + 0.25c_{12}^1 + 0.5c_{13}^1 + 0.75c_{14}^1 + c_{15}^1 \\ 0.38 = 0 \cdot c_{21}^1 + 0.25c_{22}^1 + 0.5c_{23}^1 + 0.75c_{24}^1 + c_{25}^1 \end{cases} \Leftrightarrow \begin{cases} c_{13}^1 = 0.76; & c_{14}^1 = 0.24 \\ c_{22}^1 = 0.48; & c_{23}^1 = 0.52 \end{cases}$$

Для моделі f_2 :

$$\begin{cases} 0.43 = 0 \cdot c_{11}^2 + 0.25c_{12}^2 + 0.5c_{13}^2 + 0.75c_{14}^2 + c_{15}^2 \\ 0.51 = 0 \cdot c_{21}^2 + 0.25c_{22}^2 + 0.5c_{23}^2 + 0.75c_{24}^2 + c_{25}^2 \end{cases} \Leftrightarrow \begin{cases} c_{12}^2 = 0.28; & c_{13}^2 = 0.72 \\ c_{23}^2 = 0.96; & c_{24}^2 = 0.04 \end{cases}$$

Для моделі f_3 :

$$\begin{cases} 0.12 = 0 \cdot c_{11}^3 + 0.25c_{12}^3 + 0.5c_{13}^3 + 0.75c_{14}^3 + c_{15}^3 \\ 0.82 = 0 \cdot c_{21}^3 + 0.25c_{22}^3 + 0.5c_{23}^3 + 0.75c_{24}^3 + c_{25}^3 \end{cases} \Leftrightarrow \begin{cases} c_{11}^3 = 0.52; & c_{12}^3 = 0.48 \\ c_{24}^3 = 0.72; & c_{25}^3 = 0.28 \end{cases}$$

Тож нечіткі правила та їх значимість визначаються наступним чином:

- \hookrightarrow Правило 1: якщо напрям вітру “З” тоді хвилювання помірне (0.76) або сильне (0.24)
- \hookrightarrow Правило 2: якщо напрям вітру “В” тоді хвилювання слабке (0.48) або помірне (0.52)
- \hookrightarrow Правило 3: якщо напрям вітру “Ю” тоді хвилювання слабке (0.28) або помірне (0.72)
- \hookrightarrow Правило 4: якщо напрям вітру “С” тоді хвилювання помірне (0.96) або сильне (0.04)
- \hookrightarrow Правило 5: якщо вітер слабкий тоді штіль (0.52) або слабке хвилювання (0.48)
- \hookrightarrow Правило 6: якщо вітер сильний тоді сильне хвилювання (0.72) або шторм (0.28)

4.5. Апробація ПРИАМ на екологічній моделі біомаси макрозообентосу

Основним компонентом екологічного моніторингу навколишнього середовища є моніторинг стану біосфери. Задача моніторингу полягає в оцінюванні степені уражаючої дії забруднювачів на біосферу, зокрема на гідроекосистеми. Однією з основних ланок гідроекосистем є макрозообентос. Тож відповідно одним з основних показників гідробіологічного моніторингу є щільність біомаси макрозообентосу [93]. За останні тридцять років відділом морської санітарної гідробіології зібрані дані моніторингу акваторії Севастополя. Для моделювання використані дані по біомасі макрозообентосу

(y) і дані по фізико-хімічним характеристикам донного осаду, а саме вміст хлороформ-екстрагованих сполучень (x^1) і натуральна вологість донного осаду (x^2). Наведемо приклад аналізу залежності біомаси макрозообентосу від характеристик донного осаду за 1973 рік. Навчальна вибірка за 1973 рік складається з 25 точок. Для прогнозу відібрані інші 25 точок. Кожній точці відповідає інформація з окремої станції спостереження в акваторії Севастополя.

Оптимальні значення гіперпараметрів $\beta = 10$, $\epsilon = 0.2$. Представлення моделі у формі Бернштейна:

$$f_{\text{opt}}(\mathbf{x}) = 0.17 + 0.06\phi_0^1(x^1) - 0.06\phi_1^1(x^1) - 0.07\phi_0^1(x^2) + 0.07\phi_1^1(x^2).$$

Як добре видно з моделі, спостерігається стійка тенденція зниження біомаси при збільшенні рівня забрудненості донного осаду.

4.6. Висновки до розділу

Розроблена архітектура нової інформаційної системи обробки даних спостережень, та створено її прототип. Зокрема результатом практичних досліджень нейронечіткого підходу до ідентифікації нелінійних стохастичних систем є проект В-В.NF⁸. Це програмна реалізація на мові Java алгоритмів ННМ К.Харріса (швидкий; оптимальний), МГВА (багаторядний; з частковими описами у вигляді поліномів у формі Бернштейна) з метою демонстрації їх можливостей в мережі Інтернет. Для запуску Java-аплету необхідна наявність JRE 1.4⁹. Приклад роботи аплету показано в додатку Б. Для оцінювання якості побудованих моделей використовуються критерії ІКА та ІКМ. Можна бачити відносно високу ефективність швидкої ННМ К.Харріса. Індуктивний метод побудови байєсівської моделі регресії опорних векторів у формі Бернштейна (ПРІАМ) реалізований як в програмному середовищі MATLAB 6.1

⁸В-В.NF — Bezier-Bernstein Neuro Fuzzy modeling project можна знайти за адресою <http://oleg.mytnyk.kiev.ua/nfp/>.

⁹Java Runtime Environment версії не нижче 1.4.

так і у вигляді Java-аплету.

ПРІАМ пройшов апробацію на реальних та штучних еталонних задачах лінійної та нелінійної регресій. Практично для всіх моделей прогноз побудований на основі ПРІАМ за показниками СКП та САП виявився кращим ніж на основі МГВА, НМГВА і РНМ. Прогноз на основі ПРІАМ відповідає рівню і якості прогнозу за допомогою ANFIS. Проте ПРІАМ в порівнянні з ANFIS генерує значно меншу кількість нечітких правил, і ці правила значно простіші і більш зрозумілі для експертів предметної області. ПРІАМ також пройшов апробацію на реальних економічних, метеорологічних та екологічних моделях, де показав один з найкращих результатів прогнозу. Структурованість ПРІАМ дозволяє легко визначати і інтерпретувати індивідуальні впливи факторів. Не має проблем з мультиколінеарністю вхідних даних. Деталі моделювання можна знайти в додатку А. Наведений в табл. 4.1 порівняльний аналіз підтверджує високу ефективність ПРІАМ.

Таблиця 4.1

Порівняння СКП прогнозу для різних методів

Моделі	Longley	Filippelli	Friedman	AMPG	ІСЦ	РСГ	СГМО	ІБПМ
$N_{\text{повна}}(N)$	16(11)	82(33)	1000(500)	392(300)	24(14)	24(14)	166(100)	50(25)
n	6	1	10	4	4	2	3	2
$F_{N-n, N-n}^{(80\%)}$	2.2	1.3	1.0	1.0	1.8	1.7	1.1	1.4
ПРІАМ	0.017	0.004	0.009	0.022	0.003	0.027	0.033	0.076
МГВА	0.051	0.016	0.029	0.024	0.130	0.042	0.025	0.124
НМГВА	0.001	0.039	0.037	0.026	0.008	0.098	0.043	0.053
РНМ	0.018	0.012	0.008	0.028	0.090	0.101	0.033	0.083
ANFIS	0.002	0.001	0.009	0.027	0.003	0.051	0.031	0.110
DW	2.597	1.726	2.106	1.293	1.137	1.586	1.560	1.395
Theil	0.010	0.006	0.089	0.131	0.009	0.031	0.236	0.483

Для перевірки значимості переваги одного методу над іншим необхідно порівняти відношення СКП цих методів з наведеним табличним значенням критерію Фішера [136]. Для визначення наявності чи відсутності автокореляції в залишках навчальної вибірки підрахований критерій Дарбіна-Уотсона:

$$DW = \frac{\sum_j (\delta_j - \delta_{j-1})^2}{\sum_j \delta_j^2} \quad (\text{наведений в таблиці для ПРІАМ}).$$

Зазначимо, що хоча табличні значення інтервалів значимості статистики DW з рівнем 5% в таблиці не наводяться переважно всі отримані результати свідчать про відсутність автокореляції в залишках. Точність прогнозу для різних задач можна оцінити за коефіцієнтом Тейла:

$$\text{Theil} = \frac{\sqrt{\sum_j (y_j - f_j)^2}}{\sqrt{\sum_j y_j^2} + \sqrt{\sum_j f_j^2}} \quad (\text{наведений в таблиці для ПРІАМ}).$$

Складемо таблицю рейтингу методів (табл. 4.2) за десятибальною шкалою. Метод, який показує найкращий результат, тобто найменше СКП, отримує 10 балів. Другий результат — 8 балів. Третій — 6 балів. Четвертий — 4 бали. Найгірший результат — 2 бали. Підрахунок ведеться з врахуванням значимості по критерію Фішера. Якщо два методи мають результат однакової значимості, то вони отримують однакову (середню) кількість балів. Рейтинг методу визначається як середній бал по всім моделям. Слід зазначити, що

Таблиця 4.2

Рейтинг методів

Моделі	Longley	Filippelli	Friedman	AMPG	ІСЦ	РСГ	СГМО	ІБПМ	Рейтинг
ПРІАМ	5	8	7	10	9	10	5	7	8
МГВА	2	4	4	8	2	7	10	3	5
НМГВА	9	2	2	6	6	3	2	10	5
PHM	5	6	10	2	4	3	5	7	5
ANFIS	9	10	7	4	9	7	8	3	7

цей рейтинг не носить абсолютного характеру, а відображає силу методу в конкретній його реалізації. Як можна бачити з таблиці — ПРІАМ має найвищий рейтинг. Відзначимо, що ПРІАМ жодного разу не показав значимо найгірший результат. Це свідчить про його стабільність. В результаті аналізу табл. 4.1 також було визначено, що в середньому точність прогнозу ПРІАМ вища за точність МГВА на 52%, PHM — на 40%, НМГВА — на 26%, ANFIS — на 10% (при порівнянні квадратного кореню середнього СКП).

ЗАГАЛЬНІ ВИСНОВКИ

1. Проведено аналіз методів прогнозування стохастичних процесів та обґрунтована ефективність застосування нечітких баз знань (НБЗ) побудованих на основі нейронечітких технологій. Встановлено, що основним недоліком побудованих НБЗ є їх експонентна складність, а для синтезу НБЗ пониженої складності доцільно використовувати нейронечіткі моделі у формі Бернштейна.
2. Створена нейронечітка інформаційна технологія синтезу НБЗ квадратичної складності на основі збалансованих нейронечітких моделей у формі Бернштейна, де для обґрунтування декомпозиції моделі у формі Бернштейна на незалежні підмоделі використовується принцип максимуму ентропії.
3. Розвинуто прискорений метод побудови нейронечітких моделей у формі Бернштейна, який для визначення барицентричних координат використовує швидке обернене відображення Кастельжо (ОВК). Швидкодія побудови моделей при цьому зросла в 3–4 рази. Для синтезу робастних нейронечітких моделей у формі Бернштейна вперше запропоновано використовувати еволюційний метод на основі оптимального ОVK.
4. Показано, що байєсівська регресія опорних векторів (БРОВ) дозволяє будувати моделі стохастичних процесів з домінуючою випадковою складовою. Головними перевагами БРОВ є робастність оцінок її параметрів, розв'язання задачі опуклого квадратичного програмування, побудова

довірчих інтервалів, здатність прогнозувати на коротких вибірках. Це дозволяє робити прогноз в умовах обмеженої кількості даних при сильній мультиколінеарності і шумі.

