

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ

(государственный университет)

ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ

Базовая организация - Учреждение Российской академии наук

Вычислительный центр им. А.А. Дородницына РАН

Кафедра «Интеллектуальные системы»

специализация «Интеллектуальный анализ данных»

Квалификационная работа на соискание степени бакалавра  
по направлению 010900 «Прикладная математика и физика»

**ПРОГНОСТИЧЕСКИЕ МУЛЬТИМОДЕЛИ  
РАЗНОМАСШТАБНЫХ ВРЕМЕННЫХ РЯДОВ  
ИНТЕРНЕТА ВЕЩЕЙ**

Выполнил:

студент 274 группы

*Нейчев Радослав Георгиев*

Научный руководитель:

д.ф.-м.н.

*Стрижов Вадим Викторович*

Москва, 2016

# Содержание

<b>1</b>	<b>Введение</b>	<b>3</b>
<b>2</b>	<b>Постановка задачи.</b>	<b>6</b>
<b>3</b>	<b>Построение матрицы объект-признак</b>	<b>7</b>
<b>4</b>	<b>Отбор признаков</b>	<b>9</b>
4.1	Постановка задачи отбора признаков . . . . .	9
4.2	Метод Белсли . . . . .	10
4.3	Метод отбора признаков Add-Del . . . . .	13
<b>5</b>	<b>Посроение мультимодели</b>	<b>15</b>
5.1	Смесь моделей . . . . .	16
5.2	Смесь экспертов . . . . .	19
<b>6</b>	<b>Вычислительный эксперимент</b>	<b>20</b>
6.1	Описание данных . . . . .	20
6.2	Описание эксперимента . . . . .	20
6.3	Результаты . . . . .	21
<b>7</b>	<b>Заключение</b>	<b>23</b>
	<b>Список литературы</b>	<b>24</b>

## Аннотация

В данной работе рассматривается поведение различных устройств, взаимодействующих друг с другом в рамках активно развивающегося в данный момент Интернета вещей. Множество датчиков отслеживает состояние устройства с течением времени, порождая временные ряды в различных шкалах. Информация, содержащаяся в данных рядах может быть использована для определения текущего состояния устройства и прогнозирования его поведения в ближайшее время. Предполагая, что перед сбоем внутренние параметры устройства меняются, его прогнозирование может помочь избежать поломки или минимизировать возможные потери. Решение данной задачи осложняется тем, что временные ряды могут обладать различными периодами отсчета, быть сильно скоррелированными или зашумленными. Для решения проблемы различных периодов отсчета предлагается процесс ресемплирования временных рядов оптимальным образом, приводящим отношение частот к рациональным числам. Для достижения лучших результатов предлагается использовать смеси моделей и смеси экспертов для описания данных. Область применимости данных подходов не ограничивается прогнозированием состояния устройств, они могут быть использованы для прогнозирования любых временных рядов. Вычислительный эксперимент проводился на данных о погоде и энергопотреблении в Польше [1].

# 1 Введение

**Актуальность темы.** Ввиду наличия большого количества датчиков, фиксирующих различные параметры устройств (а так же состояние окружающего мира), было бы выгодно использовать накопленную информацию для прогнозирования состояний этих устройств в будущем. Задача построения прогноза осложняется тем, что датчики, фиксирующие значения даже связанных между собой величин порождают данные различной природы. Вдобавок, частота отсчетов у различных датчиков не совпадает и может варьироваться от миллисекунд до нескольких часов. Данная работа предлагает подход к прогнозированию временных рядов, позволяющий работать с временными рядами различной природы и частоты.

