

Выбор оптимальных моделей локальной аппроксимации для классификации временных рядов

Сергей Дмитриевич Иванычев

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Научный руководитель: д.ф.-м.н. В.В. Стрижов

Выпускная квалификационная работа бакалавра

Москва 2018

Классификация временных рядов

Цель

Предложить способ построения набора локально аппроксимирующих моделей для устойчивой классификации сигналов носимых устройств.

Гипотеза

Суперпозиция локально аппроксимирующих моделей доставляет более высокое качество при меньшей сложности чем универсальные модели.

Прямая задача

Требуется выбрать такой набор моделей локальной аппроксимации, что порождающая выборка в промежуточном пространстве является *простой*.

Обратная задача

Оптимизировать структурные параметры выбираемых моделей по порождающей выборке с целью получения выборки с оптимальными свойствами.

- Кузнецов М. П., Ивкин Н. П., *Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию*, 2015.
- Карасиков М. Е., Стрижов В. В. *Классификация временных рядов в пространстве параметров порождающих моделей*, 2016.
- Артемов А. В., *Математические модели временных рядов с трендом в задачах обнаружения разладки*, 2016.

Постановка задачи классификации

Задан временной ряд

$$S : T \rightarrow \mathbb{R}, \text{ где } T = \{t_0, t_0 + d, t_0 + 2d \dots\}.$$

Определен сегмент временного ряда

$$\mathbf{x}_i = [S(t_i), S(t_i - d), S(t_i - 2d), \dots, S(t_i - (n-1)d)]^T, \quad \mathbf{x}_i \in X \equiv \mathbb{R}^n.$$

X — набор сегментов данных акселерометра

y — метки классов движения (бег, ходьба, подъем и спуск по лестнице)

То есть задана выборка $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^I, \quad y_i \in \{1, 2, \dots, K\}$

h — конечный набор моделей локальной аппроксимации.

Постановка задачи классификации

Локально аппроксимирующая модель

$$g_i(w, x) \in X, \text{ где } w \in \mathbb{R}^{n_g}.$$

Оптимальные параметры определяются образом

$$\mathbf{h}_i(x) = \arg \min_{w \in \mathbb{R}^{n_g}} \rho(g(w, x), x),$$

\mathbf{h}_i — локально аппроксимирующая модель сегмента.

Набор функций $\mathbf{h} = [\mathbf{h}_1 \dots \mathbf{h}_k] : x \mapsto [w_1^* \dots w_k^*]$ отображает пространство сегментов \mathbf{X} в *промежуточное пространство* признаков описаний \mathbf{Z} .

Классификатор

$$T \rightarrow \mathbf{X} \xrightarrow{\mathbf{h}} \mathbf{Z} \xrightarrow{a} Y,$$

\mathbf{h} — набор моделей локальной аппроксимации, $a(\cdot, \gamma)$ — многоклассовый классификатор.

Постановка задачи классификации

Оптимизация функционалов качества локально аппроксимирующих моделей.

$$\arg \min_{w \in W} \sum_{i=1}^l \sum_{k=1}^n \|g(w, \mathbf{x}_i) - \mathbf{x}_i\|_2^2$$

Оптимизация функционала качества обобщенной линейной модели.

$$\arg \min_{\theta \in \Theta} \left[- \sum_{i=1}^l \sum_{k=1}^K [y_i = k] \log P(y_i = k | \mathbf{z}_i, \theta) \right]$$

Локально-аппроксимирующие модели

Модель	Структурные параметры
SEMR	-
AR-авторегрессия	порядок
Фурье-модель FFT	количество главных частот
Вейвлет-модель SSE	количество сингулярных чисел

AR-авторегрессия

Структурный параметр: порядок m ,

$$g_{AR}(w, x) = \hat{x}, \text{ где } \hat{x}_i = \begin{cases} x_k & \text{при } k \in [1, m], \\ w_0 + \sum_{i=1}^m w_i x_{k-i} & \text{при } k \in [m+1, n]. \end{cases}$$

Фурье-модель (SSA)

Структурный параметр: количество главных собственных значений k . Сингулярное разложение траекторной матрицы,

$$S^T S = V H V^T, H = \text{diag}(\lambda_1 \dots \lambda_m),$$

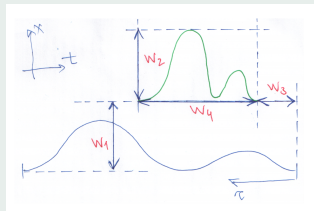
Параметры образуют k главных собственных значения.

Вейвлет-модель (FFT)

Структурный параметр: k частот из прямого преобразования Фурье, соответствующие наибольшим амплитудам

$$w_{2j} = \operatorname{Re} \sum_{k=1}^n x_k \exp \left(-\frac{2\pi i}{n} kj \right), \quad w_{2j+1} = \operatorname{Im} \sum_{k=1}^n x_k \exp \left(-\frac{2\pi i}{n} kj \right)$$

Self-Modeling Regression



$$g(x, w) = w_1 + w_2 p(w_3 + w_4 t),$$

$$w_{\text{SEMOR}} = [\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \rho].$$

- 1 Для каждого $\mathbf{h}_i \in \mathbf{h}$ вычисляем

$$[\mathbf{z}_i^1 \dots \mathbf{z}_i^k]^\top = [\mathbf{h}_i(\mathbf{x}_1) \dots \mathbf{h}_i(\mathbf{x}_k)]$$

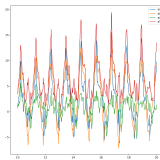
- 2 Конкатенируем вектора параметров $\mathbf{z}_i = (\mathbf{z}_i^1 \dots \mathbf{z}_i^k)$, то есть $\mathbf{z}_i = \mathbf{h}(\mathbf{x}_i)$. Получили выборку в промежуточном пространстве \mathbf{Z} .
- 3 Минимизируем функции потерь обобщенной линейной модели

$$\hat{\theta} = \arg \min_{\theta \in \Theta} L(f(\mathbf{Z}), \mathbf{y})$$

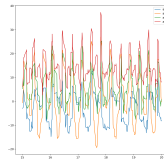
Данные с акселерометра: 4 типа движения, частота дискретизации 100 Гц.