5. Для оцінки адекватності моделей БРОВ отримано новий критерій (критерій байєсівського підтвердження — КБП), який, на відміну від відомих критеріїв, не порушує робастності моделей та має просту аналітичну форму. Ефективність критерію підтверджена аналізом залежності похибки прогнозу від значення КБП на модельних прикладах.
6. На основі БРОВ в характеристичному просторі поліноміальних функцій Без'є-Бернштейна розроблено новий індуктивний метод побудови збалансованих робастних нейронечітких моделей у формі Бернштейна (ПРІАМ), який відрізняється від аналогів підвищеною якістю прогнозування стохастичних процесів. Метод враховує апіорну структуру моделі і працює на коротких вибірках. В середньому точність прогнозу ПРІАМ вища за точність МГВА на 52%, РНМ — на 40%, НМГВА — на 26%, ANFIS — на 10%.
7. Розроблена архітектура нової інформаційної системи обробки даних спостережень, яка реалізує запропоновані методи статистичного аналізу, та створено її прототип. Унікальність системи полягає в її здатності оперувати в онлайн режимі. Впровадження прототипу інформаційної системи в Севастопольській гідрометеорологічній обсерваторії дало можливість побудувати ефективні прогнозуючі моделі, які використовуються для підтримки прийняття рішень прогнозистами з метою уточнення прогнозу параметрів хвиль.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bayes T. An essay towards solving a problem in the doctrine of chances // (1763) Reprinted in *Biometrika*. — 1958. — Vol. 45. — Pp. 293–315.
2. Laplace P. S. Memoir on the probability of the causes of events // (1774) Reprinted in *Statistical Science*. — 1986. — Vol. 1, no. 3. — Pp. 364–378.
3. Savage L. J. The foundations of statistics. — (1954) second revised edition. — New York : Dover Publications, 1972. — 310 pp.
4. Elkan C. Boosting and naive bayesian learning : technical report No. CS97-557 / University of California, SanDiego. — 1997. — 11 pp.
5. Терехов С. А. Введение в Байесовы сети // V Всероссийская науч.-техн. конф. “Нейроинформатика-2003” : материалы школы семинара “Современные проблемы нейроинформатики”. — Москва, МИФИ, 2003. — 188 с.
6. Шумский С. А. Байесова регуляризация обучения // IV Всероссийская науч.-техн. конф. “Нейроинформатика-2002” : материалы школы семинара “Современные проблемы нейроинформатики”. — Москва, МИФИ, 2002. — 30 с.
7. Райфа Г. Анализ решений. — М. : Наука, 1977. — 408 с.
8. Jaynes E. T. Prior probabilities // *IEEE Transactions On Systems Science and Cybernetics*. — 1968. — Vol. 4, no. 3. — Pp. 227–241.

9. Современные методы идентификации систем : пер. с англ. / П. Эйкхофф, А. Ванечек, Е. Савараги и др.; под ред. П. Эйкхоффа. — М. : Мир, 1983. — 400 с.
10. Gull S. Bayesian inductive inference and maximum entropy // Maximum Entropy and Bayesian Methods / Ed. by G. J. Erickson, C. R. Smith. — Dordrecht : Kluwer Academic, 1988. — Pp. 53–74.
11. Gull S. Bayesian data analysis: straight-line fitting // Maximum Entropy and Bayesian Methods / Ed. by J. Skilling. — Dordrecht : Kluwer Academic, 1989. — Pp. 511–518.
12. Mackay D. J. C. Bayesian methods for adaptive models : Ph.D. thesis / California Institute of Technology. — Pasadena, 1992. — 98 pp.
13. Mackay D. J. C. A practical Bayesian framework for backprop networks // Neural Computations. — 1992. — no. 4. — Pp. 448–472.
14. Mackay D. J. C. Bayesian non-linear modeling for the prediction competition // Maximum Entropy and Bayesian Methods / Ed. by G. R. Heidebreder. — Dordrecht the Netherlands : Kluwer Academic Publishers, 1996. — Pp. 221–234.
15. Mackay D. J. C. Comparison of approximate methods for handling hyperparameters // Neural Computation. — 1999. — Vol. 11. — Pp. 1035–1068.
16. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. — М. : Наука, 1974. — 415 с.
17. Vapnik V. N. Statistical learning theory. — New York : John Wiley and Sons Inc., 1998. — 736 pp.
18. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М. : Наука, 1979. — 448 с.

19. Алгоритмы и программы восстановления зависимостей / под ред. В. Н. Вапника. — М. : Наука, Главная редакция физико-математической литературы, 1984. — 816 с.
20. Boser B. E., Guyon I. M., Vapnik V. N. A training algorithm for optimal margin classifiers // Proc. of the Annual Conf. on Computational Learning Theory / Ed. by D. Haussler. — Pittsburgh, PA : ACM Press, 1992. — Pp. 144–152.
21. Cortes C., Vapnik V. Support vector networks // Machine Learning. — 1995. — Vol. 20. — Pp. 273–297.
22. Smola A. J. Regression estimation with support vector learning machines : master's thesis / Technische Universität München. — München, 1996. — 78 pp.
23. Vapnik V., Golowich S., Smola A. Support vector method for function approximation, regression estimation, and signal processing // In Advances in Neural Information Processing Systems. — Vol. 9. — Cambridge, MA : MIT Press, 1997. — Pp. 281–287.
24. Gunn S. R. Support vector machines for classification and regression : technical report / Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, University of Southampton. — 1998. — 66 pp.
25. Smola A. J., Schölkopf B. A tutorial on support vector regression // Statistics and Computing. — 2004. — no. 14. — Pp. 199–222.
26. Asymptotically optimal choice of ϵ -loss for support vector machines / A. J. Smola, N. Murata, B. Schölkopf, K.-R. Müller // Proc. of the Int. Conf. on Artificial Neural Networks. — 1998. — Pp. 105–110.

27. New support vector algorithms / B. Schölkopf, A. J. Smola, R. Williamson, P. Bartlett // Neural Computation. — 2000. — Vol. 12. — Pp. 1207–1245.
28. Predicting time series with support vector machines / K.-R. Müller, A. Smola, G. Rätsch et al. // Advances in Kernel Methods - Support Vector Learning. — Cambridge, MA : MIT Press, 1999. — Pp. 243–254.
29. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М. : Наука, 1965. — 656 с.
30. Mackay D. J. C. Introduction to Gaussian processes // Neural Networks and Machine Learning / Ed. by C. M. Bishop. — Berlin : Springer Verlag, 1998. — Pp. 133–165. — Ser. NATO Advanced Study Institute, vol. 168.
31. Williams C. Gaussian processes // Handbook of Brain Theory and Neural Networks / Ed. by M. A. Arbib. — second edition. — MIT Press, 2002. — Pp. 466–470.
32. Seeger M. Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations : Ph.D. thesis / Institute of Adaptive and Neural Computation, University of Edinburgh. — 2003. — 328 pp.
33. Seeger M. Gaussian processes for machine learning // International Journal of Neural Systems. — 2004. — Vol. 14, no. 2. — Pp. 69–106.
34. Kuss M., Rasmussen C. E. Assessing approximate inference for binary gaussian process classification // Journal of Machine Learning Research. — 2005. — no. 6. — Pp. 1679–1704.
35. Kuss M. Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning : Ph.D. thesis / Technische Universität Darmstadt. — 2006.
36. Rasmussen C. E., Williams C. K. I. Gaussian Processes for Machine Learning. — Cambridge, MA : MIT Press, 2006. — 248 pp.

37. Seeger M. Bayesian model selection for support vector machines, gaussian processes and other kernel classifiers // Advances in Neural Information Processing Systems. — Vol. 12. — Cambridge, MA : MIT Press, 1999. — Pp. 603–609.
38. Law M. H., Kwok J. T. Applying the bayesian evidence framework to ν -support vector regression // Proc. of the European Conf. on Machine Learning (ECML). — Freiburg, Germany, 2001. — Pp. 312–323.
39. Tipping M. E. Sparse Bayesian learning and the relevance vector machine // Journal of Machine Learning Research. — 2001. — no. 1. — Pp. 211–244.
40. Chu W., Keerthi S., Ong C. J. Bayesian support vector regression using a unified loss function // IEEE Trans. Neural Networks. — 2004. — Vol. 15, no. 1. — Pp. 29–44.
41. Chen S. Local regularization assisted orthogonal least squares regression // Neurocomputing. — 2006. — Vol. 69, no. 4-6. — Pp. 559–585.
42. Identification of nonlinear systems using generalized kernel models / S. Chen, X. Hong, C. J. Harris, X. Wang // IEEE Transactions on Control Systems Technology. — 2005. — Vol. 13, no. 3. — Pp. 401–411.
43. Мак-Каллок У. С., Питтс В. Логическое исчисление идей, относящихся к нервной деятельности // (1943) Нейрокомпьютер. — 1992. — № 3, 4. — С. 40–53.
44. Хайкин С. Нейронные сети: полный курс : пер. с англ. — 2-е изд. — М. : Издательский дом “Вильямс”, 2006. — 1104 с.
45. Harris C. J., Hong X., Gan Q. Adaptive modelling, estimation and fusion from data: a neurofuzzy approach. — Berlin : Springer, 2002. — 323 pp.
46. Mackay D. J. C. Information theory, inference, and learning algorithms. — Cambridge : Cambridge University Press, 2003. — 640 pp.

47. Малинецкий Г. Г. Математические основы синергетики. Хаос, структуры, вычислительный эксперимент. — 4-ое, суц. перераб. и доп. изд. — М. : КомКнига, 2005. — 312 с.
48. Уотермен Д. Руководство по экспертным системам : пер. с англ. — М. : Мир, 1989. — 388 с.
49. Zadeh L. A. Outline of a new approach to the analysis of complex systems and decision processes // IEEE Transactions on Systems, Man, and Cybernetics. — 1973. — Vol. 3, no. 1. — Pp. 28–44.
50. Mamdani E. H., Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller // International Journal of Man-Machine Studies. — 1975. — Vol. 7, no. 1. — Pp. 1–13.
51. Takagi T., Sugeno M. Fuzzy identification of systems and its applications to modeling and control // IEEE Transactions Systems, Man and Cybernetics. — 1985. — Vol. 15. — Pp. 116–132.
52. Brown M., Harris C. J. Neurofuzzy adaptive modelling and control. — Hemel Hempstead : Prentice Hall, 1994. — 508 pp.
53. Friedman J. H. Multivariate adaptive regression splines (with discussion) // Annals of Statistics. — 1991. — Vol. 19, no. 1. — Pp. 1–141.
54. Roger Jang J. S. ANFIS: Adaptive-network-based fuzzy inference systems // IEEE Trans. on System, Man and Cybernetics. — 1993. — Vol. 23, no. 3. — Pp. 665–685.
55. Kavli T. ASMOD — an algorithm for adaptive spline modelling of observation data // International Journal of Control. — 1993. — Vol. 58, no. 4. — Pp. 947–967.

56. Weyer E., Kavli T. Theoretical properties of the ASMOD algorithm for empirical modelling // International Journal of Control. — 1997. — Vol. 67, no. 5. — Pp. 767–789.
57. Gunn S. R., Brown M. SUPANOVA — a sparse, transparent modelling approach // In Proc. of IEEE Int. Workshop on Neural Networks for Signal Processing. — Madison, Wisconsin, 1999. — Pp. 21–30.
58. Hong X., Harris C. J. Generalized neurofuzzy network modeling algorithms using Bézier-Bernstein polynomial functions and additive decomposition // IEEE Trans. Neural Networks. — 2000. — Vol. 11, no. 4. — Pp. 889–902.
59. Harris C. J., Hong X. Neurofuzzy network model construction using Bézier-Bernstein polynomial functions // IEE Proc. D Control Theory and Applications. — 2000. — Vol. 147, no. 3. — Pp. 337–343.
60. Новицький Д. В. Алгебраїчні методи у теорії нейронної асоціативної пам'яті : автореф. дис... канд. фіз.-мат. наук : 01.05.02 / Ін-т кібернетики ім. В.М.Глушкова НАН України. — К., 2005. — 20 с.
61. Мітюшкін Ю. І. Синтез і настройка баз нечітких знань для моделювання багатовимірних залежностей : автореф. дис... канд. техн. наук : 01.05.02 / Вінниц. держ. техн. ун-т. — Вінниця, 2002. — 19 с.
62. Колодяжний В. В. Керування стохастичними об'єктами в умовах невизначеності на основі нечітких нейромережових моделей : автореф. дис... канд. техн. наук : 05.13.03 / Харк. нац. ун-т радіоелектрон. — Х., 2002. — 20 с.
63. Тазетдінов В. А. Нейромережева технологія аналізу та прогнозування процесів на ринку нерухомості : автореф. дис... канд. техн. наук : 05.13.06 / Черкас. держ. технол. ун-т. — Черкаси, 2005. — 21 с.

64. Трипольська В. В. Автоматизована ідентифікація підписів як бінарних зображень у просторі ознак перетворення Радона на основі нейромережевого підходу : автореф. дис... канд. техн. наук : 05.13.06 / Херсон. держ. техн. ун-т. — Х., 2003. — 17 с.
65. Сетлак Г. Інтелектуальні системи підтримки прийняття рішень в управлінні виробництвом у нечітких умовах : автореф. дис... д-ра. техн. наук : 05.13.06 / Нац. техн. ун-т України “Київ. політехн. ін-т”. — К., 2005. — 36 с.
66. Побережник С. М. Автоматизація процесів відтворення функціональних залежностей в системах інформаційної підтримки прийняття рішень : автореф. дис... канд. техн. наук : 05.13.06 / Одес. нац. політехн. ун-т. — О., 2003. — 19 с.
67. Чуи Ч. Введение в вейвлеты : пер. с англ. — М. : Мир, 2001. — 412 с.
68. Кирушев В. А., Малоземов В. Н., Певный А. Б. Вейвлетное разложение пространства дискретных периодических сплайнов // Матем. заметки. — 2000. — Т. 67, № 5. — С. 712–720.
69. Астафьева Н. М. Вейвлет анализ: основы теории и примеры применения // Успехи физических наук. — 1996. — Т. 166, № 11. — С. 1145–1170.
70. Дремин И. М., Иванов О. В., Нечитайло В. А. Вейвлеты и их использование // Успехи физических наук. — 2001. — Т. 171, № 5. — С. 465–501.
71. Новиков И. Я., Стечкин С. Б. Основы теории всплесков // Успехи математических наук. — 1998. — № 6. — С. 53–128.
72. Смоленцев Н. К. Основы теории вейвлетов. Вейвлеты в MATLAB. — М. : ДМК Пресс, 2005. — 304 с.
73. Li J. Image compression — the mechanics of the JPEG 2000 / Microsoft Research, Signal Processing, One Microsoft Way, Bld. 113/3374, Redmond,

WA 98052. — 2002.

http://research.microsoft.com/~jinl/paper_2002/msri_jpeg.htm.