В данной работе предполагается, что каждый из рассматриваемых временных рядов имеет фиксированную частоту отсчетов и достаточно длинную предысторию. Задача анализа разномасштабных временных рядов возникает в различных предметных областях, таких как диагностика заболеваний, прогнозирование погоды, анализ показаний различных датчиков оборудования [2–5]. Основной предпосылкой к анализу разномасштабных временных рядов является предположение о наличии сложных зависимостей между различными процессами, имеющими, возможно, различную частоту и природу. Обычно при анализе разномасштабных временных рядов, модель для каждого ряда строится в его собственной временной шкале. Подобный подход применяется в различных задачах классификации временных рядов, поиска аномалий и прогнозирования ошибок [4, 6, 7]. Для получения значений временных рядов с меньшей частотой как правило производится усреднение [8], а для повышения частоты отсчетов временного ряда применяется операция дифференцирования [9]. Операции усреднения и дифференцирования, являющиеся по-сути применением вейвлет-преобразования Хаара [9], могут быть заменены любыми другими подобными вейвлет-преобразованиями [10]. При работе с разномасштабными временными рядами как правило определяется оптимальная частота дискретизации [3, 11].

В случаях, когда необходимо получить прогноз на несколько шагов вперед, возможно использование итеративного, прямого, или же ММО-подхода [12]. В случае итеративного подхода, прогноз строится поэтапно с применением одной и той же модели: на следующем шаге прогноз используется как реальные исторические данные.

Подобный подход достаточно прост вычислительно и в реализации, но ему свойственно накопление ошибки с каждым шагом прогнозирования. Прямой подход подразумевает построение собственной модели для каждого шага вперед [13]. Этим обеспечивается более высокая точность и отсутствие накопления ошибки, но обучение нескольких моделей приводит к значительному повышению вычислительной сложности относительно итеративного подхода. Последний рассматриваемый подход, ММО (Multiple input, multiple output) позволяет строить прогноз сразу на несколько шагов вперед, но его реализация может быть значительно сложнее, чем для случая с одним шагом прогнозирования. В работе [12] производится сравнение всех трех подходов при построении прогноза на несколько шагов вперед с помощью SVR (метод опорных векторов для задачи регрессии). Согласно результатам этой работы, прямой и ММО подходы доставляют большую точность независимо от длины прогноза, причем ММО-подход дает лучшие результаты в большинстве случаев.

В данной работе задача прогнозирования рассматривается как задача регрессии и для ее решения используются следующие подходы: многомерная линейная регрессия, случайный решающий лес (Random Forest) [14, 15], метод опорных векторов [16–18] и искусственные нейронные сети [19, 20]. Случайный лес использует для решения задач классификации и регрессии набор из большого числа случайно сгенерированных решающих деревьев, что дает лучшие результаты, чем использование одного сложного дерева [21]. При решении задачи регрессии, узлы деревьев разделяют признаковое пространство на два подпространства, а каждый лист представляет собой некоторую регрессионную модель, которая используется для построения прогноза если признаковое описание предыстории попало в соответствующую область признакового пространства. Прогнозы всех деревьев усредняются, или же, в более общем случае, усреднение производится в соответствии с весами (вероятностью, что его результат является верным) каждого дерева. Благодаря использованию случайных подвыборок признаков при решении задач регрессии и классификации, случайный лес показывает хорошие результаты в случаях, когда размерность признакового пространства высока. Также случайный лес обладает высокой обобщающей способностью благодаря своей структуре: его можно рассматривать как бустинг над композицией простых моделей. Метод опорных векторов (SVR) также показывает

хорошие результаты при прогнозировании временных рядов [16, 18]. Данный подход обладает высокой устойчивостью при работе с данными высокой размерности. Также возможно его обобщение на случай ММО-подхода [22]. Искусственные нейронные сети позволяют моделировать достаточно сложные зависимости между данными без непосредственного расширения признакового пространства. Их популярность при решении самых разных задач [20, 23] во многом обуславливается именно этим их качеством.

Более подробно перечисленные методы решения задачи прогнозирования рассматриваются в работах [23, 24]. Подходы к работе с разномасштабными временными рядами также приводятся в работах [6, 10, 25–27].

**Цель работы.** Предложить комплексный подход к прогнозированию временных рядов, позволяющий использовать дополнительные временные ряды различной природы и частоты при построении прогноза для улучшения его качества.

**Научная новизна.** Применение операции ресемплирования для построения матрицы объект-признак, позволяющее работать с временными рядами различной природы.

Использование мультимоделей в прогнозировании, что позволяет описать достаточно сложные зависимости с помощью простых моделей.

