Классификация временных рядов

Цель

Предложить способ построения набора моделей локальной аппроксимации для устойчивой классификации сигналов носимых устройств.

Гипотеза

Суперпозиция моделей локальной аппроксимации доставляет более высокое качество при меньшей сложности чем универсальные модели.

Прямая задача

Исследование статистических свойств промежуточного параметрического пространства, строящегося моделями локальной аппроксимации.

Обратная задача

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

Постановка задачи классификации

Задан временной ряд

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

Определен сегмент временного ряда

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

Задана выборка $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}, \quad y_i \in \{1, 2, \dots K\}.$

$\mathbf{X}$ — набор сегментов данных акселерометра,

$\mathbf{y}$ — метки классов движения (бег, ходьба, подъем и спуск по лестнице).

Задан $\mathbf{h}$ — конечный набор моделей локальной аппроксимации.

The *local approximation model* g is a model, which approximates the time series $\mathbf{x}(t)$ within the time segment $\in [t, t-\Delta t]$:

$$g : [t, t-\Delta t] \to \hat{\mathbf{x}},$$

or element-vise,

$$g : \tau \mapsto \hat{\mathbf{s}}.$$

The model $g = g(\mathbf{b})$ has parameters $\mathbf{b}$ to optimize them according to the error function:

$$\hat{\mathbf{b}} = \arg\min S(\mathbf{b}) = \|\hat{\mathbf{x}} - \mathbf{x}\|.$$

The local approximation *method* generates the feature description $\boldsymbol{\varphi}$ (notation of C. Bishop) of the time tick t as the description of this approximation problem solution:

$$\boldsymbol{\varphi} = \boldsymbol{\varphi}(\mathbf{g}, \mathbf{b}, \mathbf{x}).$$

Note that the feature description is taken at the time t, so that the time series $\mathbf{x}_t$ produces the description $\boldsymbol{\varphi}_t$. A trivial example of the model $\mathbf{g}$ is identity function,

$$\boldsymbol{\varphi} = \mathbf{x} = \mathbf{g}(\mathbf{x}).$$

Постановка задачи классификации

Модель локальной аппроксимации

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

Оптимальные параметры определяются как

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

$\mathbf{h}_i$ — модель локальной аппроксимации.

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 признаковых описаний* $\mathbf{Z}$.

Модель классификации

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

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

Постановка задачи классификации

Минимизация функций ошибки каждой модели локальной аппроксимации

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

Оптимизация функции ошибки обобщенной линейной модели

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

Построение промежуточного пространства

Модели локальной аппроксимации

Модель	Структурные параметры
SEMOR	-
AR-авторегрессия	порядок
Фурье-модель (FFT)	количество главных частот
Сингулярного спектр (SSA)	количество сингулярных чисел

Модели локальной аппроксимации

AR-авторегрессия

Структурный параметр: порядок m,

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

Сингулярный спектр (SSA)

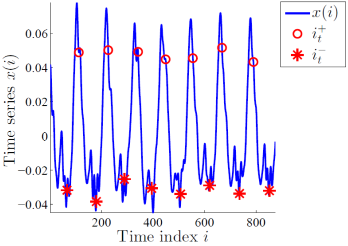
Структурный параметр: количество главных собственных значений k. Сингулярное разложение траекторной матрицы,

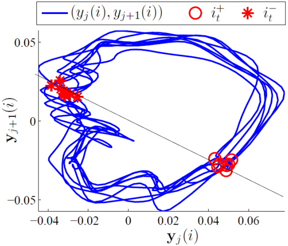
$$S^{\mathsf{T}} S = V H V^{\mathsf{T}}, H = \mathrm{diag}(\lambda_1 \ldots \lambda_m),$$

параметры образуют k главных собственных значения.

Find dissection of the trajectory of principal components $\mathbf{y}_j = \mathbf{H}\mathbf{v}_j$, where $\mathbf{H}$ is the Hankel matrix and $\mathbf{v}_j$ are its eigenvectors:

$$\frac{1}{N}\mathbf{H}^{\mathsf{T}}\mathbf{H} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathsf{T}}, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N).$$





Motrenko A.P., Strijov V.V. Extracting fundamental periods to segment human motion time series // IEEE Journal of Biomedical and Health Informatics, 2016, 20(6) : 1466 - 1476.

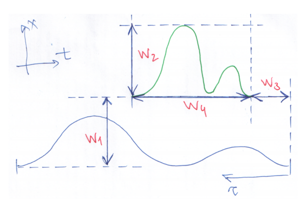
Фурье-модель (FFT)

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

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

Self-Modeling Regression

Накладывание шаблона линейными преобразованями x и t.



$$g(\mathbf{x}, \mathbf{w}) = w_1 + w_2 p(w_3 + w_4 t),$$

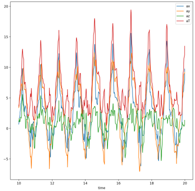
$$w_{\mathrm{SEMOR}} = [\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \rho].$$

Построение промежуточной выборки и оптимизация функции потерь обобщенной линейной модели