74. Kitada Y. Identification of nonlinear structural dynamic systems using wavelets // Journal of Engineering Mechanics. — 1998. — Vol. 124, no. 10. — Pp. 1059–1066.
75. Prazenica R. J., Kurdila A. J. Volterra kernel identification using triangular wavelets // Journal of Vibration and Control. — 2004. — no. 10. — Pp. 597–622.
76. Hasiewicz Z., Pawlak M. Nonlinear system identification via wavelet expansions // IFAC System Identification. — Santa Barbara, California, USA, 2000. — 6 pp.
77. Thuillard M. A review of wavelet networks, wavenets, fuzzy wavenets and their applications // Advances in Computational Intelligence and Learning: Methods and Applications. — London : Springer, 2002. — Pp. 43–60.
78. Thuillard M. Fuzzy logic in the wavelet framework // Proc. Toolmet'2000 — Tool Environments and Development Methods for Intelligent Systems. — 2000. — Pp. 15–36.
79. Popoola A., Ahmad S., Ahmad K. A fuzzy-wavelet method for analyzing non-stationary time series // Proc. of The 5th Int. Conf. on Recent Advances in Soft Computing. — Nottingham, UK, 2004. — 6 pp.
80. Митник О. Ю. Поліноміальні функції Без'є-Бернштейна в нечітких нейронних моделях // Системний аналіз та інформаційні технології : тези V міжнар. наук.-практ. конф. студентів, аспірантів та молодих вчених, 1–3 лип. 2003 р. — К. : НТУУ “КПІ”, 2003. — С. 80–81.
81. Митник О. Ю. Еволюційний метод побудови оптимального оберненого відображення Кастельжо // Системний аналіз та інформаційні техноло-

- гії : тези VI міжнар. наук.-практ. конф. студентів, аспірантів та молодих вчених, 1–3 лип. 2004 р. — К. : НТУУ “КПІ”, 2004. — С. 66–68.
82. Оцінювання хлорофілу в рослинності на основі спектрального аналізу / П. І. Бідюк, І. В. Баклан, О. Ю. Митник, В. І. Литвиненко // Стан та перспективи розвитку новітніх науково-освітніх комп’ютерних технологій : матеріали наук.-практ. конф., 14–15 лист. 2003 р. — Миколаїв, 2003. — С. 62–65.
 83. Митник О. Ю., Бідюк П. І. Нейронечіткий підхід до ідентифікації нелінійних динамічних систем // Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій : матеріали міжнар. наук. конф., 18–21 трав. 2005 р., Євпаторія. Т. 5. — Херсон : Вид-во Херсонського морського ін-ту, 2005. — С. 58–60.
 84. Митник О. Ю. Нейронечіткий підхід до ідентифікації нелінійних динамічних систем // Моделювання та дослідження стійкості динамічних систем : тези міжнар. конф., 23–25 трав. 2005 р. — К. : Вісник КНУ, 2005. — С. 204.
 85. Митник О. Ю. Швидкий метод наближення маргінальної правдоподібності для оцінки адекватності моделей нелінійних стохастичних процесів // Моделювання та дослідження стійкості динамічних систем : тези міжнар. конф., 22–25 трав. 2007 р. — К. : Вісник КНУ, 2007. — С. 70.
 86. Митник О. Ю. Локальное сглаживание ϵ -нечувствительной функции потерь в байесовской модели регрессии опорных векторов // Интеллектуальный анализ информации : сборник трудов VI междунар. науч. конф., 16–19 мая, 2006 г. — К., 2006. — С. 189–198.
 87. Митник О. Ю. Байєсівський умовивід регресії опорних векторів // матеріали XI міжнар. наук. конф. імені академіка М. Кравчука, 18–20 трав. 2006 р. — К. : ТОВ “Задруга”, 2006. — С. 729.

88. Митник О. Ю., Бідюк П. І. Обернене відображення Кастельжо в нечітких нейронних моделях // Системні дослідження та інформаційні технології. — 2004. — № 2. — С. 24–34.
89. Бідюк П. І., Митник О. Ю. Застосування генетичного алгоритму в задачі оцінювання вмісту хлорофілу в рослинності // Наукові вісті НТУУ “КПІ”. — 2004. — № 4. — С. 65–70.
90. Бідюк П. І., Литвиненко В. І., Митник О. Ю. Порівняльний аналіз двох методів дистанційного оцінювання стану рослинності // Інформаційні технології і системи. — 2004. — Т. 7, № 1. — С. 108–116.
91. Митник О. Ю. Регуляризація поліноміальних функцій Без’є-Бернштейна і її застосування в задачі ідентифікації нелінійних систем // Системні дослідження та інформаційні технології. — 2006. — № 1. — С. 94–105.
92. Мытник О. Ю. Построение байесовской регрессии опорных векторов в характеристическом пространстве полиномиальных функций Безье-Бернштейна // Кибернетика и системный анализ. — 2007. — № 4. — С. 179–188.
93. Мытник О. Ю., Алёмов С. В. Использование байесовской модели регрессии опорных векторов для анализа данных гидробиологического мониторинга // Экология моря. — 2007. — Т. 73. — С. 64–69.
94. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем. — К. : Наукова думка, 1982. — 296 с.
95. Ивахненко А. Г., Степашко В. С. Помехоустойчивость моделирования. — К. : Наукова думка, 1985. — 216 с.
96. Зайченко Ю. П., Заєць І. О. Синтез та адаптація нечітких прогнозуючих моделей на основі методу самоорганізації // Наукові вісті НТУУ “КПІ”. — 2001. — № 2. — С. 32–41.

97. Горский В. Г., Адлер Ю. П., Талалай А. М. Планирование промышленных экспериментов (модели динамики). — М. : Металлургия, 1978. — 112 с.
98. Теория систем. Математические методы и моделирование. Сборник статей : пер. с англ. — М. : Мир, 1989. — 384 с.
99. Безручко Б. П., Смирнов Д. А. Математическое моделирование и хаотические временные ряды. — Саратов : ГосУНЦ “Колледж”, 2005. — 320 с.
100. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов : пер. с англ. — М. : Мир, 1988. — 488 с.
101. Усов А. В., Дубров О. М., Дмитришин Д. В. Моделирование систем с распределёнными параметрами. — Одесса : Астропринт, 2002. — 664 с.
102. Leontaritis I. J., Billings S. A. Input-output parametric models for nonlinear systems – part 1: deterministic nonlinear systems; part 2: stochastic nonlinear systems // International Journal of Control. — 1985. — Vol. 41, no. 1. — Pp. 303–344.
103. Foley D. C., Sadegh N. Modelling of nonlinear systems from input-output data for state space realization // Proc. of the 40th IEEE Conf. on Decision and Control. — Orlando, Florida USA, 2001. — Pp. 2980–2985.
104. Rivals I., Personnaz L. Black-box modeling with state-space neural networks // Neural Adaptive Control Technology / Ed. by R. Zbikowski, K. J. Hunt. — World Scientific, 1996. — Pp. 237–264. — World Scientific Series in Robotics and Intelligent Systems, vol. 15.
105. Verdult V. Nonlinear system identification: a state-space approach : Ph.D. thesis / University of Twente. — 2002. — 292 pp.
106. Guo F. A new identification method for Wiener and Hammerstein systems : Ph.D. thesis / Institut für Angewandte Informatik. — 2004. — 98 pp.

107. Geman S., Bienenstock E., Doursat R. Neural networks and the bias/variance dilemma // Neural Computations. — 1992. — no. 4. — Pp. 1–58.
108. Vapnik V. Principles of risk minimization for learning theory // In Advances in Neural Information Processing Systems. — Vol. 4. — Cambridge, MA : MIT Press, 1992. — Pp. 831–838.
109. Bousquet O., Boucheron S., Lugosi G. Introduction to statistical learning theory // Advanced Lectures on Machine Learning. — 2003. — Pp. 169–207.
110. Балакришнан А. В. Теория фильтрации Кальмана. — М. : Мир, 1988. — 168 с.
111. Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики. — 2004. — № 13. — С. 5–36.
112. Воронцов К. В. Комбинаторные оценки качества обучения по прецедентам // Докл. РАН. — 2004. — Т. 394, № 2. — С. 175–178.
113. Воронцов К. В. Комбинаторные обоснования обучаемых алгоритмов // ЖВ-МиМФ. — 2004. — Т. 44, № 11. — С. 2099–2112.
114. Handbook of discrete and computational geometry / Ed. by J. E. Goodman, J. O'Rourke. — New York : CRC-Press, 1997. — 1008 pp.
115. Wikipedia contributors. VC dimension // Wikipedia, The Free Encyclopedia. Retrieved 17:48, January 10, 2006.
http://en.wikipedia.org/w/index.php?title=VC_dimension&oldid=34381014.
116. Тихонов А. Н., Арсенин В. Я., Тимонов А. А. Математические задачи компьютерной томографии. — М. : Наука. Гл. ред. физ.-мат. лит., 1987. — 160 с. — Проблемы науки и технического прогресса.
117. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. — М. : Наука, 1986. — 288 pp.

118. Bossley K. M. Regularisation theory applied to neurofuzzy modelling : technical report ISIS-TR3 / University of Southampton, Dept of ECS. — 1995. — 64 pp.
119. Панова Н. В. Методи розв'язку некоректних задач ідентифікації дискретних динамічних систем : автореф. дис... канд. фіз.-мат. наук : 01.05.04 / Ін-т косм. дослідж. НАН та НКА України. — К., 2002. — 20 с.
120. Остропицький В. М. Розробка інформаційної технології нелінійної робастної сплайн-обробки даних в системі моніторингу : автореф. дис... канд. техн. наук : 05.13.06 / Київ. міжнар. ун-т цив. авіації. — К., 1999. — 17 с.
121. Василенко В. А. Сплайн функции: теория, алгоритмы, программы. — Новосибирск : Наука, 1983. — 210 с.
122. Енюков И. С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА. — М. : Финансы и статистика, 1986. — 232 с.
123. Згуровский М. З., Бидюк П. И. Анализ и управление большими космическими конструкциями. — К. : Наукова Думка, 1997. — 452 с.
124. Moody J. E. The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems // In Advances in Neural Information Processing Systems. — Vol. 4. — Cambridge, MA : MIT Press, 1992. — Pp. 847–854.
125. Neal R. M. Monte Carlo implementation of Gaussian process models for Bayesian classification and regression : technical report 9702 / University of Toronto, Department of Statistics. — 1997. — 24 pp.
126. Minka T. P. A family of algorithms for approximate Bayesian inference : Ph.D. thesis / Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science. — 2001. — 75 pp.

127. Городецкий А. Я. Информационные системы. Вероятностные модели и статистические решения. — СПб. : Изд-во СПбГПУ, 2003. — 326 с.
128. Воронин А. Н. Многокритериальный синтез динамических систем. — К. : Наукова думка, 1992. — 160 с.
129. Jaynes E. T. Probability Theory: The Logic of Science. — Cambridge : Cambridge University Press, 2003. — 758 pp.
130. Закс Ш. Теория статистических выводов. — М. : Мир, 1975. — 776 с.
131. Де Гроот М. Оптимальные статистические решения. — М. : Мир, 1974. — 491 с.
132. Киричков В. Н. Идентификация объектов систем управления технологическими процессами / Автоматика и управление в технических системах. — К. : Выща шк., 1990. — Т. 2. — 263 с.
133. Purser R. J., Parrish D. F. A Bayesian technique for estimating continuously varying statistical parameters of a variation assimilation // Meteorology and Atmospheric Physics. — 2003. — Vol. 82. — Pp. 209–226.
134. Mackay D. J. C. Bayesian model comparison and backprop nets // In Advances in Neural Information Processing Systems. — Vol. 4. — Cambridge, MA : MIT Press, 1992. — Pp. 839–846.
135. Бендат Д., Пирсол А. Измерение и анализ случайных процессов. — М. : Мир, 1974. — 463 с.
136. Пустыльник Е. И. Статистические методы анализа и обработки наблюдений. — М. : Наука, 1968. — 288 с.
137. Информационные системы и технологии в экономике / Т. П. Барановская, В. И. Лойко, М. И. Семенов, А. И. Трубилин. — М. : Финансы и статистика, 2003. — 418 с.