**Практическая ценность.** На основе анализа прогнозов достаточной точности можно сделать полезные выводы. Например, в промышленности своевременный прогноз нештатной ситуации может помочь избежать сбоя или минимизировать потери. Или же прогноз может помочь оптимизировать некоторые процессы и повысить прибыль.

Ввиду того, что используемые подходы достаточно общие, их применимость не ограничивается только прогнозированием внутренних параметров различных устройств, они применимы для построения прогнозов любых временных рядов.

**Положения, выносимые на защиту:**

- Предложен комплексный подход к прогнозированию временных рядов, позво-

ляющий работать с разномасштабными данными.

- Теоретически обоснована применимость мультимodelей к задаче прогнозирования и произведена практическая проверка данного подхода в рамках вычислительного эксперимента.

## 2 Постановка задачи.

Пусть задан набор временных рядов  $\mathfrak{D} = \{\mathbf{s}^{(q)} \mid q = 1, \dots, Q\}$ , каждый временной ряд  $\mathbf{s}$  имеет вид

$$\mathbf{s} = [s_1, \dots, s_i, \dots, s_T], \quad s_i = s(t_i), \quad 0 \leq t_i \leq t_{\max}$$

и представляет собой набор значений  $s_i = s(t_i)$  некоторой зависящей от времени функции  $s(t)$ . Пусть каждый временной ряд  $\mathbf{s}^{(q)}$  имеет свой период  $\tau^{(q)}$ :

$$t_i^{(q)} = i \cdot \tau^{(q)}.$$

Используя имеющиеся временные ряды из набора  $\mathfrak{D}$  необходимо спрогнозировать значения целевого ряда  $\mathbf{s}_1$  на некоторый отрезок времени вперед (Рис. 1):  $\hat{s}_1(t_i)$  при  $\Delta t_r < t_i \leq T_{\max} + \Delta t_r$  с минимальной ошибкой. В данной работе в качестве функции ошибки используется симметричная версия MAPE:

$$SMAPE(\mathbf{s}, \hat{\mathbf{s}}) = \frac{1}{r} \sum_{i=1}^r \frac{2|s_i - \hat{s}_i|}{|s_i + \hat{s}_i|}. \quad (1)$$

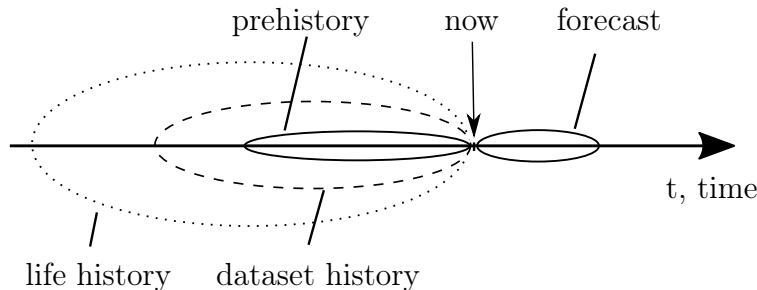


Рис. 1: Представление временного ряда в рамках задачи прогнозирования.

При построении прогноза логичным шагом является использование информации, содержащейся в других рядах (в случае их наличия). Поэтому задача построения прогноза рассматривается как задача регрессии. Для ее решения необходимо перейти к признаковому описанию временных событий в виде матрицы объект-признак.

### 3 Построение матрицы объект-признак

Пусть объект — вектор, содержащий предысторию временного ряда длиной  $\Delta t_p$ , ответ — вектор содержащий следующие  $\Delta t_r$  значений ряда. Тогда можно ввести следующее понятие:

**Определение:** Семпл — отрезок последовательных значений временного ряда длиной  $(\Delta t_r + \Delta t_p)$ .

Для построения матрицы объект-признак предлагается применить операцию семплирования временных рядов.

Пусть задан временной ряд  $\mathbf{s}$ . Возможны следующие случаи:

- Ряд  $\mathbf{s}$  обладает ярко выраженной периодикой. В таком случае предлагается использовать семплы, не перекрывающие друг друга. Начала семплов располагаются в соответствии с периодом данного ряда.
- Ряд  $\mathbf{s}$  обладает слабо выраженной периодикой или имеет несколько различных периодов. В данном случае семплы могут перекрываться для достижения лучших результатов. Начала семплов также располагаются в соответствии с периодикой.
- Если же ряд  $\mathbf{s}$  не обладает периодикой, сильно зашумлен или же отсутствует информация о природе описываемых данных, предлагается использовать семплы, начала которых распределены случайным образом равномерно по временной шкале рассматриваемого временного ряда.