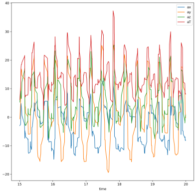
1. Для каждого $\mathbf{h}_i \in \mathbf{h}$ вычисляем
$$[\mathbf{z}_i^1 \dots \mathbf{z}_i^k]^\mathsf{T} = [\mathbf{h}_i(\mathbf{x}_1) \dots \mathbf{h}_i(\mathbf{x}_k)]$$
2. Конкатенируем вектора параметров $\mathbf{z}_i = (\mathbf{z}_1^i \dots \mathbf{z}_k^i)$, то есть $\mathbf{z}_i = \mathbf{h}(\mathbf{x}_i)$. Получили выборку в промежуточном пространстве $\mathbf{Z}$.
3. Минимизируем функции потерь обобщенной линейной модели
$$\hat{\theta} = \arg\min_{\theta \in \Theta} L\left(f(\mathbf{Z}), \mathbf{y}\right).$$

Решение задачи: генерация данных

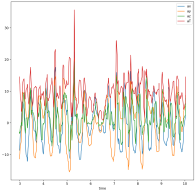
Данные с акселерометра: 4 типа движения, частота дискретизации 100 Гц.
Сегментация: локальные экстремумы с окном и квантиль по длине сегментов.
Нормализация: приведение к одной размерности с помощью кубических сплайнов.



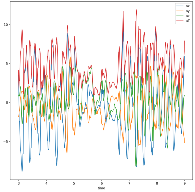
Ходьба



Бег



Вверх

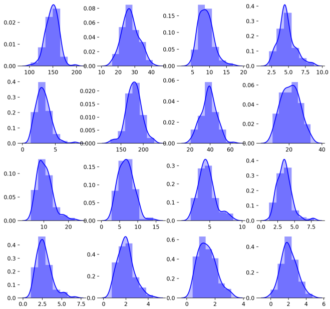


Вниз

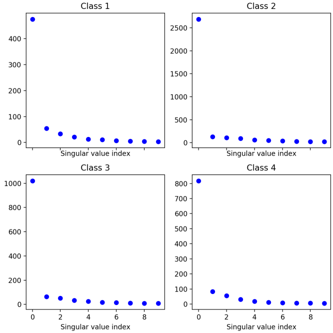
Решение задачи: проверка гипотезы простоты выборки

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

анализ унимодальности распределений



анализ спектра выборки

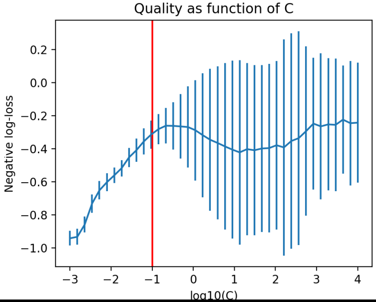


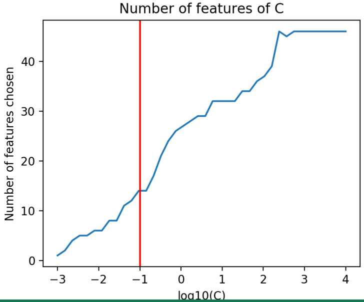
Обобщенная линейная модель: отбор признаков

Сравниваем обобщающую способность обобщенной линейной модели (GLE) с универсальной моделью при одинаковой сложности.
Определим сложность модели как

$$\mathrm{Comp}(\mu) = \#|\text{neurons in the hidden layer}|$$

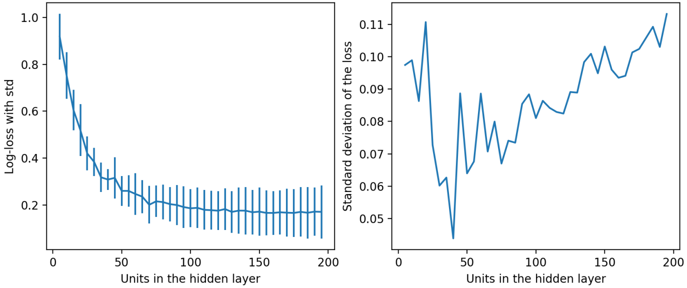
Отбираем признаки в $(\mathbf{Z}, \mathbf{y})$ для обобщенной линейной модели.
Логистическая регрессия с L_1 регуляризацией.





На выборке $(\mathbf{X}, \mathbf{y})$ оптимизируем параметры двуслойной нейронной сети (NN). Получаем зависимости

$$L(\mathrm{Comp}), D_L(\mathrm{Comp}).$$



Сравнение ошибки при разных сложностях универсальной модели

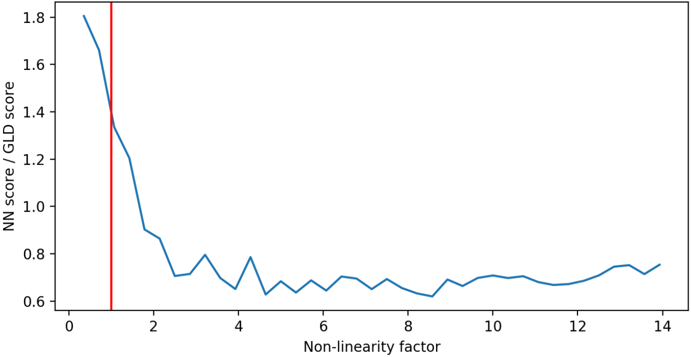


Рис.: Отношение ошибок от отношения сложностей

Результат: при $\mathrm{Comp}(\mathrm{GLE}) = \mathrm{Comp}(\mathrm{NN})$, имеем $\frac{L(\mathrm{NN})}{L(\mathrm{GLE})} = 1.4$, $\frac{D_L(\mathrm{NN})}{D_L(\mathrm{GLE})} > 1$.