138. Шеннон К. Работы по теории информации и кибернетике : пер. с англ. / под ред. Р. Л. Добрушина, О. Б. Лупанова. — М. : Из-во иностр. лит., 1963. — 829 с.
139. Кульбак С. Теория информации и статистики. — М. : Наука, 1967. — 408 с.
140. Akaike H. A new look at the statistical model identification // IEEE Trans. Automat. Contr. — 1974. — no. 19. — Pp. 716–723.
141. Schwarz G. Estimating the dimension of a model // Annals of Statistics. — 1978. — Vol. 6. — Pp. 461–464.
142. Rissanen J. Modeling by shortest data description // Automatica. — 1978. — Vol. 14. — Pp. 465–471.
143. Bossley K. M. Neurofuzzy construction algorithms : transfer report / University of Southampton, Dept of ECS. — 1996. — 59 pp.
144. Murata N., Yoshizawa S., Amari S. Network information criterion determining the number of hidden units for an artificial neural network model // IEEE Transactions on Neural Networks. — 1994. — Vol. 5, no. 6. — Pp. 865–872.
145. Golub G. H., Heath M., Wahba G. Generalized cross-validation as a method for choosing good ridge parameter // Technometrics. — 1979. — Vol. 21, no. 2. — Pp. 215–223.
146. Зайченко Ю. П. Основи проектування інтелектуальних систем. — К. : Видавничий дім “Слово”, 2004. — 352 с.
147. Вознесенский В. А., Ковальчук А. Ф. Принятие решений по статистическим моделям. — М. : Статистика, 1978. — 192 с.
148. Вольтерра В. Теория функционалов, интегральных и интегро-дифференциальных уравнений. — М. : Наука, 1982. — 304 с.

149. Винер Н. Нелинейные задачи в теории случайных процессов. — М. : Изво иностр. лит., 1961. — 128 с.
150. Ли Ю., Щецен М. Определение ядер Винера-Хопфа методом взаимной корреляции // Техническая кибернетика за рубежом. — М. : Машиностроение, 1968. — С. 166–185.
151. Щецен М. Моделирование нелинейных систем на основе теории Винера // ТИИЭР. — 1981. — Т. 69, № 12. — С. 44–62.
152. Maas S. A. Analysis and optimization of nonlinear microwave circuits by Volterra series // Microwave J. — 1990. — Vol. 33, no. 4. — P. 245.
153. Пинчук В. М. Идентификация динамических объектов на основе итерационного построения ряда Вольтерра : дис... канд. техн. наук : 05.13.01 / КПИ. — К., 1985. — 169 с.
154. Газі А. Методи і засоби комп'ютерної ідентифікації динамічних об'єктів на основі інтегральних моделей : автореф. дис... канд. техн. наук : 05.13.06 / Ін-т пробл. моделювання в енергетиці ім. Г.Є.Пухова НАН України. — К., 2005. — 20 с.
155. Иванов А. И. Идентификация нелинейных динамических объектов с параллельно-рекурсивной структурой. — Челябинск : ЧПИ, 1985. — 21 с. — Рукопись деп. в ВИНТИ №8616-В от 12.12.85.
156. Иванов А. И. Быстрое обучение искусственных нейронных сетей в системах биометрической аутентификации личности : дис... докт. техн. наук : 05.13.01 / ПГУ. — Пенза, 2000. — 350 с.
157. Круг Г. К., Сосулин Ю. А., Фатуев В. А. Планирование эксперимента в задачах идентификации и экстраполяции. — М. : Наука, 1977. — 207 с.

158. Ланцош К. Практические методы прикладного анализа. Справочное руководство: Перевод с англ. М.З. Кайнера / под ред. А. М. Лопшица. — М. : Физматгиз, 1961. — 524 с.
159. Ivakhnenko A. G. Heuristic self-organization in problems of engineering cybernetics // Automatica. — 1970. — Vol. 6, no. 2. — Pp. 207–219.
160. Глибовець М. М., Олецький О. В. Штучний інтелект. — К. : Вид. дім “КМ Академія”, 2002. — 366 с.
161. Коршевніук Л. А., Бидюк П. И. Решение задачи распределения инвестиций на основе нечеткого логического вывода // Системні дослідження та інформаційні технології. — 2003. — № 2. — С. 34–42.
162. Roger Jang J. S., Sun C. T. Functional equivalence between radial basis function networks and fuzzy inference systems // IEEE Transactions on Neural Networks. — 1993. — Vol. 4, no. 1. — Pp. 156–159.
163. Castellano G. A neurofuzzy methodology for predictive modeling : Ph.D. thesis / Department of Computer Science, University of Bari. — Bari, Italy, 2001. — 224 pp.
164. Hartmut P., Wolfgang B., Marco P. Bézier- and B-spline techniques. — Berlin; New York : Springer, 2002. — 290 pp.
165. О’Квин Д. Допечатная подготовка. Руководство дизайнера : пер. с англ. — М. : Издательский дом “Вильямс”, 2002. — 592 с.
166. Harris C. J., Hong X. Data based constructive identification — overcoming the curse of dimensionality : ISIS report / Dept of Electronic and Computer Science, University of Southampton. — 2000. — 12 pp.
167. Chen S., Billings S. A., Luo W. Orthogonal least squares methods and their application to non-linear system identification // International Journal of Control. — 1989. — Vol. 50, no. 5. — Pp. 1873–1896.

168. Wang L. X., Mendel J. M. Fuzzy basis function, universal approximation, and orthogonal least squares learning // IEEE Transactions on Neural Networks. — 1992. — Vol. 3, no. 5. — Pp. 807–814.
169. Chen S., Cowan C. F. N., Grant P. M. Orthogonal least squares learning algorithm for radial basis function networks // IEEE Transactions on Neural Networks. — 1991. — Vol. 2, no. 2. — Pp. 302–309.
170. Денискин Ю. И. Геометрическое моделирование криволинейных объектов с использованием барицентрических координат // МАИ. Электронный журнал “Прикладная геометрия”. — 1999. — Т. 1, № 1. — С. 1–20.
http://www.mai.ru/~apg/Volume1/v1_n1.htm.
171. Малютов М. Б., Заиграев А. Ю. Современные задачи оптимального планирования регрессионных экспериментов. — К. : Выща шк. Головное изд-во, 1989. — 64 с. — Соврем. достижения математики и её приложений.
172. Hardin R. H., Sloane N. J. A. A new approach to the construction of optimal designs // Journal of statistical planning and inference. — 1993. — Vol. 37. — Pp. 339–369.
173. Chen S., Hong X., Harris C. J. Sparse kernel regression modeling using combined locally regularized orthogonal least squares and D-optimality experimental design // IEEE Transactions on Automatic Control. — 2003. — Vol. 48, no. 6. — Pp. 1029–1036.
174. Genetic algorithms can improve the construction of D-optimal experimental designs / J. Poland, A. Mitterer, K. Knödler, A. Zell // Advances In Fuzzy Systems and Evolutionary Computation / Ed. by N. Mastorakis. — 2001. — Pp. 227–231.
175. Паклин Н. Б. Адаптивные модели нечеткого вывода для идентификации нелинейных зависимостей в сложных системах : дис... канд. техн. наук : 05.13.18 / ИЖГТУ. — Ижевск, 2004. — 162 с.

176. Rogers A., Prugel-Bennett A. Genetic drift in genetic algorithm selection schemes // IEEE Transactions on Evolutionary Computation. — 1999. — Vol. 3, no. 4. — Pp. 298–303.
177. Бидюк П. И., Литвиненко В. И., Токарь А. А. Параллельные генетические алгоритмы // Системні дослідження та інформаційні технології. — 2002. — № 4. — С. 7–16.
178. Eiben A. E., Aarts E. H. L., Van Hee K. M. Global convergence of genetic algorithms: a markov chain analysis // Parallel Problem Solving from Nature. — Berlin : Springer, 1991. — Pp. 4–12.
179. Феллер В. Введение в теорию вероятностей и её приложения: В 2 т. — М. : Мир, 1984. — Т. 1. — 528 с.
180. Vapnik V. N. The nature of statistical learning theory. — New York : Springer, 1999. — 304 pp.
181. Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. — М. : Наука, 1970. — 383 с.
182. Chu W., Sathya Keerthi S. New approaches to support vector ordinal regression // Proc. of the 22nd Int. Conf. on Machine Learning. — Bonn, Germany, 2005. — 8 pp.
183. Smola A. J., Schölkopf B. From regularization operators to support vector kernels // In Advances in Neural Information Processing Systems. — Vol. 10. — Cambridge, MA : MIT Press, 1998. — Pp. 343–349.
184. Козлов М. К., Тарасов С. П., Хачиян Л. Г. Полиномиальная разрешимость выпуклого квадратичного программирования // ДАН СССР. — 1979. — Т. 248, № 5. — С. 1049–1051.

185. Procedures for optimization problems with a mixture of bounds and general linear constraints / P. E. Gill, W. Murray, M. A. Saunders, M. H. Wright // ACM Trans. Math. Software. — 1984. — Vol. 10. — Pp. 282–298.
186. Musicant D. R., Feinberg A. Active set support vector regression // IEEE Transactions on Neural Networks. — 2004. — Vol. 15, no. 2. — Pp. 268–275.
187. Girosi F. An equivalence between sparse approximation and Support Vector Machines / MIT Artificial Intelligence Laboratory. — A.I. Memo 1606. — 1997. — 25 pp.
188. Platt J. Fast training of support vector machines using sequential minimal optimization // Advances in Kernel Methods - Support Vector Learning / Ed. by B. Schölkopf, C. J. C. Burges, A. J. Smola. — Cambridge, MA : MIT Press, 1999. — Pp. 185–208.
189. Hornstein A., Parlitz U. Bias reduction for time series models based on support vector regression // International Journal of Bifurcation and Chaos. — 2004. — Vol. 14, no. 6. — Pp. 1947–1956.
190. Song X., Zhou T., Zhang H. Support vector regression estimation based on non-uniform lost function // Proc. of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conf. — Shanghai, China, 2005. — 4 pp.
191. Altun Y., Tsochantaridis I., Hofmann T. Hidden Markov support vector machines // Proc. of the 20th Int. Conf. on Machine Learning. — Washington DC, 2003. — 8 pp.
192. Comparing support vector machines with gaussian kernels to radial basis function classifiers / B. Schölkopf, K. Sung, C. Burges et al. // IEEE Trans. Sign. Processing. — 1997. — no. 45. — Pp. 2758–2765.
193. Nonlinear estimation and modeling of fMRI data using spatio-temporal support vector regression / Y. Wang, R. T. Schultz, R. T. Constable,

- L. H. Staib // Information Processing in Medical Imaging Proceedings. — 2003. — Pp. 647–659.
194. Yang H., Chan L., King I. Support vector machine regression for volatile stock market prediction // IDEAL, LNCS 2412. — 2002. — Pp. 391–396.
 195. Tsakonas A., Dounias G., Shtovba S. Forecasting football match outcomes with support vector machines // In Proc of Int. Conf. Computer-Based Information Technology. — Zhitomir, Ukraine, 2002. — 5 pp.
 196. Sollich P. Bayesian methods for support vector machines: Evidence and predictive class probabilities // Machine Learning. — 2002. — Vol. 46, no. 1-3. — Pp. 21–52.
 197. Coleman T. F., Li Y. On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds // Mathematical Programming. — 1994. — Vol. 67, no. 2. — Pp. 189–224.
 198. Оцінювання ієрархічних моделей із застосуванням методу Монте-Карло / П. І. Бідюк, В. І. Литвиненко, І. В. Баклан, Я. І. Баклан // Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій : матеріали міжнар. наук. конф., Євпаторія. Т. 1. — Херсон : Вид-во Херсонського морського ін-ту, 2005. — С. 41–46.
 199. Mytnik O. Y. Construction of Bayesian support vector regression in the feature space spanned by Bezier-Bernstein polynomial functions // Cybernetics and Sys. Anal. — 2007. — Vol. 43, no. 4. — Pp. 613–620.
 200. A probabilistic framework for SVM regression and error bar estimation / J. B. Gao, S. R. Gunn, C. J. Harris, M. Brown // Machine Learning. — 2002. — Vol. 46, no. 1-3. — Pp. 71–89.
 201. UCI Repository of machine learning databases / D. J. Newman, S. Hettich, C. L. Blake, C. J. Merz; Department of Information and Computer Science,

University of California, Irvine, CA. — 1998.

<http://www.ics.uci.edu/~mllearn/MLRepository.html>.

