Смеси моделей векторной авторегрессии в задаче прогнозирования временных рядов

Дмитрий Сергеевич Федоряка

Московский физико-тенический институт

20 апреля 2017 г.

Цель исследования

Цель

Строить предсказание периодических временных рядов методом векторной авторегрессии

Проблема

Линейная регрессия даёт низкую точность и переобучается

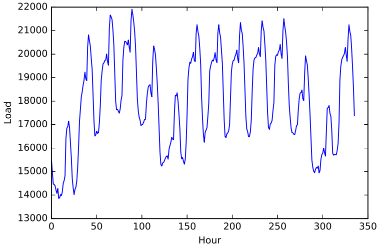
Решение

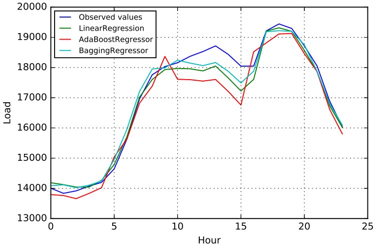
Построить композицию линейных моделей

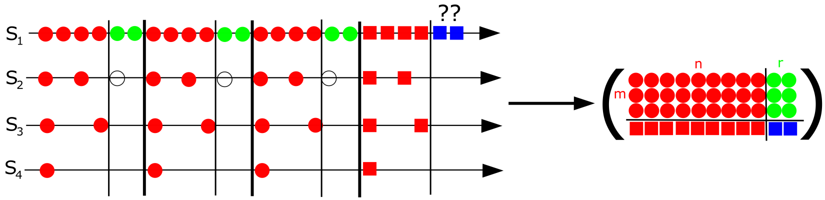
Литература

- Стрижов, Нейчев, Катруца — Выбор оптимального набора признаков из мультикоррелирующего множества в задаче прогнозирования, 2015
- Bishop — Pattern Recognition and Machine Learning, 2006
- Freund Y., Schapire R. E —Experiments with a new boosting algorithm, 1996

Работа в одном слайде







Постановка задачи

Набор временных рядов:

$$\mathfrak{D} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_p\}$$

$$\mathbf{s}_i = \{s_i(\Delta t_i), s_i(2\Delta t_i), \dots s_i(L_i \Delta t_i)\} = \{s_i^1, s_i^2, \dots s_i^{L_i}\}$$

Прогноз:

$$f(\mathfrak{D}) = \{\hat{s}_1^{L_1-r+1}, \hat{s}_1^{L_1-r+2}, \dots, \hat{s}_1^{L_1}\}$$

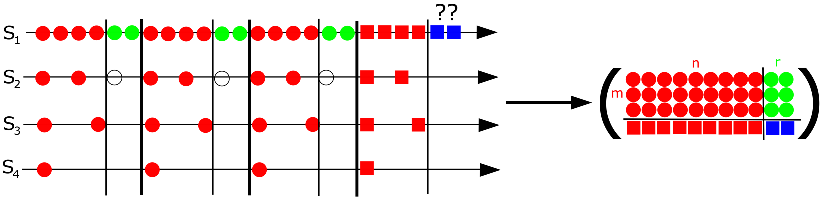
Функция потерь:

$$S(f) = \sum_{i=L_1-r+1}^{L_1} (s_1^i - \hat{s}_1^i)^2$$

Задача прогнозирования:

$$f^* = \arg\min_f S(f, \mathfrak{D})$$

Решение: сведение к задаче регрессии



$$f^* = \arg\min_f \sum_{j=1}^{m-1} ||f(X_i) - Y_i||_2^2$$

Решение

Линейная модель:

$$f(\mathbf{x}) = \mathbf{x}W$$

W - матрица коэффициентов:

$$W = X^+ Y = (X^T X)^{-1} X^T Y$$

Проблемы:

- низкая точность
- переобучение

Решение

Композиция моделей:

$$f(\mathbf{x}) = \sum_{i=1}^{K} \pi_i(\mathbf{x}) f_i(\mathbf{x})$$

Алгоритмы:

- Бэггинг и метод случайных подпространств
- Адаптивный бустинг
- Кластерная регресия

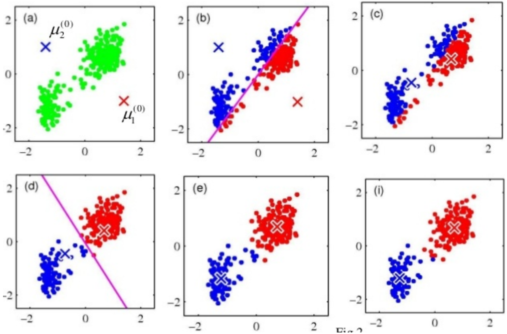
Кластерная регрессия

Гипотеза: обучающая выборка разбивается на кластеры, так что объекты i-го класса описываются некоторой своей моделью f_i:

$$\overline{1, m} = \bigsqcup_{j=1}^{K} C_j$$

Предложение: использовать эти модели как базовые модели смеси экспертов.