В общем случае, способ семплирования зависит от природы данных и должен выбираться в соответствии с их происхождением.

Перед операцией семплирования временной ряд нормируется таким образом, чтобы все его значения лежали в интервале  $[0, 1]$ :

$$s_i = \frac{s_i^{old} - subt}{div}, \quad subt = \min(s^{old}), \quad div = \max(s^{old} - subt)$$

Пусть для временного ряда  $\mathbf{s}$  заданы частота семплирования  $\frac{1}{\tau}$ , длина предыстории  $\Delta t_p$  и горизонт прогнозирования  $\Delta t_r$ . Определим аппроксимирующую функцию



следующим образом:

$$g(\mathbf{s}, t_i) = \mathbf{s}(\hat{t}), \quad (2)$$

где  $\hat{t}$  — ближайший к  $t_i$  момент времени, в котором определено значение ряда  $\mathbf{s}$ . То есть временной ряд аппроксимируется кусочно-постоянной функцией, принимающей значение  $\mathbf{s}(\hat{t})$  на полуинтервале  $[\hat{t} - T; \hat{t} + T)$ , где  $2T$  — период временного ряда  $\mathbf{s}$  (Рис. 2). В таком случае для порождения семпла размера  $l$  достаточно задать его конец  $t_i$ , по которому с помощью преобразования будет получен семпл (Рис. 3):

$$\{g(t_j^*)\}_{j=1}^l, \quad t_j^* \in [t_i - (\Delta t_p + \Delta t_r), t_i], \quad j = \{1, \dots, l\}. \quad (3)$$

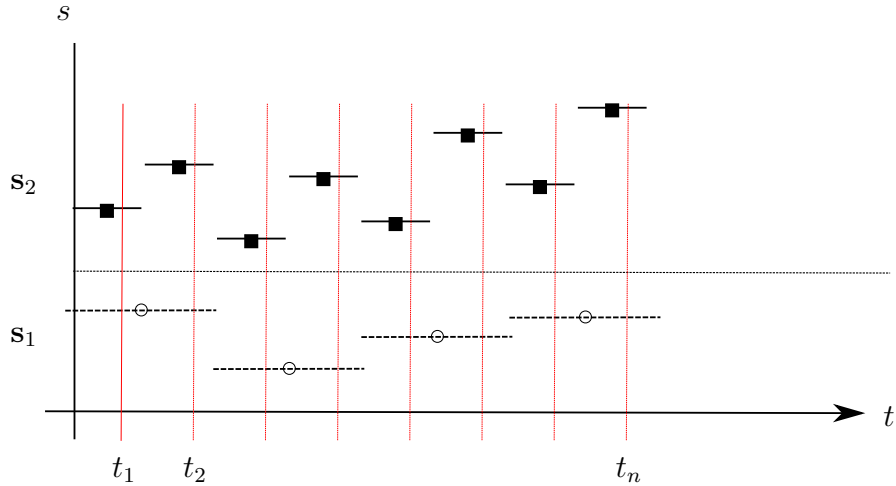


Рис. 2: Процесс ресемплирования временного ряда.

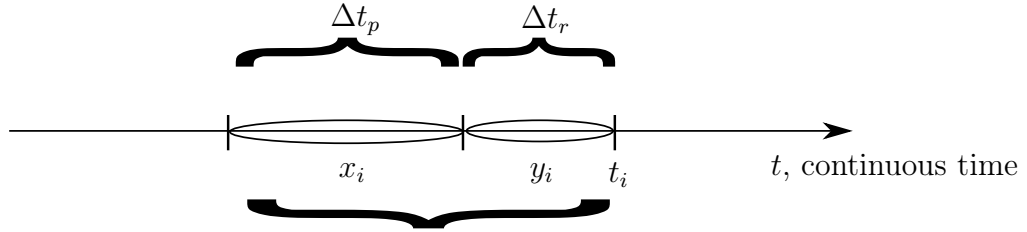


Рис. 3: Порождение пары объект-ответ.