202. Longley J. W. An appraisal of least squares programs for the electronic computer from the viewpoint of the user // Journal of the American Statistical Association. — 1967. — no. 62. — Pp. 819–841.
203. Соколова Г. Н. Информационные технологии экономического анализа. — М. : Экзамен, 2002. — 320 с.
204. Sultan K., Lukyanenko I., Gorodnicheko Y. The Methodological aspects of development and application of macroeconometric models: The Case of Ukraine // Journal of the American Statistical Association. — Kyiv : NaUKMA Publishing House, 2000.
205. Gorodnichenko Y. Effects of intergovernmental aid on fiscal behavior of local governments: the case of Ukraine : EERC MA thesis / NaUKMA. — 2001. — 101 pp.
206. Kosse V. Interest rates and thier role in the economy during transition. The problem of high interest rates. Case of Ukraine : EERC MA thesis / NaUKMA. — 2002. — 90 pp.
207. Кордзадзе Т. З. Коінтеграційні процеси та їх застосування в економетричному аналізі // Системні дослідження та інформаційні технології. — 2003. — № 1. — С. 97–111.
208. Исследование нечётких нейронных сетей в задачах макроэкономического прогнозирования / Ю. П. Зайченко, Ф. Севае, К. М. Титаренко, Н. В. Титаренко // Системні дослідження та інформаційні технології. — 2004. — № 2. — С. 70–86.
209. Зайченко Ю. П., Севае Ф., Келестин Ю. В. Сравнительный анализ эффективности нечётких нейронных сетей в задачах прогнозирования в

економике и финансовой сфере // Системні дослідження та інформаційні технології. — 2006. — № 1. — С. 56–70.

210. Дербенцев В. Д. Аналіз та моделювання інфляційних процесів в економіці України : автореф. дис... канд. екон. наук : 08.03.02 / Київ. нац. екон. ун-т. — К., 2001. — 19 с.
211. Половцев О. В. Аналіз, економетричне моделювання та прийняття оптимальних рішень в умовах невизначеності : автореф. дис... канд. техн. наук : 05.13.06 / Херсон. держ. техн. ун-т. — Х., 1999. — 17 с.
212. Демківська Т. І. Методи та засоби прогнозування екологічних та економічних процесів на основі комп'ютерного моделювання часових рядів : автореф. дис... канд. техн. наук : 05.13.06 / Ін-т пробл. моделювання в енергетиці ім. Г.Є.Пухова НАН України. — К., 2001. — 19 с.

ДОДАТОК А

РЕЗУЛЬТАТИ ПОБУДОВИ ПРОГНОЗІВ ЗА ДОПОМОГОЮ ПРИАМ

Модель Лонглі

Таблиця А.1

Повний розв'язок оптимальної моделі Лонглі

№	1	2	3	4	5	6	7	8	9	10	11
α	11.2	0	20.4	0	0	0	0	0	0	48.8	0
α^*	0	0	0	38.5	0	12	0	0	0	0	29.8
b	0.56										

№	КБП	$\tilde{\epsilon}$	β	N_{sv}	\tilde{R}_{emp}	\tilde{R}_{reg}	СКП _н	САП _н
1	-19.70	0.030	20.0	5	0.0009	0.4616	0.0169	0.1191
2	-22.54	0.027	30.0	5	0.0013	0.7462	0.0121	0.0979
3	-25.35	0.023	48.8	6	0.0021	1.4613	0.0141	0.1039
4	-25.49	0.022	48.8	6	0.0024	1.6597	0.0160	0.1101
5	-25.56	0.021	48.8	6	0.0026	1.7645	0.0170	0.1132
ϵ							СКП	САП
219.942							ПРИАМ	1833920 1174.86
							МГВА	5468339 2068.64
							PHM	1892395 1182.36
							НМГВА	137580 324.91

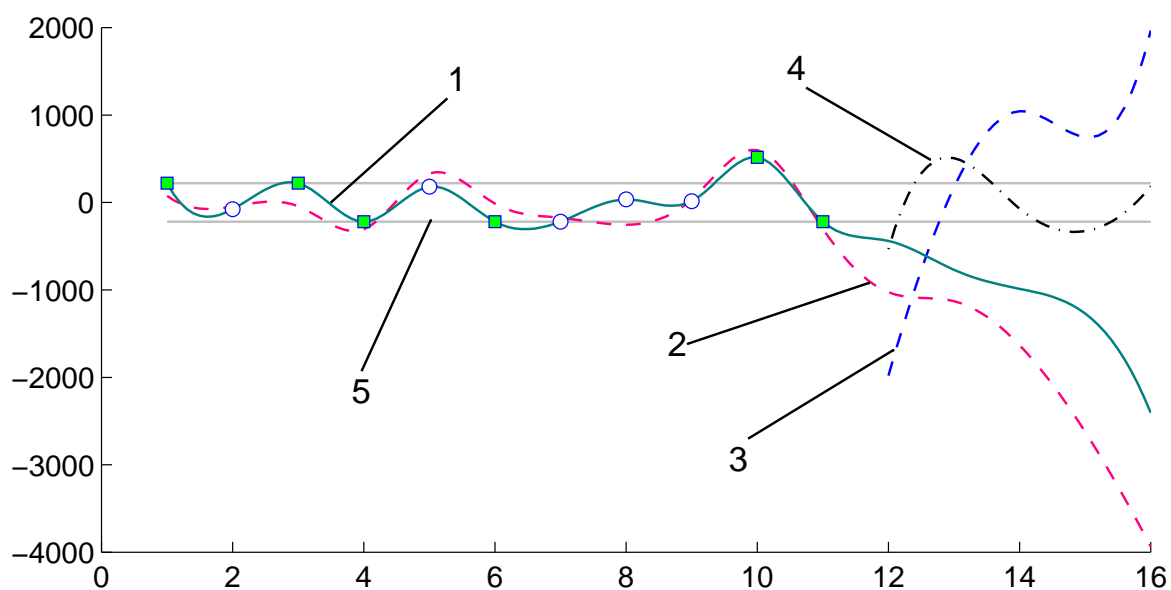


Рис. А.1. Залишки для моделі Лонглі: 1 — ПРІАМ, 2 — МГВА, 3 — РНМ, 4 — НМГВА, 5 — смуга нечутливості.

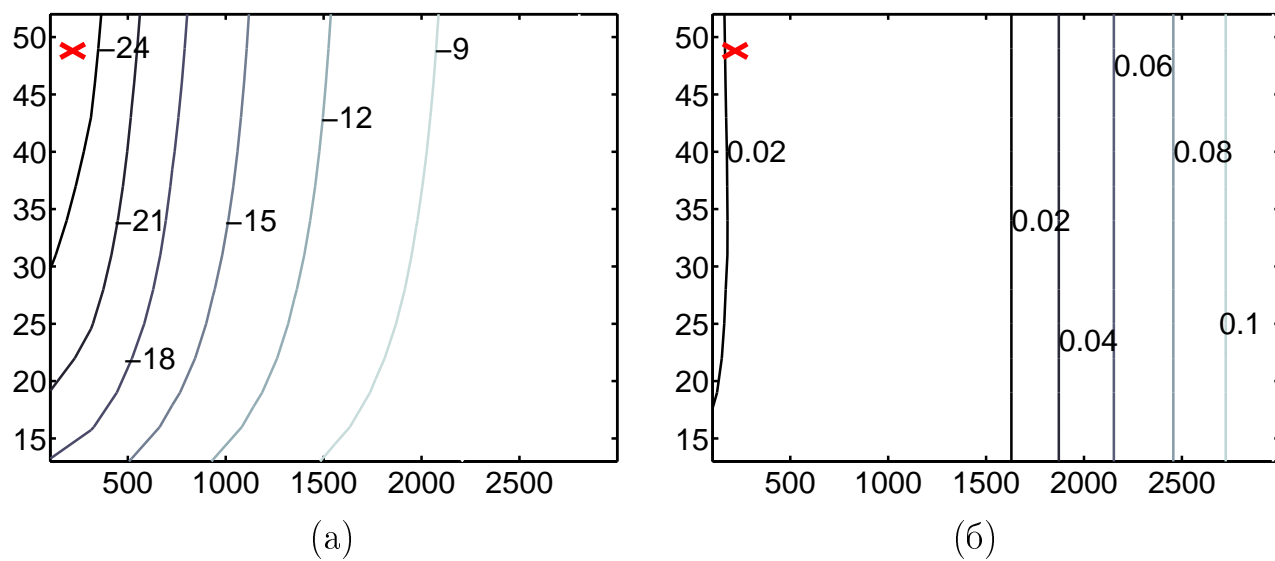


Рис. А.2. Залежність (а) КБП і (б) СКПн моделі Лонглі від ширини смуги нечутливості ϵ для різних значень гіперпараметру β .

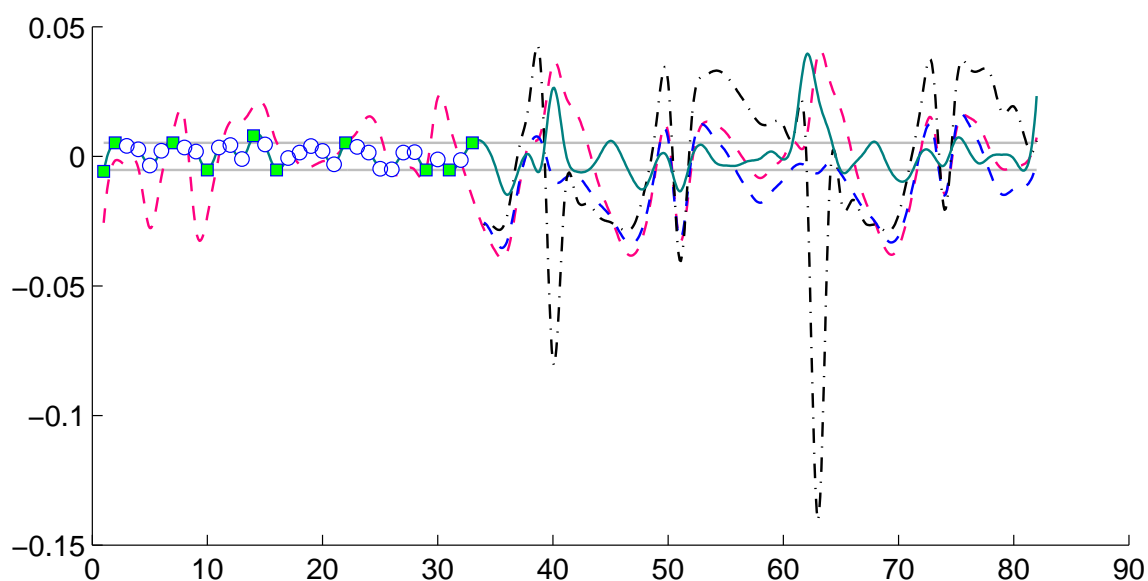


Рис. А.3. Залишки для моделі Філіппеллі.

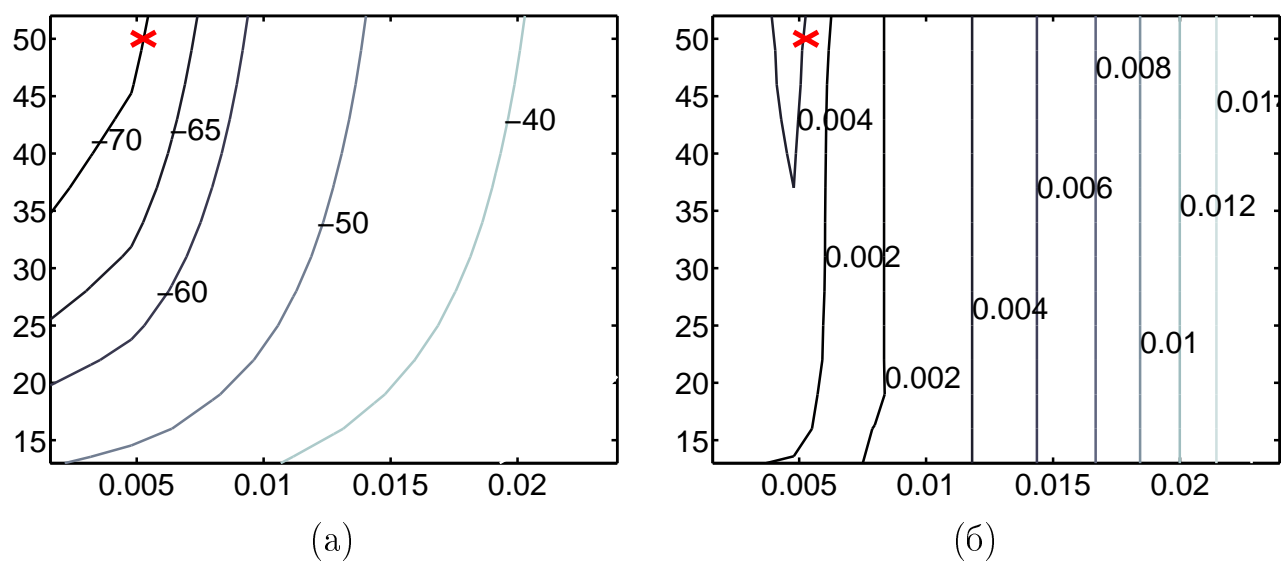


Рис. А.4. Залежність (а) КБП і (б) СКПн моделі Філіппеллі від ширини смуги нечутливості ϵ для різних значень гіперпараметру β .