Алгоритм кластеризации K-means



Центры → модели

Поиск центров → обучение моделей

Ближайший центр → наиболее точно описывающая объект модель

Кластерная регрессия: обучение

$\forall j \in \overline{1, m}\ C_j = rand(1, m)$

для $iter \in \overline{1, N_{it}}$

  для $j \in \overline{1, K}$

    Обучаем модель f_j на объектах с индексами C_j.

  для $i \in \overline{1, m}$

    для $j \in \overline{1, K}$

      $S(i, j) = ||f_j(X_i) - Y_i||$

  для $i \in \overline{1, m}$

    для $j \in \overline{1, K}$

      $C_j = \{i \in \overline{1, m} | \underset{k}{argmin}(S(i, k)) = j\}$

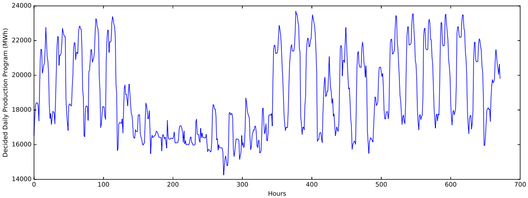
Кластерная регрессия: предсказание

Обучаем алгоритм f_0 на всей выборке.
$\tilde{Y}^0 = f_0(X^0)$ — предварительный ответ.
$\forall\, j \in \overline{1, K}\ \pi'_j = \mathfrak{E}(\frac{||f_j(X^0) - Y^0||}{||Y^0||})$ — вычисление вероятностей.
$\forall\, j \in \overline{1, K}\ \pi_j = \frac{\pi'_j}{\sum_{k=1}^{K} \pi'_k}$ — нормировка.
$Y^0 = \sum_{j=1}^{K} \pi_j \cdot f_j(X^0)$ – ответ.

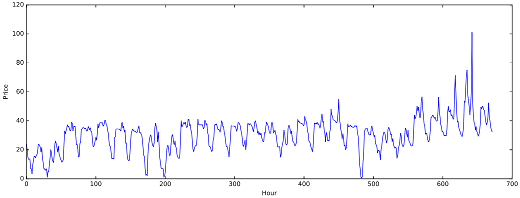
Примеры возможных функций $\mathfrak{E}(S)$: $\frac{1}{S+\varepsilon}$, e^{-S}, $1 - \frac{1}{1+e^{-S}}$

Эксперимент

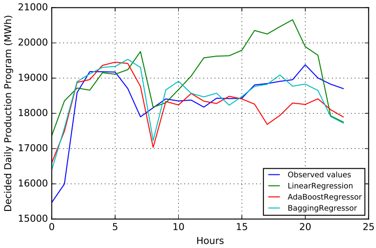
Турция - потребление энергии

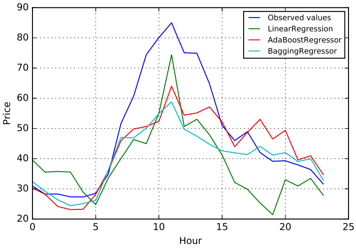


Германия - рыночные цены на электричество



Ошибки предсказаний

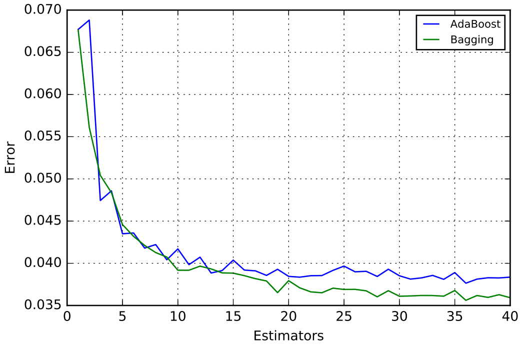




	Турция 1	Турция 2	Германия 1	Германия 2	Польша
Лин. рег.	6.3%	4.0%	30.1%	28.8%	2.0%
Бустинг	3.8%	3.6%	24.1%	23.0%	3.6%
Бэггинг	3.6%	3.0%	27.5%	26.4%	1.7%

$$MSE(\hat{s}) = \frac{\sqrt{< (s_1 - \hat{s}_1)^2 >}}{< s_1 >} = \frac{1}{\sqrt{r}} \frac{\sqrt{\sum_i (s_1^i - \hat{s_1^i})^2}}{\sum_i s_1^i}$$

Зависимость ошибки от числа базовых моделей



Заключение

- Задача прогнозирования временных рядов с периодичностью сведена к задаче регрессии.
- Задача регрессии решается с помощью линейной модели.
- Построены композиции моделей (бэггинг и бустинг), повышающие точность прогноза.
- Предложен эвристический итерационный алгоритм композиции моделей, который не оказался эффективным в данной задаче из-за его требований к размеру обучающей выборки и из-за свойств временных рядов.