Таким образом, порожденные семплы — новые объекты  $\mathbf{x}'$  и ответы  $\mathbf{y}'$  к ним. Объект — вектор, содержащий первые  $\Delta t_r$  элементов семпла, следующие  $\Delta t_p$  элементов — ответ (в терминах авторегрессионной матрицы).

$$\underbrace{[(t_i), \dots, s(t_i - \Delta t_r)]}_{\mathbf{y}'_i}, \dots, \underbrace{[s(t_i - \Delta t_r - \Delta t_p)]}_{\mathbf{x}'_i} = [\mathbf{y}'_i, \mathbf{x}'_i],$$

$$\mathbf{X}^* = \left[ \begin{array}{c|ccc} \hat{\mathbf{y}}' & \mathbf{x}'_0 & \mathbf{x}''_0 & \dots \\ \hline \mathbf{y}'_1 & \mathbf{x}'_1 & \mathbf{x}''_1 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}'_m & \mathbf{x}'_m & \mathbf{x}''_m & \dots \end{array} \right] = \left[ \begin{array}{c|c} \hat{\mathbf{y}} & \mathbf{x}_0 \\ \hline \mathbf{Y}_{1 \times r} & \mathbf{X}_{1 \times n} \\ \hline \mathbf{Y}_{m \times r} & \mathbf{X}_{m \times n} \end{array} \right].$$

При наличии нескольких временных рядов  $2, \dots, j$  матрица объект-признак включает в себя конкатенацию объектов всех рассматриваемых временных рядов между собой  $[\mathbf{x}' \cup \mathbf{x}'' \cup \dots]$ , а в качестве ответов рассматриваются соответственные части семплов целевого ряда  $\mathbf{y}'$ .

В результате применения данного подхода может быть получено нужное количество объектов, признаковое описание которых может быть расширено благодаря применению различных преобразований, выбираемых в зависимости от решаемой задачи.

## 4 Отбор признаков

### 4.1 Постановка задачи отбора признаков

Представим матрицу ответов  $\mathbf{Y}$  в виде

$$\mathbf{y} = f(\mathbf{X}, \mathbf{w}) + \boldsymbol{\varepsilon}(\mathbf{X}),$$

где  $\boldsymbol{\varepsilon}$  — матрица регрессионных остатков. Пару  $(\mathbf{X}, \mathbf{Y})$  назовем выборкой и обозначим как  $\mathcal{D}$ . Признаки, которым соответствуют ненулевым строкам матрицы  $\mathbf{W}$ , будем называть *активными*, а остальные признаки будем считать *исключенными*. Множество индексов элементов выборки обозначим  $\mathcal{I}$  и разобьем на непересекающиеся подмножества:  $\mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$ .

Зафиксируем функцию ошибки  $S(\mathbf{f}(\mathbf{X}, \mathbf{W}), \mathbf{Y})$ . Подмножество индексов активных признаков обозначим  $\mathcal{C}$ , где  $\mathcal{I}$  есть множество индексов всех признаков. Назовем *моделью* пару  $(f, \mathbf{w})$  и обозначим ее как  $f$ . При фиксированной модели необходимо найти множество индексов  $\mathcal{C}$ , минимизирующее функцию ошибки  $S$  на элементах выборки  $\mathcal{D}_{\mathcal{C}}$ , состоящей из элементов выборки  $\mathcal{D}$  с индексами из множества  $\mathcal{C}$ :

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subseteq \mathcal{I}} S(\mathcal{A} | \mathbf{w}^*, \mathcal{D}_{\mathcal{C}}). \quad (4)$$

Запись вида  $S(|\mathcal{D})$  означает, что выборка  $\mathcal{D}$  фиксирована, а выборка  $\mathcal{L}$  меняется. Для решения задачи регрессии требуется найти такие параметры  $\mathbf{W}^*$ , которые доставляют минимум функции ошибки  $S$  на элементах выборки  $\mathcal{D}_{\mathcal{L}}$ , определяемой аналогично  $\mathcal{D}$ :