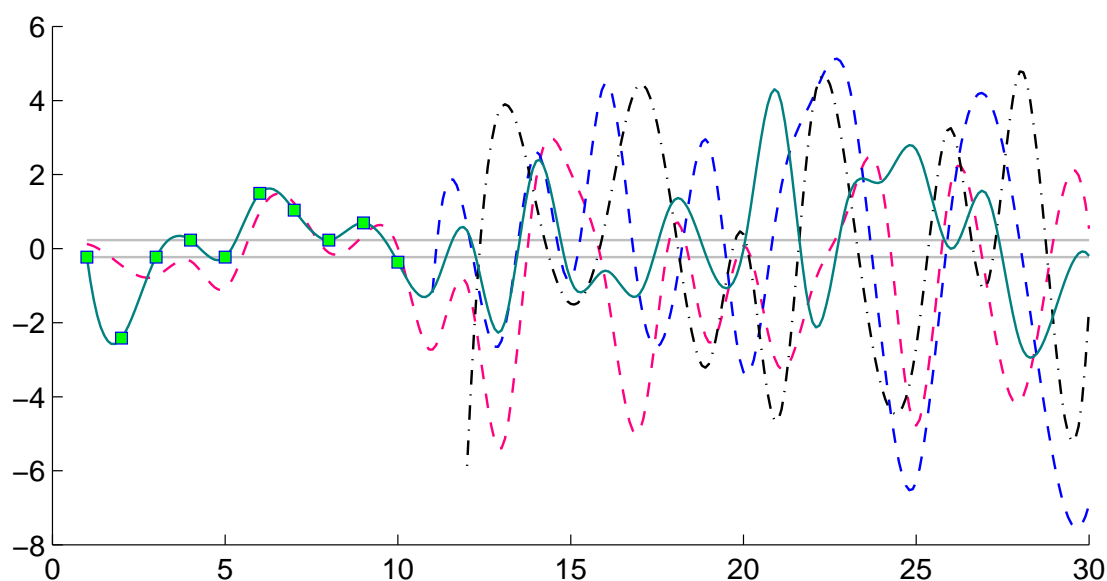


Рис. А.5. Залишки для спрощеної моделі Фрідмана.

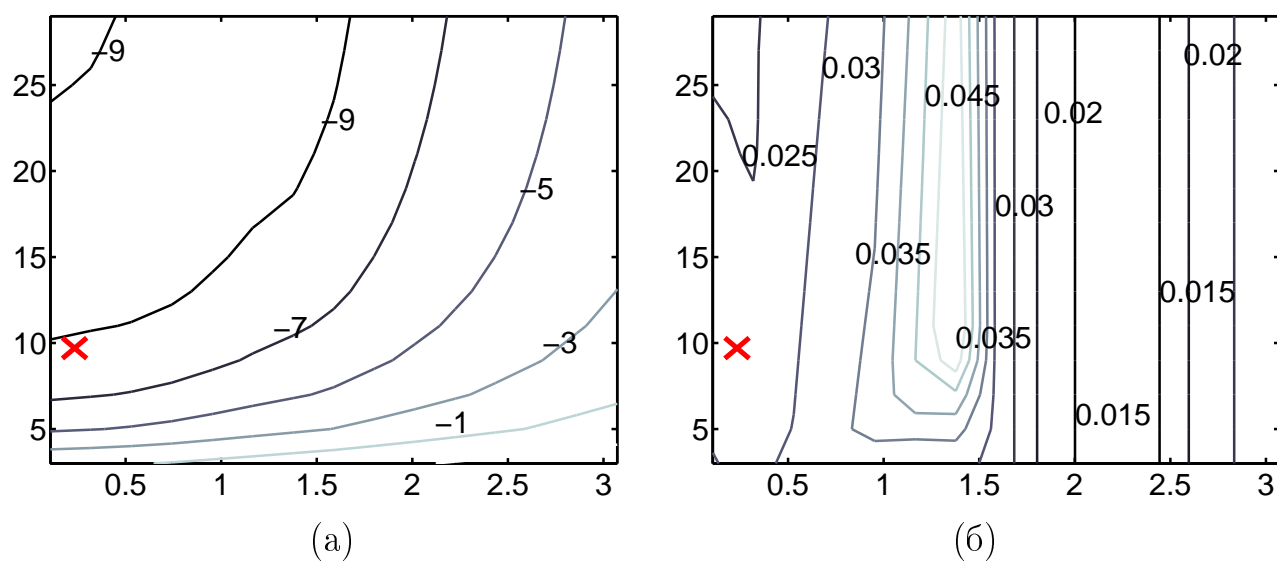


Рис. А.6. Залежність (а) КБП і (б) СКПн спрощеної моделі Фрідмана від ширини смуги нечутливості ϵ для різних значень гіперпараметру β .

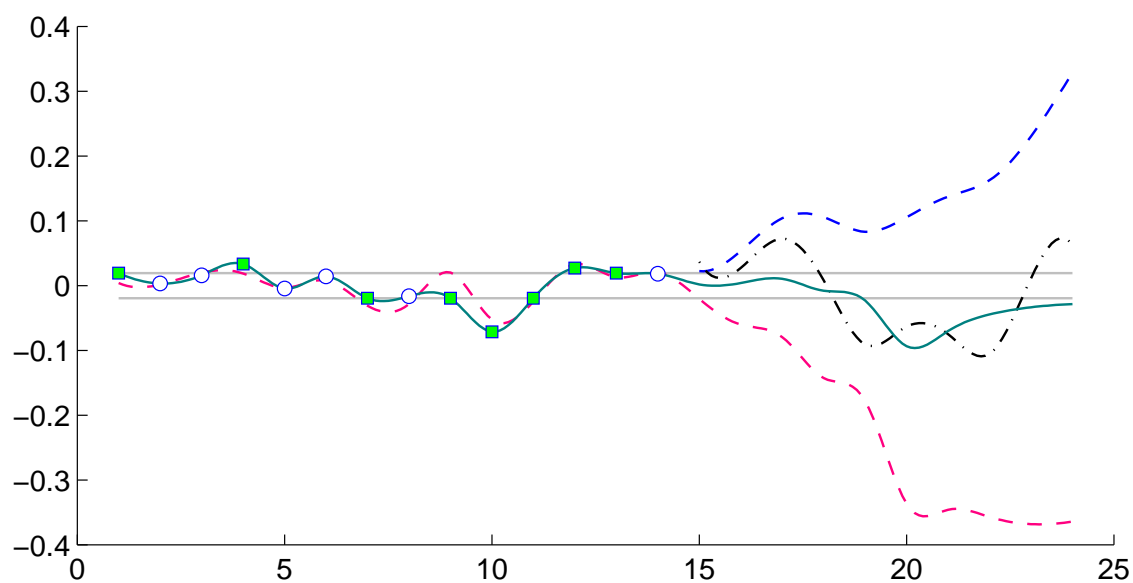
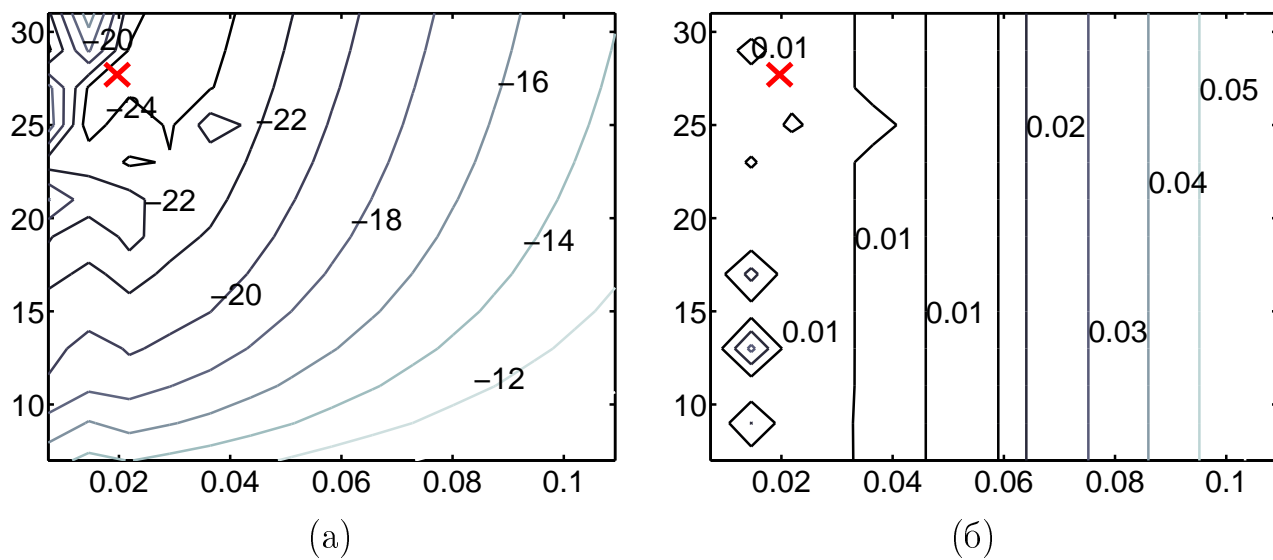


Рис. А.7. Залишки для моделі ІСЦ.

Рис. А.8. Залежність (а) КБП і (б) СКПн моделі ІСЦ від ширини смуги нечутливості ϵ для різних значень гіперпараметру β .

Моделювання реального споживання

Таблиця А.5

Повний розв'язок оптимальної моделі РСГ

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
α	0	0	0	0	0	0	0	6.6	25.5	0	0	0	0.3	0
α^*	0	0	0	0	0	0	0	0	0	0	29.8	2.6	0	0
b	0.43													

№	КБП	$\tilde{\epsilon}$	β	N_{sv}	\tilde{R}_{emp}	\tilde{R}_{reg}	СКП _н	САП _н	
1	-8.96	0.210	20.0	3	0.0000	0.1943	0.0275	0.1534	
2	-9.49	0.200	21.0	5	0.0002	0.3451	0.0274	0.1477	
3	-9.60	0.190	21.0	5	0.0020	0.8185	0.0245	0.1417	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
30	-10.28	0.203	30.0	4	0.0000	0.2396	0.0268	0.1489	
31	-10.28	0.199	29.7	5	0.0003	0.4115	0.0272	0.1474	
32	-10.28	0.199	29.8	5	0.0004	0.4334	0.0271	0.1472	
33	-10.29	0.200	29.8	5	0.0001	0.3387	0.0275	0.1480	
ϵ							СКП	САП	
0.116							ПРИАМ	0.009	0.086
							МГВА	0.014	0.101
							РНМ	0.034	0.165
							НМГВА	0.033	0.136

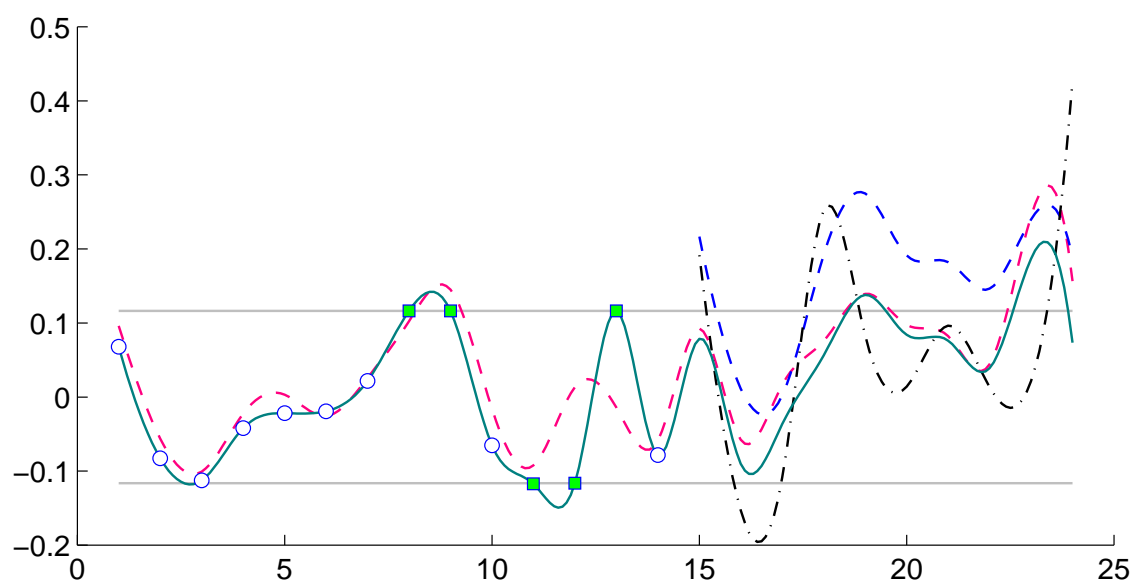


Рис. А.9. Залишки для моделі РСГ.

