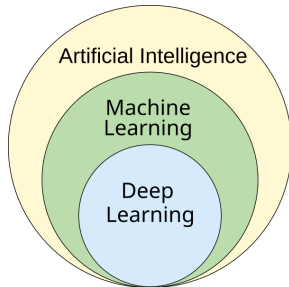


Основные понятия и задачи машинного обучения

Виктор Китов

[victorkitov.github.io](https://github.com/victorkitov)



Курс поддержан
фондом
'Интеллект'



Победитель
конкурса VK среди
курсов по IT



Содержание

- 1 Постановка задачи
- 2 Другие постановки задачи
- 3 Функциональный класс
- 4 Оценка параметров модели

Актуальность

Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, искусственном интеллекте и машинном обучении.

Клаус Мартин Шваб,
президент Всемирного экономического форума.



Идея машинного обучения

- Машинное обучение - наука, позволяющая компьютерам принимать сложные решения без явной спецификации алгоритма принятия решения.
 - алгоритм находится в большом параметризованном семействе моделей.
- Компьютер учится на заданном опыте решать некоторый класс задач, относительно некоторого показателя качества, если показатель качества растет на классе задач после получения опыта.

Примеры

- Фильтрация спама

- если отправитель осуществляет массовую рассылку, на которую пользователи не отвечают, а тело письма содержит ключевые слова "уникальное предложение" и "приобретите сейчас" -> спам
- если пользователь уже отвечал на письмо -> не спам
- если в подписи письма отправитель и получатель из одной организации -> не спам
- ...

- Разметка частей речи.

пр.	сущ.	глагол.	прил.	сущ.
	С	гор	побежали	звонкие ручейки.

- если слово заканчивается на "еть" -> глагол (смотреть, хотеть)
- если предыдущее слово "к" -> существительное (к солнцу, к делу)
- ...

Ручной подход и машинное обучение¹

Ручной подход:

- нужны специалисты, дорого, медленно
- неточно (люди извлекают лишь простые правила)

Эти проблемы решает машинное обучение.

- нужен "опыт" (обучающая выборка размеченных данных)
- вычислительные ресурсы для настройки моделей

¹Источник иллюстрации.

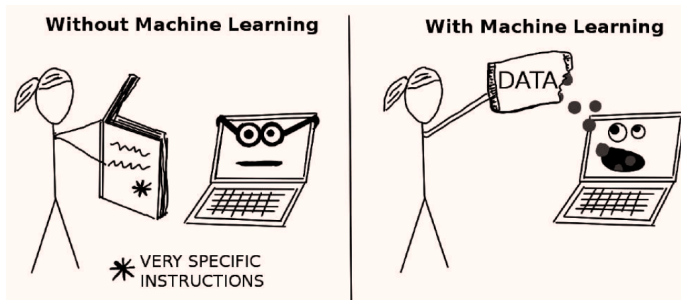
Ручной подход и машинное обучение¹

Ручной подход:

- нужны специалисты, дорого, медленно
- неточно (люди извлекают лишь простые правила)

Эти проблемы решает машинное обучение.

- нужен "опыт" (обучающая выборка размеченных данных)
- вычислительные ресурсы для настройки моделей



¹Источник иллюстрации.

Примеры задач машинного обучения

- Предсказать, уйдёт ли клиент к конкурентам? (churn prediction)
- Является ли последовательность финансовых транзакций мошеннической? (fraud detection)
- Предсказание пробок и времени в пути при планировании маршрута (traffic prediction).
- Стоит ли показывать заданный товар покупателю в качестве рекомендации? (recommender systems)
- Рекомендовать ли человека в качестве друга в социальной сети?
- Голосовой ассистент: распознавание речи, автоматический ответ на вопросы, генерация речевого ответа.
- Идентификация человека по лицу. Распознавание номера машины на камерах.

Примеры задач машинного обучения

- Подсчёт и отслеживание людей по камерам видеонаблюдения (object tracking) Обнаружение неправомерных действий (activity recognition).
- Автоматическое управление машинами (self-driving cars): распознавание ситуации, планирование маршрута.
- Автоматическая торговля на бирже (algorithmic trading).
- Перевод с одного языка на другой (machine translation).
- Автоматическая игра в шахматы, управление игровыми персонажами.
- Постановка медицинских диагнозов по жалобам пациента и результатам обследований.
- Рекомендация веб-страниц по поисковому запросу (information retrieval).
- Автоматическая оценка недвижимости, зарплаты по резюме.