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} S(\mathbf{W} | \mathcal{D}_{\mathcal{L}}, \mathcal{A}), \quad (5)$$

## 4.2 Метод Белсли

Между признаками возможно существование мультиколлинеарной зависимости. Например, если временной ряд обладает посуточной периодичностью, его значение в один и тот же час сильно коррелирует со значениями в соседние часы. В случае мультиколлинеарности признаков оценка параметров (5) является неустойчивой. Для устранения данной проблемы необходимо найти мультиколлинеарные признаки. Применим для этого метод Белсли. Рассмотрим сингулярное разложение матрицы  $\mathbf{X}$ :

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \quad (6)$$

$\mathbf{U}$  и  $\mathbf{V}$  — ортогональные матрицы,  $\mathbf{D}$  — диагональная матрица, состоящая из сингулярных чисел  $\lambda_j$  таких, что

$$\lambda_1 < \lambda_2 < \dots < \lambda_r,$$

где  $r$  — ранг матрицы  $\mathbf{D}$ . В нашей задаче мы предполагаем, что матрица невырождена, поэтому  $r = n$ . Предположим, что столбцы  $\mathbf{Y}_{*k}$  матрицы  $\mathbf{Y}$  являются независимыми гауссовскими векторами с вектором средних  $(\mathbf{X}\mathbf{W})_{*k}$  с матрицей ковариаций  $\sigma_k^2 \mathbf{I}_m$  соответственно. В высказанных предположениях, матрица ковариаций вектора  $\mathbf{W}_{*k}$  равняется  $\mathbf{D}(\mathbf{W}_{*k}) = \sigma_k^2 (\mathbf{X}^T \mathbf{X})^{-1}$ . Положим  $\mathbf{A} = \mathbf{X}^T \mathbf{X}$ . Сингулярные числа  $\lambda_i$ ,  $1 \leq i \leq n$  являются собственными числами матрицы  $\mathbf{A}$ , а столбцы матрицы  $\mathbf{V}$  — собственными векторами матрицы  $\mathbf{A}$ . Используя соотношение (6), получаем следующее выражение для матрицы  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{X}^T \mathbf{X} = \mathbf{V}\mathbf{D}^T \mathbf{U}^T \mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^T \mathbf{D}\mathbf{V}^T. \quad (7)$$

Определим как  $i$ -ый индекс обусловленности отношение  $\eta_i = \lambda_{\max}/\lambda_i$ , где  $\lambda_{\max}$  — максимальное сингулярное число. Большое значение  $\eta_i$  указывает на близкую к

линейной зависимость между признаками. Максимальный индекс обусловленности матрицы  $\mathbf{X}$  характеризует, насколько велико будет изменение компонент вектора параметров  $\mathbf{w}$  при изменении матрицы признаков  $\mathbf{X}$ . Назовем его *числом обусловленности*  $\theta$ . Заметим, что число обусловленности матрицы  $\hat{\mathbf{A}}$  есть квадрат числа обусловленности матрицы  $\mathbf{X}$ .

В рамках нашей задачи для нахождения максимально коррелирующих признаков необходимо найти индекс  $i^*$  вида

$$i^* = \operatorname{argmax}_{i \in \mathcal{A}} \eta_i, \quad (8)$$

где  $\mathcal{A}$  — текущее множество активных признаков.

Оценками дисперсии параметров будут диагональные элементы матрицы  $\hat{\mathbf{A}}^{-1}$ .

$$\mathbf{var}(\mathbf{w}) = (\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{V}^T)^{-1} \mathbf{D}^{-2} \mathbf{V}^{-1} = \mathbf{V} \mathbf{D}^{-2} \mathbf{V}^T, \quad (9)$$

$$\mathbf{var}(w_j) = \sum_{i=1}^n v_{ij}^2 / \lambda_i^2,$$

где  $[v_{ij}] = \mathbf{V}$ . *Дисперсионные доли*  $q_{ij}$  определим как вклад  $j$ -го признака в дисперсию  $i$ -го элемента вектора параметров  $\mathbf{w}$ :

$$q_{ij} = \frac{v_{ij}^2 / \lambda_j^2}{\sum_{j=1}^n v_{ij}^2 / \lambda_j^2}. \quad (10)$$