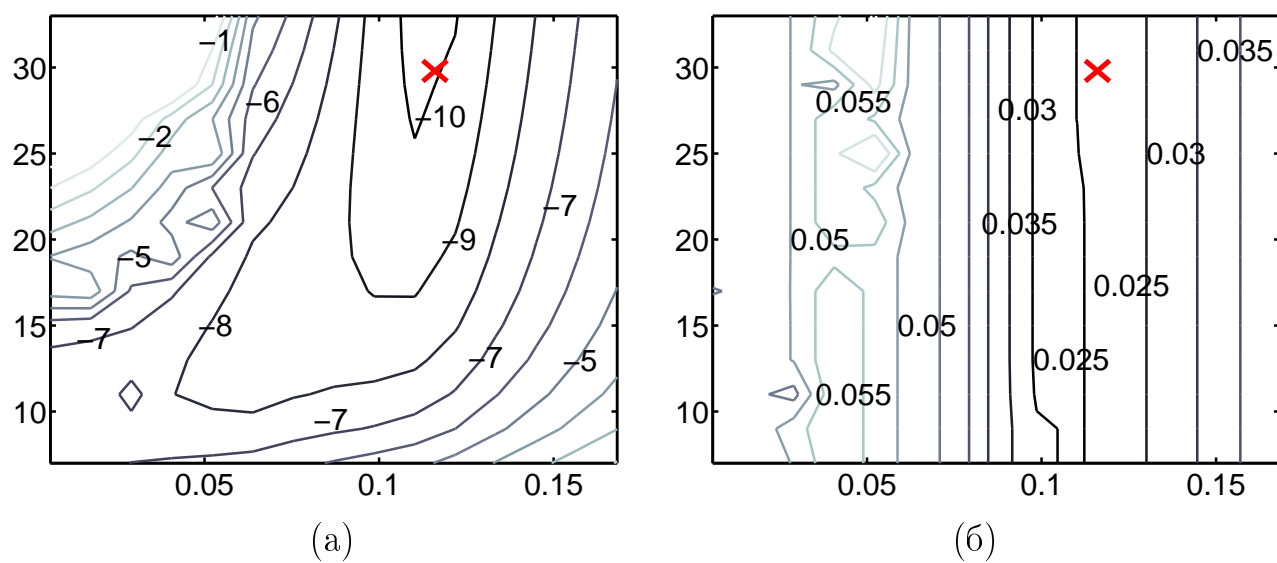


Рис. А.10. Залежність (а) КБП і (б) СКПн моделі РСГ від ширини смуги нечутливості ϵ для різних значень гіперпараметру β .

ДОДАТОК Б

ІНФОРМАЦІЙНА СИСТЕМА ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ

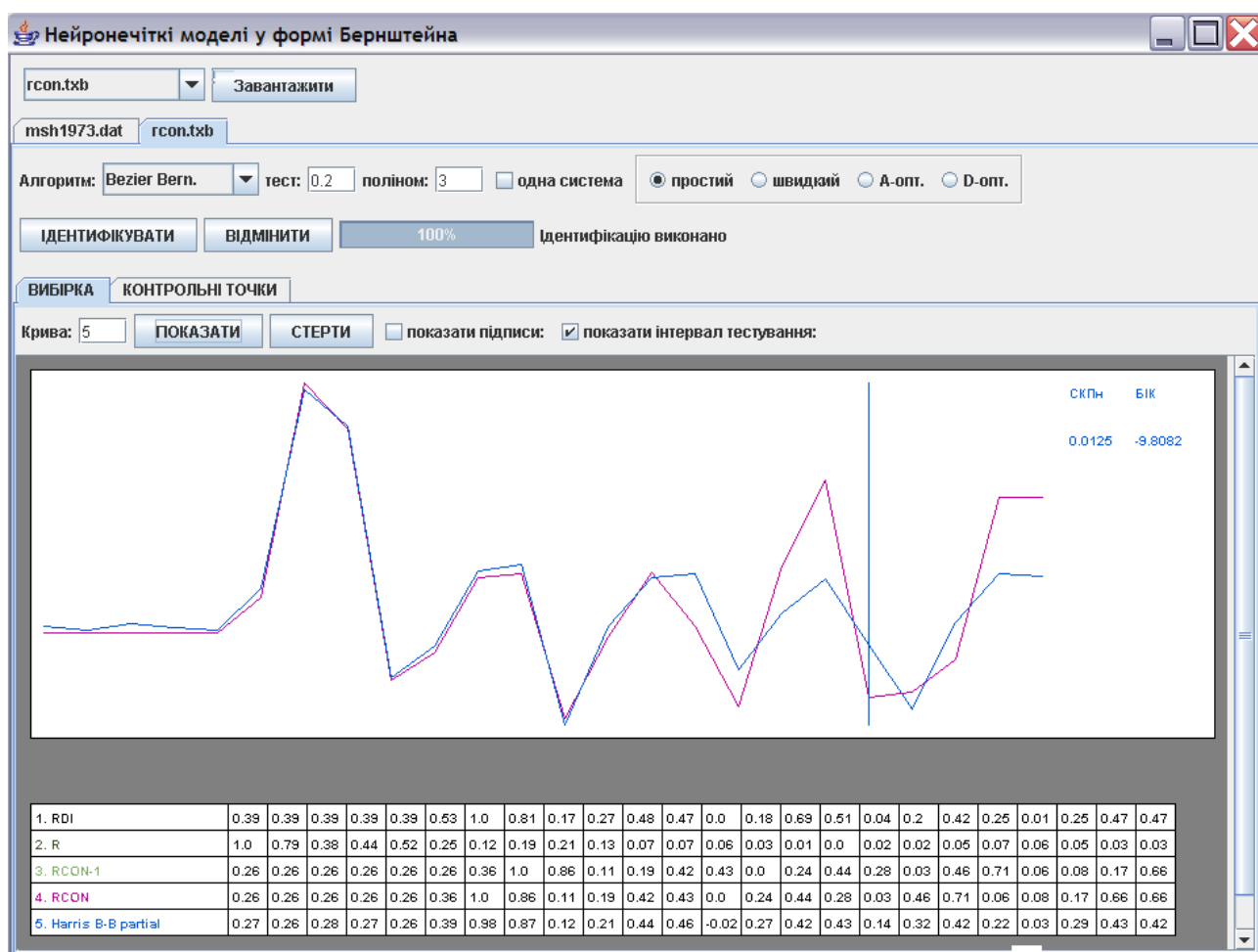


Рис. Б.1. Приклад побудови моделі (скріншот моделі у формі Бернштейна).

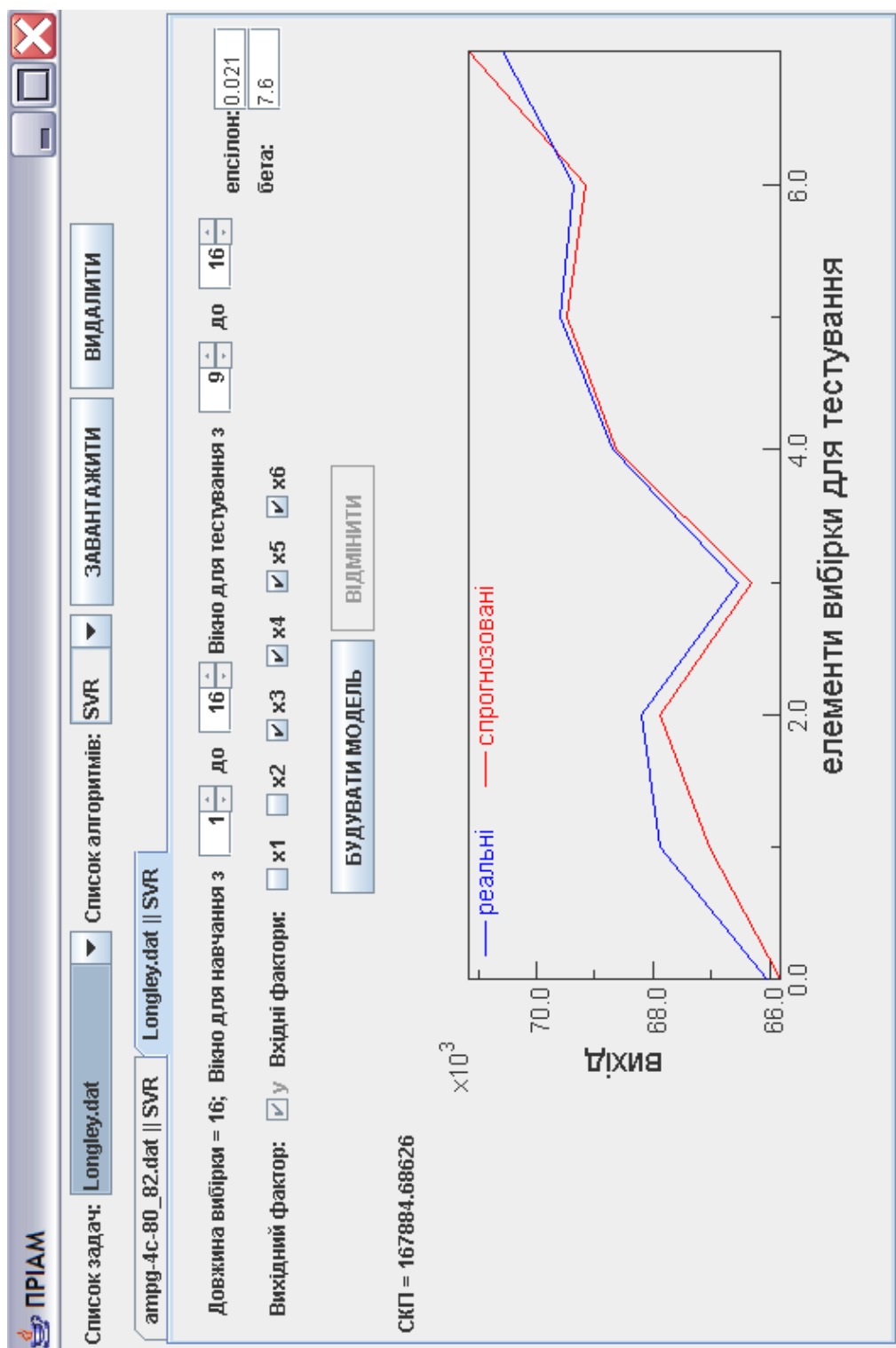


Рис. Б.2. Приклад побудови моделі (скріншот ПРІАМ).

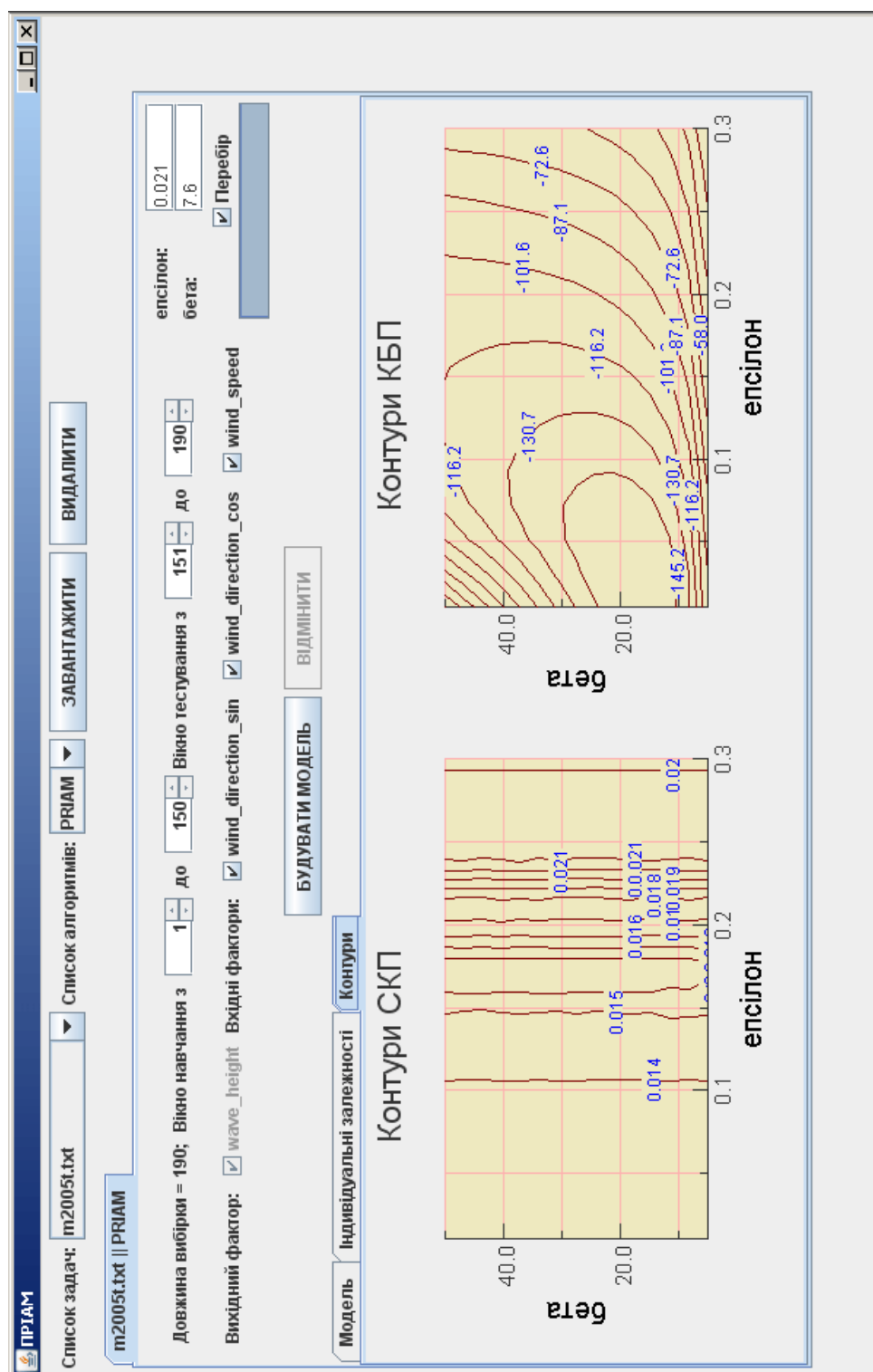


Рис. Б.3. Приклад побудови моделі (скріншот аналізу залежності КБП і СКП).