Примеры задач машинного обучения

- Хвалит или ругает пользователь товар в своём отзыве? (sentiment analysis)
- Генерация иллюстраций к тексту. Текстовое описание, что показано на изображении.
- Прогноз погоды. Рекомендации фермерам, когда сажать/поливать/удобрять посевы.
- Автоматическое написание программного кода (no code).
- Автоматический выбор, каким пользователям какую онлайн-рекламу показать (targeted ads).
- Генерация химических соединений, обладающих требуемыми свойствами:
 - крепкий, но легкий и термостойкий материал с повышенной проводимостью (material design)
 - препарат, обеспечивающий лечение и обладающий минимальными побочными эффектами (drug discovery)

Виды обучения

- **Машинное обучение** (machine learning) - в целом про настройку прогнозирующих алгоритмов.
- **Глубокое обучение** (deep learning) - подраздел про сложные многоуровневые модели (нейросети).
- **Обучение с подкреплением** (reinforcement learning) - прогнозируем не один раз, а вырабатываем интерактивную стратегию поведения в изменяемой среде.
 - шахматы, игровые персонажи, управление машинами, дронами и роботами-ассистентами.

Виды обучения

- **Машинное обучение** (machine learning) - в целом про настройку прогнозирующих алгоритмов.
- **Глубокое обучение** (deep learning) - подраздел про сложные многоуровневые модели (нейросети).
- **Обучение с подкреплением** (reinforcement learning) - прогнозируем не один раз, а вырабатываем интерактивную стратегию поведения в изменяемой среде.
 - шахматы, игровые персонажи, управление машинами, дронами и роботами-ассистентами.

Машинное обучение сводит теорию и практику воедино.
Мат. модели работают для решения реальных задач!

- Чтобы запрограммировать сложную логику принятия решений, нужно самому разобраться, как человек принимает решения.

Где машинное обучение дает преимущество

- **сложно сформулировать явную зависимость**
 - слишком много наблюдений
 - логи сайта
 - слишком много признаков
 - категоризация текстов
 - сложные взаимосвязи признаков
 - классификация изображений
 - сложный результат генерации
 - изображение, видео, длинный текст
- **нужна быстрая скорость адаптации к изменяющимся условиям**
 - предсказание цен на акции
- **необходимо построить много моделей**
 - для голосового помощника - свою под каждого пользователя

Формальные определения

- Есть класс объектов Z
- Каждый объект описывается вектором известных характеристик (признаков) $x \in \mathcal{X}$ и предсказываемых характеристик (откликов) $y \in \mathcal{Y}$.

$$z = (x, y) \in Z$$

- Задача: найти отображение f , которое бы точно описывало взаимосвязь $\mathcal{X} \rightarrow \mathcal{Y}$.
 - используя конечный набор известных пар (x, y) обучающей выборки.
 - для применения к любым новым x в тестовой выборке.
- Тестовая выборка может быть известна или неизвестна заранее.

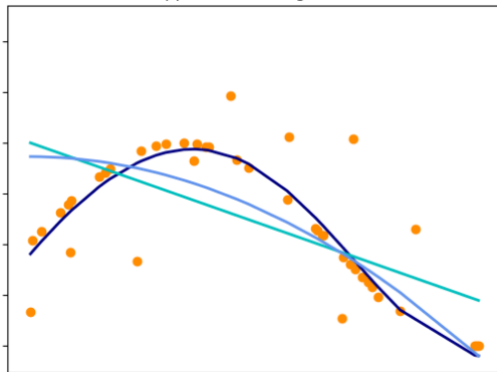
Основные типы признаков

- Входное описание объекта $x \in \mathcal{X}$ состоит из индивидуальных признаков $x^i \in \mathcal{X}_i$.
- Типы каждого признака (например, в задаче кредитного скоринга):
 - $\mathcal{X}_i = \mathbb{R}$ - вещественный (количественный) признак
 - например, возраст, зарплата.
 - $\mathcal{X}_i = \{0, 1\}$ - бинарный признак
 - пример: есть ли у должника просрочки по платежам?
 - $|\mathcal{X}_i| < \infty$ - дискретный категориальный признак
 - пример: профессия.
 - $|\mathcal{X}_i| < \infty$ и \mathcal{X}_i упорядоченный дискретный (порядковый) признак
 - пример: уровень образования.