Из определения дисперсионных долей следует, что их большие значения указывают на наличие зависимости между признаками. Определим индекс  $j^*$ , вносящий максимальный вклад в дисперсию  $i$ -го элемента вектора  $\mathbf{w}$ :

$$j^* = \operatorname{argmax}_{j \in \mathcal{A}} q_{i^*j}, \quad (11)$$

где максимальный индекс обусловленности  $i^*$  определяется (8).

Будем называть модель  $\mathbf{f}$  *неустойчивой*, если число обусловленности матрицы признаков  $\mathbf{X}$  велико:  $\theta \gg 1$ .

Исключение из модели  $\mathbf{f}$  параметров, максимально влияющих на минимальное сингулярное значение матрицы  $\mathbf{X}$ , доставляет максимальную устойчивость модели.

Действительно, пусть  $N$  объектов описываются моделью  $\mathbf{f}$ , состоящей из  $n$  признаков  $\chi_i$ . Некоторые признаки мультиколлинеарны. Обозначим максимальный индекс обусловленности  $\eta_{\max}$ , из ковариационной матрицы  $\hat{\mathbf{A}}$  найдем максимальную

соответствующую ему дисперсионную долю. Пусть она соответствует признаку  $\chi_{j^*}$ . Это указывает на принадлежность признака  $\chi_{j^*}$  к множеству мультикоррелирующих признаков. Мультиколлинеарность влечет близость к нулю одного или нескольких сингулярных значений  $\lambda_k$ .

Следовательно, исключение  $j^*$ -го признака из модели  $\mathbf{f}$  приведет к увеличению минимального сингулярного числа и уменьшению числа обусловленности  $\theta'$  новой матрицы  $\mathbf{X}'$ . Меньшее число обусловленности характеризует лучшую устойчивость модели, а значит, исключение  $j^*$ -го признака ее повышает.

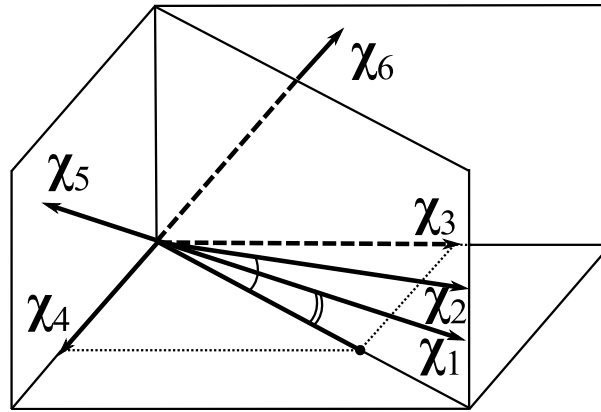


Рис. 4: Мультикоррелирующие, ортогональные векторы значений свободных переменных.

Для лучшего понимания принципа работы метода Белсли, рассмотрим его на конкретном примере. На Рис. 4 приведены 3 объекта, каждый из которых описывается шестью признаками  $\chi_1, \dots, \chi_6$ . Расстоянием от объекта до признака будем считать величину угла между ними. Табл. 1 содержит разложение вектора дисперсий  $\mathbf{var}(\chi_j)$  по индексам обусловленности  $\eta_i$ . На Рис. 5 приведена визуализация оценки ковариационной матрицы параметров  $\hat{\mathbf{A}}$ .

Согласно Табл. 1, максимальный индекс обусловленности  $\eta_6 = 4.87$ . В соответствующей ему шестой строке дисперсионные доли в первом и втором столбцах (выделены красным) максимальны. Они соответствуют признакам  $\chi_1$  и  $\chi_2$ , между которыми есть ярко выраженная зависимость, что наглядно видно на Рис. 4.