ДОДАТОК В

МАТЕМАТИЧНИЙ ДОДАТОК

Приклад простого байєсівського висновку

Припустимо, що необхідно оцінити яка частина виборців r (учасників голосування) країни C скаже “за” на референдумі з питання Q . Нехай було проведено соціологічне опитування серед $n = 10$ чоловік, $m = 7$ з яких проголосують “за”. Згідно теореми Байєса ми можемо обчислити апостеріорну щільність розподілу для r :

$$f(r|n = 10, m = 7) = \frac{f(m = 7|r, n = 10)f(r)}{\int_0^1 f(m = 7|r, n = 10)f(r)dr},$$

де $f(r)$ — апріорна щільність розподілу, $f(m = 7|r, n = 10)$ — функція правдоподібності. Припустимо (момент суб’єктивізму), що апріорний ймовірнісний розподіл рівномірний на $[0; 1]$, тобто $f(r) = 1$. Припускаючи, що експерти випадкові, функція правдоподібності $L(r) = f(m = 7|r, n = 10)$ має вигляд біноміального розподілу для 7 успіхів з 10 спроб:

$$Pr(m = 7|r, n = 10) = \binom{10}{7} r^7 (1 - r)^3.$$

Нормалізуючий множник (знаменник, підтвердження) складе:

$$\int_0^1 f(m = 7|r, n = 10)f(r)dr = \int_0^1 \binom{10}{7} r^7 (1 - r)^3 1dr = \binom{10}{7} \frac{1}{1320}.$$

Таким чином апостеріорна щільність розподілу:

$$f(r|n=10, m=7) = \frac{\binom{10}{7} r^7 (1-r)^3}{\binom{10}{7} \frac{1}{1320}} = 1320 r^7 (1-r)^3.$$

Насправді, цікаво отримати ймовірність того, що більше половини виборців проголосують “за” і питання Q буде прийнято. Зважаючи на симетрію рівномірного розподілу, апріорна ймовірність такої події $1/2$. Але після опитування апостеріорна ймовірність такої події:

$$\int_{1/2}^1 1320 r^7 (1-r)^3 dr \approx 0.887.$$

що приблизно відповідає 89% успіху.

Гібридизація нечіткої логіки і нейронних мереж

Доведення теореми 2.2. Задамо ймовірність (силу) правила “якщо $\mathbf{x} \in A^i$, тоді $y \in B^j$ ” як c_{ij} . При цьому $\sum_j c_{ij} = 1$.

Етап фазифікації. Для всіх значень входу \mathbf{x} і виходу y обчислюються значення $\mu_{A^i}(\mathbf{x})$, $\mu_{B^j}(y)$.

Етап нечіткої імплікації. Оскільки імплікація реалізується операцією добутку (а не мінімуму), то результат імплікації для кожного правила є $\mu_{A^i}(\mathbf{x})\mu_{B^j}(y)$.

Етап агрегації. Агрегація реалізується операцією зваженої суми, тож нечітке значення виходу складе:

$$\mu_{A^i \rightarrow B^j}(y) = \sum_{ij} c_{ij} \mu_{A^i}(\mathbf{x}) \mu_{B^j}(y) = \sum_i \mu_{A^i}(\mathbf{x}) \sum_j c_{ij} \mu_{B^j}(y).$$

Етап дефазифікації. Для центроїдного методу:

$$\begin{aligned}
 y(\mathbf{x}) &= \frac{\int \mu_{A^i \rightarrow B^j}(y) y dy}{\int \mu_{A^i \rightarrow B^j}(y) dy} = \frac{\sum_i \mu_{A^i}(\mathbf{x}) \sum_j c_{ij} \int \mu_{B^j}(y) y dy}{\underbrace{\sum_i \mu_{A^i}(\mathbf{x})}_{=1} \underbrace{\sum_j c_{ij}}_{=1} \underbrace{\int \mu_{B^j}(y) dy}_{=const}} = \\
 &= \sum_i \mu_{A^i}(\mathbf{x}) w_i, \quad \text{де} \quad w_i = \sum_j c_{ij} y_j^c, \quad \text{де} \quad y_j^c = \frac{\int \mu_{B^j}(y) y dy}{\int \mu_{B^j}(y) dy}.
 \end{aligned}$$

□

Приклад генерації нечітких правил за схемою К.Харріса

Покажемо яким чином можна сформулювати базу нечітких правил для нейро-нечіткої моделі РСГ, отриманої в підрозділі 4.3.2. Збалансована нейронечітка модель у формі Бернштейна має вигляд:

$$\begin{aligned}
 f_{\text{opt}}(\mathbf{x}) &= 0.43 + 0.06\phi_0^1(x^1) - 0.06\phi_1^1(x^1) - \\
 &\quad - 0.58\phi_0^2(x^2) + 0.13\phi_1^2(x^2) + 0.45\phi_2^2(x^2). \quad (\text{B.1})
 \end{aligned}$$

Тож нейронечіткі підмоделі в канонічній формі відповідно визначаються як:

$$\begin{aligned}
 f_1(x^1) &= 0.49\phi_0^1(x^1) + 0.37\phi_1^1(x^1), \\
 f_2(x^2) &= -0.15\phi_0^2(x^2) + 0.56\phi_1^2(x^2) + 0.88\phi_2^2(x^2).
 \end{aligned}$$

Для простоти визначимо нечіткі множини для змінної виходу *RCON* через чотири В-сплайни першого порядку $N_0^1, N_1^1, N_2^1, N_3^1$ (2.3) визначені на вузлах $\{-1; -0.5; 0; 0.5; 1; 1.5\}$, які відповідають нечітким змінним: дуже низьке, низьке, середнє, високе з центрами в точках $\{-0.5; 0; 0.5; 1\}$. Модель (4.1) визначає нечіткі множини для змінних входу $x^1 \equiv R$, $x^2 \equiv RDI$ через базисні поліноми Бернштейна першого ϕ_0^1, ϕ_1^1 та другого $\phi_0^2, \phi_1^2, \phi_2^2$ порядків відповідно. Вони відповідають нечітким змінним: низька та висока відсоткова ставка; низькі, середні, високі реальні доходи.

Ймовірності правил, що індуковані нейронечіткою моделлю f_1 , можна

знайти розв'язавши систему:

$$\begin{cases} 0.49 = -0.5c_{11}^1 + 0 \cdot c_{12}^1 + 0.5c_{13}^1 + 1 \cdot c_{14}^1 \\ 0.37 = -0.5c_{21}^1 + 0 \cdot c_{22}^1 + 0.5c_{23}^1 + 1 \cdot c_{24}^1 \end{cases} \Leftrightarrow \begin{cases} c_{12}^1 = 0.02; \quad c_{13}^1 = 0.98 \\ c_{22}^1 = 0.26; \quad c_{23}^1 = 0.74 \end{cases}$$

з обмеженнями $\sum_j c_{ij}^1 = 1$, $c_{ij}^1 \in [0; 1]$ і максимум два суміжні коефіцієнти (c_{ij}^1 і c_{ij+1}^1) ненульові для заданого i .

Аналогічно для моделі f_2 можна скласти і розв'язати систему:

$$\begin{cases} -0.15 = -0.5c_{11}^2 + 0 \cdot c_{12}^2 + 0.5c_{13}^2 + 1 \cdot c_{14}^2 \\ 0.56 = -0.5c_{21}^2 + 0 \cdot c_{22}^2 + 0.5c_{23}^2 + 1 \cdot c_{24}^2 \\ 0.88 = -0.5c_{31}^2 + 0 \cdot c_{32}^2 + 0.5c_{33}^2 + 1 \cdot c_{34}^2 \end{cases} \Leftrightarrow \begin{cases} c_{11}^1 = 0.30; \quad c_{12}^1 = 0.70 \\ c_{23}^1 = 0.88; \quad c_{24}^1 = 0.12 \\ c_{33}^1 = 0.24; \quad c_{34}^1 = 0.76 \end{cases}$$

Тож нечіткі правила та їх значимість визначаються наступним чином:

- \hookrightarrow Правило 1: якщо R низька тоді RCON низьке (0.02) або середнє (0.98)
- \hookrightarrow Правило 2: якщо R висока тоді RCON низьке (0.26) або середнє (0.74)
- \hookrightarrow Правило 3: якщо RDI низькі тоді RCON дуже низьке (0.3) або низьке (0.7)
- \hookrightarrow Правило 4: якщо RDI середні тоді RCON середнє (0.88) або високе (0.12)
- \hookrightarrow Правило 5: якщо RDI високі тоді RCON середнє (0.24) або високе (0.76)

Користуючись цими правилами можна робити висновки на основі неповної інформації. Наприклад, якщо немає даних щодо RDI, а відомо тільки, що R висока, то працює правило 2. Якщо відомі обидва фактори, то відповідні два правила можна комбінувати. На основі збалансованої моделі можна навіть зробити висновок щодо вихідної змінної в умовах повної невизначеності. В нашому випадку це $y = 0.43$.

ДОДАТОК Г

ДОКУМЕНТИ, ЯКІ ПІДТВЕРДЖУЮТЬ АПРОБАЦІЮ ОСНОВНИХ
ВИСНОВКІВ І РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ

Севастопольская гидрометеорологическая обсерватория

Севастопольская гидрометеорологическая обсерватория

г. Севастополь, ул. Советская, 61

☎ (0692) 54-36-81(факс) (0692) 54-83-28 (0692) 54-02-96 ✉ pogoda@ukrpost.ua

✉ г. Севастополь, п / я 260 ☐ Телеграф 197117 море

30 мая 2008г. № 125

СПРАВКА

о внедрении результатов диссертационной работы Мытника Олега Юрьевича «Информационные технологии синтеза робастных нейронечётких моделей стохастических процессов», поданной на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии.

Одной из основных задач Севастопольской гидрометеорологической обсерватории является обслуживание потребителей информацией о состоянии погоды и моря по закрепленной за обсерваторией зоне ответственности, в том числе по акватории ряда морских портов, размещенных на южном побережье Крыма. Для обеспечения безопасного мореплавания в акватории морских портов большее значение имеет суточный прогноз высоты волны. Основным фактором, определяющим распространение и характеристики волн на море, является ветер. Для получения прогноза параметров ветра используются прогностические карты распределения ветра на основе численных прогнозов полей атмосферного давления на уровне моря. Севастопольская гидрометеорологическая обсерватория использует численную схему прогноза полей волнения, основанную на методе Шулейкина из теории поля ветровых волн. Этот метод успешно используется для прогнозирования элементов волн глубокого открытого моря. Но для построения прогнозов на мелководье или возле берега не существует эффективных теоретически обоснованных процедур, которые бы учитывали особенности местности. Поэтому на практике все зависит от опыта прогнозистов на местах.

Мытник О.Ю. разработал универсальный метод математического моделирования и прогнозирования стохастических процессов с нелинейностями разных типов. Использование в гидрометеорологической обсерватории разработанного диссертантом статистического метода дало возможность построить высокоэффективные прогнозирующие модели, которые используются для поддержки принятия решений прогнозистами с целью уточнения прогноза параметров волн. Для автоматизации процесса принятия решений диссертантом также создана информационная система

обработки накопленных нами данных ежедневных наблюдений за параметрами ветровых волн (высота, период, длина, направление) и показателями ветра (направление, скорость, продолжительность действия) с 1960 по 2006 год.

Директор Севастопольской ГМО



В.А.Наумова