Возможные постановки задачи

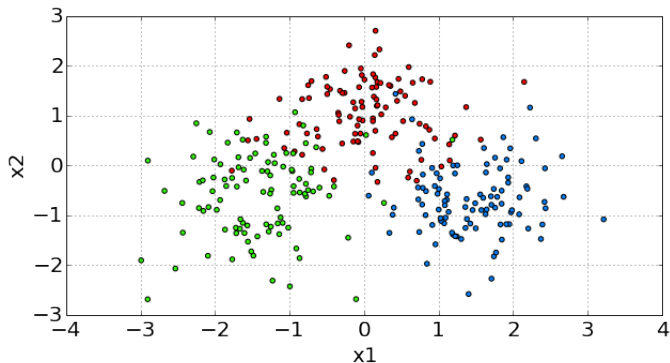
- **Обучение с учителем** (supervised learning):
 - обучающая выборка: $(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$
 - **трансдуктивное обучение**: входы x тестовой выборки известны заранее.
- **Обучение без учителя** (unsupervised learning):
 - обучающая выборка: $x_1, x_2, \dots x_N$

Обучение с учителем - регрессия



Предсказать $y \in \mathbb{R}$ для любого x .

Обучение с учителем - классификация



Предсказать дискретный y , обозначенный цветом, в каждой точке.

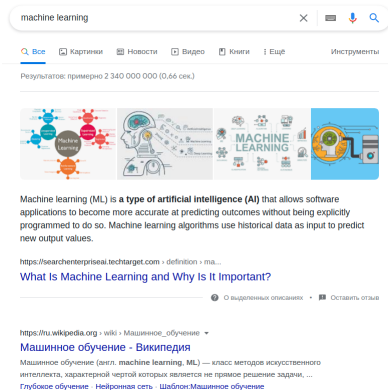
Пример: медицинские приложения

- **объекты:** пациенты
- **признаки:**
 - вещественные: возраст, пульс, артериальное давление, содержание гемоглобина в крови, доза лекарства.
 - бинарные: пол, наличие головной боли, слабости, тошноты
 - категориальные: перенесенные болезни
 - порядковые: тяжесть состояния
- **возможные отклики:**
 - классификация: определить тип болезни, способ лечения.
 - регрессия: длительность лечения и выздоровления.

Пример: прогнозирование поведения клиентов

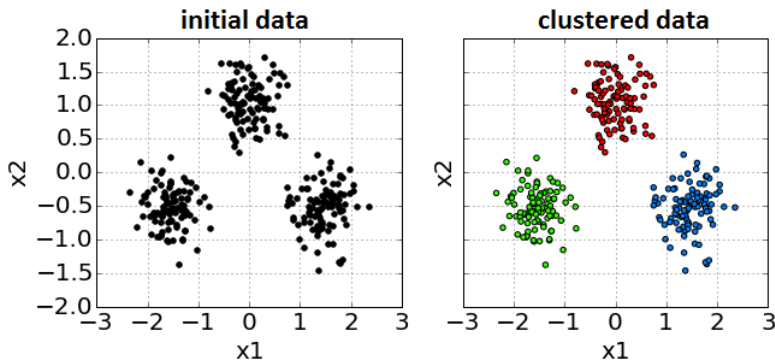
- **объекты:** клиент в текущий момент времени.
- **признаки:**
 - вещественные: возраст, историческая частота пользования услугами и траты на услуги.
 - бинарные: пол, были ли задолженности по платежам.
 - категориальные: какими услугами пользуется.
 - порядковые: оценка компании по мнению клиента.
 - возможные отклики:
- **классификация:** уйдет ли клиент к конкурентам?
подключит ли услугу?
- **регрессия:** сколько раз воспользуется услугой? сколько денег внесет на счёт?

Обучение с учителем - ранжирование



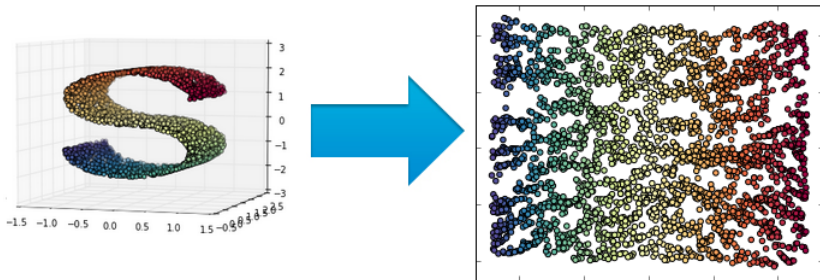
Как можно было бы решить задачу ранжирования через регрессию или классификацию?

Обучение без учителя - кластеризация



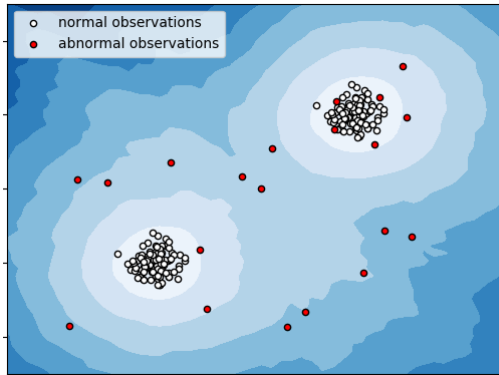
Разбивка объектов на похожие группы.