Сегментация: локальные экстремумы с окном + квантиль по длине сегментов.

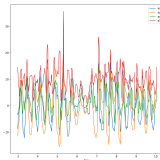
Нормализация: приведение к одной размерности с помощью кубических сплайнов.



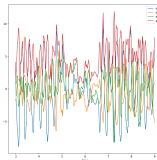
Ходьба



Бег



Вверх



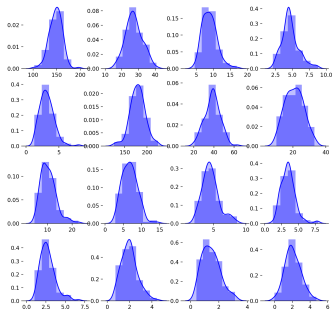
Вниз

Решение задачи: проверка простоты выборки

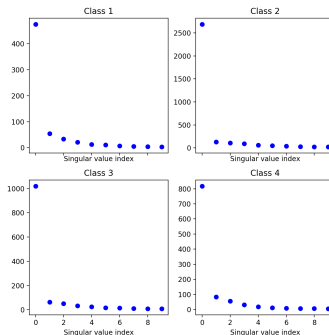
Данные: сегменты временного ряда акселерометра.

Тесты простоты выборки: (T -тест) $\mathbb{E}\varepsilon = 0$, $D\varepsilon = \text{const}$, а также

Анализ унимодальности распределений



Анализ спектра выборки

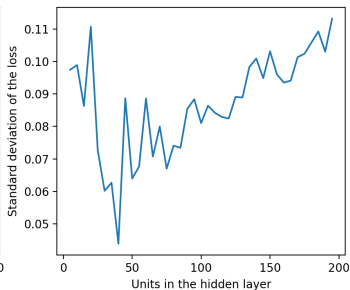
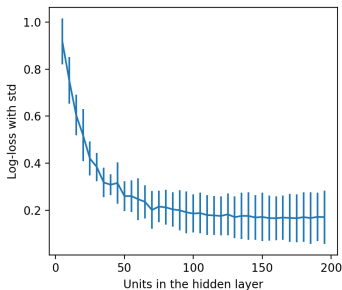


Универсальная модель

Определим сложность модели как количество нейронов на скрытом слое.

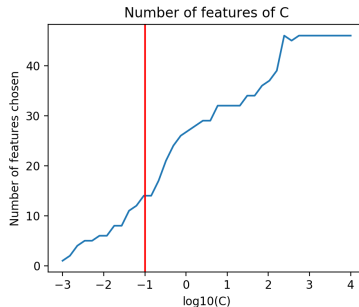
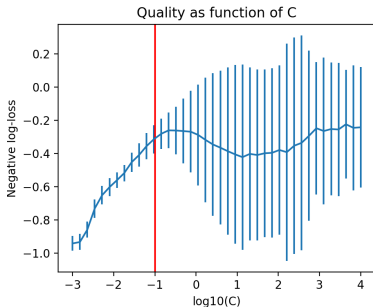
На выборке (\mathbf{X}, \mathbf{y}) обучаем универсальную модель — нейронную сеть, варьируя количество нейронов на скрытом слое.

Получаем зависимость качества и дисперсии качества от сложности (количества нейронов).



Обобщенная линейная модель: отбор признаков

Пользуемся логистической регрессией с L_1 регуляризацией.



Выбираем порог C , с устраивающим нас качеством и дисперсией. Эти признаки используем в обобщенной линейной модели.

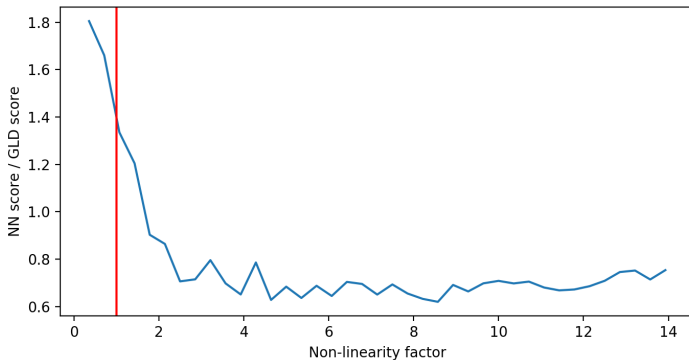


Рис.: Отношение ошибок от отношения сложностей

Получили, что при одинаковой сложности, универсальная модель показывает в 1.4 раз худшее качество, чем универсальная линейная модель.

- Выборка в промежуточном пространстве простая, а аппроксимирующие ее линейная модель являются адекватной.
- GLM адекватнее разделяет выборку чем универсальная модель, то есть при одинаковой сложности обеспечивает более высокое качество и меньше переобучается.