Обучение без учителя - снижение размерности



Переход из 3D в 2D с минимальным искажением геометрии.

Обучение без учителя - обнаружение аномалий



Выделение нетипичных объектов.

Поиск ассоциативных правил

rhs	lhs	support	confidence	lift
bottled beer	liquor, red/blush wine	0.0016268429	0.9411765	19.53
	liquor, red/blush wine, soda	0.0006100661	1.0000000	20.75
	liquor, soda	0.0010167768	0.7692308	15.96
bottled water	bottled beer, misc. beverages	0.0005083884	0.5000000	8.29
citrus fruit	meat, turkey	0.0004067107	0.5714286	6.97
coffee	condensed milk, sugar	0.0004067107	0.8000000	25.71
frankfurter	liver loaf, sausage	0.0005083884	0.5000000	8.48
liquor	bottled beer, red/blush wine	0.0016268429	0.6666667	81.96
	bottled beer, red/blush wine, s...	0.0006100661	1.0000000	122.94
	red/blush wine, soda	0.0006100661	0.5454545	67.06

Анализ потребительских корзин (market basket analysis)

По наборам множеств $\{a, b, c\}$, $\{a, d, e\}$, $\{a, b\}$, $\{a, b, g, h\}$
генерировать правила: $a \rightarrow b$, $b \rightarrow a, \dots$

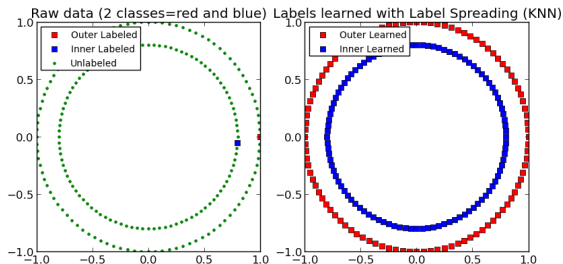
Содержание

- 1 Постановка задачи
- 2 Другие постановки задачи
- 3 Функциональный класс
- 4 Оценка параметров модели

Частичное обучение - классификация

Частичное обучение (semi-supervised learning):

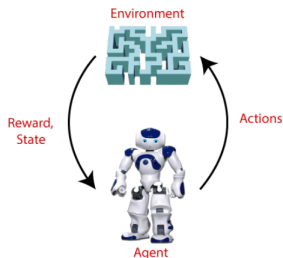
- обучающая выборка:
 $(x_1, y_1), (x_2, y_2), \dots (x_N, y_N), x_{N+1}, x_{N+2}, \dots x_{N+M}$
 - positive-unlabelled (PU) learning - все $y_1 = \dots y_N = +1$
(применяется в выделении нетипичных наблюдений)
- Пример (предполагаем, что близкие точки принадлежат одному классу):



Обучение с подкреплением

Обучение с подкреплением (reinforcement learning):

- агент совершает действия в среде
- действия переводят в среду в др. состояния
- среда дает вознаграждения за состояния
- $(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$ формируется динамически по действиям агента



Source: <https://www.microsoft.com/en-us/research/blog/training-deep-control-policies-for-the-real-world/>

Обучение с учителем - задача

- Требуется найти отображение $f(x) : X \rightarrow Y$.
- Варианты использования:
 - предсказание y по x .
 - анализ зависимости $X \rightarrow Y$ на качественном уровне
 - обнаружение нетипичных объектов (где модель ошибается)
- Вопросы в настройке модели:
 - какую целевую переменную y нужно прогнозировать?
 - какие использовать признаки x ?
 - в каком классе искать отображение f ?
 - в каком смысле отображение f должно приближать зависимость $X \rightarrow Y$?
 - как алгоритмически подбирать параметры f ?

Основные типы откликов

- $\mathcal{Y} = \mathbb{R}$ - регрессия
 - например, цена акции
- $\mathcal{Y} = \mathbb{R}^M$ - векторная регрессия
 - например, динамика цен на квартиры
- $\mathcal{Y} = \{\omega_1, \omega_2, \dots, \omega_C\}$ - классификация.
 - $C=2$: бинарная классификация.
 - например, спам/не спам
 - $C>2$: многоклассовая классификация
 - например, идентификация пользователя
- \mathcal{Y} - множественная классификация из $\{\omega_1, \omega_2, \dots, \omega_C\}$.²
 - например, категоризация новостей

²Можно ли ее решить используя обычную классификацию?

Содержание

- 1 Постановка задачи
- 2 Другие постановки задачи
- 3 Функциональный класс**
- 4 Оценка параметров модели

Пример линейного класса функций.

- Регрессия: $\hat{y} = g(x)$, $g(x)$ параметризовано θ .

³Однозначно ли определены дискриминантные ф-ции?

Пример линейного класса функций.

- Регрессия: $\hat{y} = g(x)$, $g(x)$ параметризовано θ .
- Многоклассовый классификатор ($y \in \{1, 2, \dots, C\}$)³:

$$\hat{y}(x) = \arg \max_c g_c(x), \quad g(x) \text{ параметризовано } \theta.$$

$$\{x : g_i(x) = g_j(x)\}, \quad \text{граница между классами } i, j.$$

$$M(x, y) = g_y(x) - \max_{c \neq y} g_c(x), \quad \text{отступ (качество классификации)}$$

³Однозначно ли определены дискриминантные ф-ции?

Пример линейного класса функций.

- Регрессия: $\hat{y} = g(x)$, $g(x)$ параметризовано θ .
- Многоклассовый классификатор ($y \in \{1, 2, \dots, C\}$)³:

$$\hat{y}(x) = \arg \max_c g_c(x), \quad g(x) \text{ параметризовано } \theta.$$

$$\{x : g_i(x) = g_j(x)\}, \quad \text{граница между классами } i, j.$$

$$M(x, y) = g_y(x) - \max_{c \neq y} g_c(x), \quad \text{отступ (качество классификации)}$$

- Бинарный классификатор ($y \in \{+1, -1\}$):

$$\hat{y}(x) = \arg \max_{c \in \{+1, -1\}} g_c(x) = \text{sign}(g_{+1}(x) - g_{-1}(x)) = \text{sign}(g(x))$$

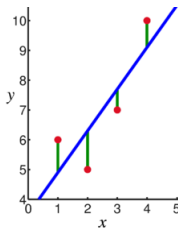
$$M(x, y) = g_y(x) - g_{-y}(x) = y(g_{+1}(x) - g_{-1}(x)) = yg(x)$$

³Однозначно ли определены дискриминантные функции?

Примеры

линейная регрессия $y \in \mathbb{R}$:

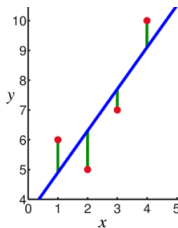
$$f(x|w) = w_0 + w_1x$$



Примеры

линейная регрессия $y \in \mathbb{R}$:

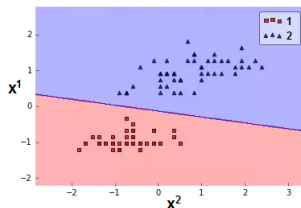
$$f(x|w) = w_0 + w_1x$$



линейная классификация
 $y \in \{1, 2\}$:

$$g_c(x|w) = w_c^0 + w_c^1x^1 + w_c^2x^2, \quad c = 1, 2.$$

$$f(x|w) = \arg \max_c g_c(x|\theta)$$



Функция качества / потерь

- Точность предсказаний может оцениваться:
 - **критерием качества** (score function, выше->лучше)
 - **функцией потерь** (loss function, ниже->лучше)
- $loss = F(score)$, $score = F^{-1}(loss)$ для некоторой убывающей $F(\cdot)$.
 - например, ф-ция потерь = - ф-ция качества.
- $loss=loss(\text{ошибка})$, $score=score(\text{ошибка})$, где ошибка:
 - регрессия: $(\hat{y} - y)$
 - классификация: $-M(x, y)$.

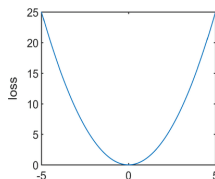
Функции потерь регрессии $F(\hat{y} - y)^4$

$$\text{MAE: } |\hat{y} - y|$$

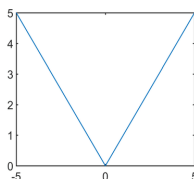
$$\text{MSE: } (\hat{y} - y)^2$$

$$\text{Huber} \begin{cases} \frac{1}{2}(\hat{y} - y)^2, & |\hat{y} - y| \leq \delta \\ \delta (|\hat{y} - y| - \frac{1}{2}\delta) & |\hat{y} - y| > \delta \end{cases}$$

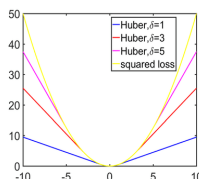
$$\epsilon\text{-insensitive } \max \{|\hat{y} - y| - \epsilon, 0\}$$



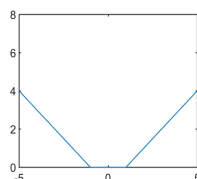
(a) square loss



(b) absolute loss



(c) Huber loss

(d) ϵ -insensitive loss ($\epsilon = 1$)

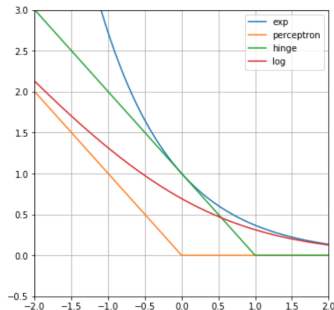
⁴ Выбор ф-ции потерь на практике должен исходить от бизнес-задачи.

Функции потерь классификации $F(M)$

- $Loss = F(M)$ для некоторой убывающей $F(\cdot)$ [выше отступ->лучше].

$$\mathcal{L}_{exp}(M) = e^{-M} \quad \mathcal{L}_{perceptron}(M) = [-M]_+$$

$$\mathcal{L}_{hinge}(M) = [1 - M]_+ \quad \mathcal{L}_{log}(M) = \ln(1 + e^{-M})$$

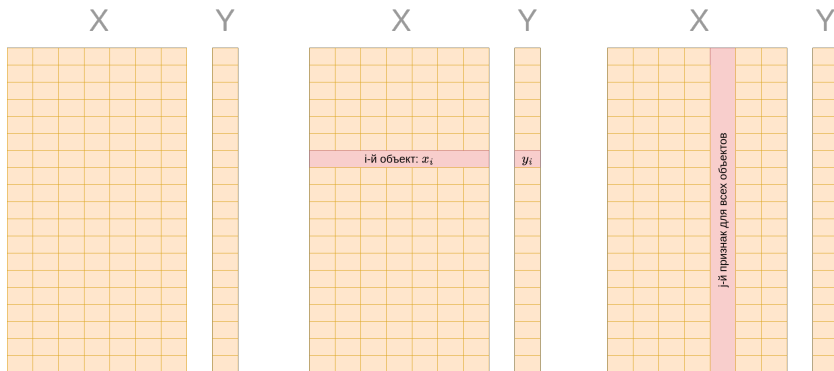


Содержание

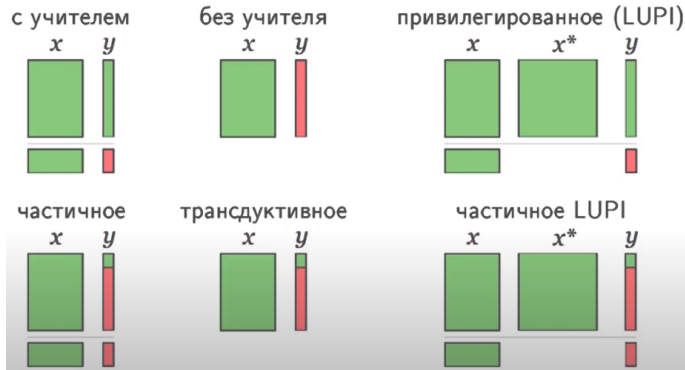
- 1 Постановка задачи
- 2 Другие постановки задачи
- 3 Функциональный класс
- 4 Оценка параметров модели

Обучающая выборка

Обучающая выборка (training set): $(x_1, y_1), \dots, (x_N, y_N)$, задаётся матрицей объекты-признаки (design matrix) $X \in \mathbb{R}^{N \times D}$, и вектором откликов (targets) $Y = [y_1, \dots, y_M]^T$.



Виды обучения

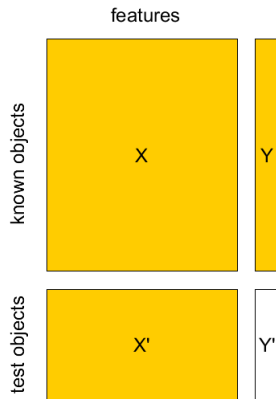


LUPI - Learning Using Priveledged Information⁵.

⁵Vapnik V., Vashist A. A new learning paradigm: Learning Using Priveledged Information // Neural Networks. 2009.

Обучающая и тестовая выборка

- Обучающая выборка $X, Y: (x_1, y_1), \dots (x_M, y_M)$
- Тестовая выборка $X', Y': (x'_1, y'_1), \dots (x'_K, y'_K)$



Критерий оптимизации параметров модели

- Необходимо минимизировать **теоретический риск**:

$$\int \int \mathcal{L}(f_w(x), y) p(x, y) dx dy \rightarrow \min_w$$

⁶Предполагаем что объекты независимы и одинаково распределены.

Критерий оптимизации параметров модели

- Необходимо минимизировать **теоретический риск**:

$$\int \int \mathcal{L}(f_w(x), y) p(x, y) dx dy \rightarrow \min_w$$

- Но мы можем минимизировать только **эмпирический риск**⁶:

$$L(w|X, Y) = \frac{1}{N} \sum_{n=1}^N \mathcal{L}(f_w(x_n), y_n)$$

- Параметры находим из условия:

$$\hat{w} = \arg \min_w L(w|X, Y)$$

⁶Предполагаем что объекты независимы и одинаково распределены.

Метод максимального правдоподобия

Безусловная статистика: $y \sim p_w(y)$, тогда вероятность пронаблюдать обучающую выборку⁷:

$$p(Y) = p(y_1)p(y_2)...p(y_N) \rightarrow \max_w$$

⁷ Выведите вероятность выпадения каждой грани многогранника с K сторонами, если после N подбрасываний стороны 1, 2, ..., K выпали N_1, N_2, \dots, N_K раз.

Метод максимального правдоподобия

Безусловная статистика: $y \sim p_w(y)$, тогда вероятность пронаблюдать обучающую выборку⁷:

$$p(Y) = p(y_1)p(y_2)...p(y_N) \rightarrow \max_w$$

Условная статистика: предсказываем y по x , поэтому

$$y \sim p_w(y|x)$$

$$p(Y|X) = p(y_1|x_1)p(y_2|x_2)...p(y_N|x_N) \rightarrow \max_w$$

⁷ Выведите вероятность выпадения каждой грани многогранника с K сторонами, если после N подбрасываний стороны 1, 2, ..., K выпали N_1, N_2, \dots, N_K раз.

Метод максимального правдоподобия

Безусловная статистика: $y \sim p_w(y)$, тогда вероятность пронаблюдать обучающую выборку⁷:

$$p(Y) = p(y_1)p(y_2)...p(y_N) \rightarrow \max_w$$

Условная статистика: предсказываем y по x , поэтому

$$y \sim p_w(y|x)$$

$$p(Y|X) = p(y_1|x_1)p(y_2|x_2)...p(y_N|x_N) \rightarrow \max_w$$

Для численной устойчивости (избежать машинного нуля) максимизируется логарифм правдоподобия (log-likelihood):

$$\log p(Y|X) = \log p(y_1|x_1) + \log p(y_2|x_2) + ... + \log p(y_N|x_N) \rightarrow \max_w$$

⁷ Выведите вер-ть выпадения каждой грани многогранника с K сторонами, если после N подбрасываний стороны 1,2,... K выпали $N_1, N_2, ...N_K$ раз.

Связь методов

- Принцип максимума правдоподобия:

$$\hat{\theta} = \arg \max_w \frac{1}{N} \sum_{n=1}^N \log p_w(y_n | x_n)$$

- Принцип минимизации эмпирического риска:

$$\hat{\theta} = \arg \min_w \frac{1}{N} \sum_{n=1}^N \mathcal{L}(f_w(x_n), y_n)$$

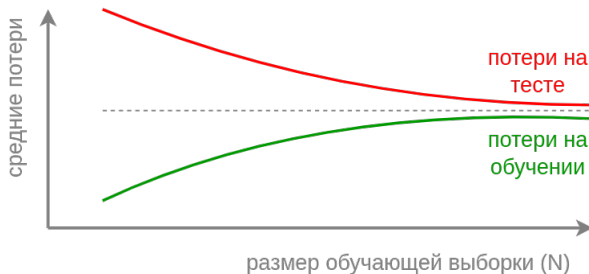
- Связь методов⁸:

$$\mathcal{L}(f_w(x_n), y_n) = -\log p_w(y_n | x_n)$$

⁸Откуда появился знак минуса в связи?

Кривая обучения (learning curve)

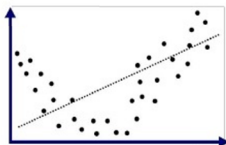
Типичная зависимость от N для параметрических моделей:



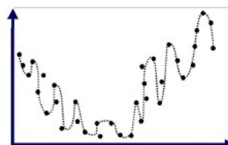
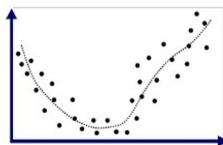
- Параметрические модели: $\#$ параметров не зависит от N .
- Непараметрические: $\#$ параметров растёт с N .

Проблемы недообучения и переобучения

- Недообучение: модель слишком простая для реальных данных.
 - не улавливает тонких закономерностей
- Переобучение: модель слишком сложная для реальных данных.
 - настраивается на шум в измерениях



underfitting



overfitting

Дополнительные ресурсы

Обучающие материалы

- [Документация sklearn](#): во-многом как учебник!
- Соревнования по машинному обучению [kaggle.com](#)
 - живые задачи на соревнованиях, форум с обсуждением идей
 - много данных и обучающих материалов
- [paperswithcode.com](#): описание/код SOTA методов
 - только нейросети

Данные

- [Репозиторий UCI](#):
 - классические небольшие датасеты (>450 задач).
- [openml.org](#): >5000 датасетов, есть большие
 - есть описания признаков, их статистики, распределения

Дополнительные ресурсы

Обучающие материалы

- [Документация sklearn](#): во-многом как учебник!
- Соревнования по машинному обучению [kaggle.com](#)
 - живые задачи на соревнованиях, форум с обсуждением идей
 - много данных и обучающих материалов
- [paperswithcode.com](#): описание/код SOTA методов
 - только нейросети

Данные

- [Репозиторий UCI](#):
 - классические небольшие датасеты (>450 задач).
- [openml.org](#): >5000 датасетов, есть большие
 - есть описания признаков, их статистики, распределения

Советую самим учиться и экспериментировать,
не ограничиваться материалами курса!

Этапы решения задачи⁹

- Этапы решения задачи машинного обучения:
 - Понять бизнес-проблему
 - Формализация задачи
 - Сбор данных
 - Предобработка данных
 - **Генерация признаков**
 - **Подбор модели**
 - **Оценка качества модели**
 - Внедрение модели
 - Поддержка модели

⁹Жирным выделены этапы на [kaggle.com](https://www.kaggle.com)

Теория и практика

- Неясные критерии качества модели
 - заказчик не определился с целями и бизнес-процессом
- Противоречивые критерии
 - например, доходность-риск в биржевой торговле
- Грязные данные
 - ошибки измерений, сбора и обработки
- Неполные данные
 - важные признаки не собираются
- Неструктурированные данные
 - например, отчеты испытаний в свободной текстовой форме
- Данные устаревают
 - важна регулярная адаптация модели

Темы курса

- Популярные классы моделей: метрические, линейные, деревья решений.
- Нелинейное обобщение линейных методов.
- Композиции алгоритмов.
- Предобработка данных, генерация признаков.
- Оценка качества моделей.
- Отбор признаков.
- Снижение размерности.
- Кластеризация.
- Детекция аномалий.
- Рекомендательные системы.
- Ранжирование.
- Специальные темы.

Обозначения в курсе

- **Объекты и целевые переменные:**

- x - вектор признаков (вход)
- y - предсказываемая величина, отклик (выход)
- x_i - i -й объект выборки X , y_i - i -й отклик Y .
- x^k - k -й признак объекта x
- x_i^k - k -й признак i -го объекта выборки x_i

- **Обучающая выборка:**

- X - матрица объектов (объекты x признаки), $X \in \mathbb{R}^{N \times D}$
- $Y \in \mathbb{R}^N$ - вектор откликов для каждого объекта

Обозначения в курсе

- **Количественные характеристики:**

- D - размерность признакового пространства: $x \in \mathbb{R}^D$
- N - число ($\#$) объектов обучающей выборки
- C - число классов в классификации.

- **Возможные классы:** $\{1, 2, \dots, C\}$ либо $\{\omega_1, \omega_2, \dots, \omega_C\}$

- **Оптимизация:**

- \hat{w} - оценка w на обучающей выборке, \hat{y} - оценка y , и т.д.
- $\mathcal{L}(\hat{y}, y)$ - функция потерь для одного объекта
 - y - истинный отклик, \hat{y} - прогноз.
- $L(w) = \sum_{n=1}^N \mathcal{L}(f_w(x_n), y_n)$ - функция потерь на всей выборке.

Обозначения

- **Специальные функции:**

- $\#[\text{объектов}]$ = число объектов, $\#[\text{признаков}]$ = число признаков
- $[x]_+ = \max\{x, 0\}$ - положительная срезка
- $\mathbb{I}[\text{условие}] = \begin{cases} 1, & \text{если условие выполнено} \\ 0, & \text{если условие не выполнено} \end{cases}$
- $\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$

- **Прочие обозначения:**

- $A \succcurlyeq 0$ неотрицательно определенная матрица A .
- Все вектора - столбцы.
- $O(\cdot)$ - асимптотическая сложность алгоритма.
 - Например $O(D)$ - сложность скалярного произведения D -мерных векторов

Заключение

- Задача машинного обучения найти такие w чтобы $f_w(x)$ точнее предсказывала y .
- Обучение бывает: с учителем, без учителя, частичное, привилегированное.
- Зависимость восстанавливается функцией $\hat{y} = f_{\hat{w}}(x)$ из класса $\{f_w(x), w \in W\}$.
- \hat{w} выбирается, чтобы минимизировать эмпирический риск

$$\frac{1}{N} \sum_{n=1}^N \mathcal{L}(f_w(x_n), y_n) \rightarrow \min_w$$

- Модели бывают
 - слишком простые (недообучение)
 - слишком сложные (переобучение)