Таблица 1: Разложение  $\text{var}(\chi_j)$

	Индекс обусловленности	Дисперсия параметров					
		$\text{var}(\chi_1)$	$\text{var}(\chi_2)$	$\text{var}(\chi_3)$	$\text{var}(\chi_4)$	$\text{var}(\chi_5)$	$\text{var}(\chi_6)$
$\eta_1$	0	0	0	0	0	0	0
$\eta_2$	0	0	0	0	0	0	0
$\eta_3$	0	0	0	0	0	0	0
$\eta_4$	1	0.02	0.05	0	0.2	0.01	0.53
$\eta_5$	2.29	0.08	0.1	0.15	0.39	0.48	0.23
$\eta_6$	4.87	0.9	0.85	0.84	0.4	0.5	0.24

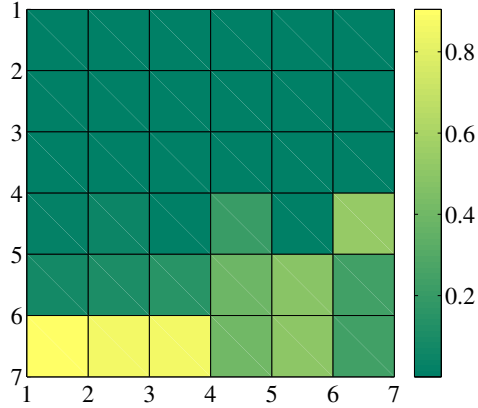


Рис. 5: Визуализация оценки ковариационной матрицы параметров  $\hat{\mathbf{A}}$ .

### 4.3 Метод отбора признаков Add-Del

Применим метод Белсли для отбора признаков. Предлагаемый метод состоит из двух этапов **Add** и **Del**. Метод предполагает, что регрессионные остатки имеют нормальное распределение с нулевым матожиданием и произвольной неотрицательно определенной ковариационной матрицей. В начальный момент множество активных признаков  $\mathcal{A}_0 = \emptyset$ . Рассмотрим работу алгоритма.

Этап **Add**. Выбираем признак  $j^*$  из множества исключенных признаков согласно

$$j^* = \arg \min_{j \in \mathcal{J} \setminus \mathcal{A}} S(\mathbf{w} | \mathcal{A} \cup \{j\}, \mathfrak{D}) \quad (12)$$

и добавляем его к модели:

$$\mathcal{A} = \mathcal{A} \cup \{j^*\}. \quad (13)$$

Признаки добавляются к активному набору  $\mathcal{A}$ , пока функция ошибки не удовлетворит критерию останова:

$$S(\mathbf{w}_{\mathcal{A}}|\mathcal{D}) \geq S_{min} + \delta S_1. \quad (14)$$

Этап **Del**. В соответствии с критерием (11) находим признак  $j^*$ , максимально коррелирующий с другими, и удаляем его из множества :

$$\mathcal{A} = \mathcal{A} \setminus \{j^*\}. \quad (15)$$

Повторяем данную операцию, пока полученная ошибка  $S(f_{\mathcal{A}_k}|\mathbf{w}^*, \mathcal{D})$  превосходит свое минимальное значение на данном этапе не более, чем на заданную величину  $\delta S_2$ . Критерий останова этапа *Del* на данном шаге

$$S(\mathbf{w}_{\mathcal{A}}|\mathcal{D}_{\mathcal{L}}) \geq S_{min} + \delta S_2. \quad (16)$$

Этапы **Add** и **Del** повторяются до тех пор, пока значение ошибки не стабилизируется, то есть ее значение будет слабо изменяться от итерации к итерации. Величина колебания значения от итерации к итерации должна составлять величину порядка  $\frac{1}{100}S$ , более конкретную величину можно подобрать опытным путем.

Остановимся подробнее на критериях останова (14),(16). Для определения величин  $\delta S_{1,2}$  рассмотрим свойства суммы квадратов регрессионных остатков

$$\mathbf{S} = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}. \quad (17)$$

Вектор регрессионных остатков  $\boldsymbol{\varepsilon}$ , определяется выражением

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\mathbf{w} = \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y}, \quad (18)$$

где матрица  $(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) = \mathbf{P}$  симметрична и идемпотентна:  $\mathbf{P}^T = \mathbf{P}$  и  $\mathbf{P}^2 = \mathbf{P}$ .

Подставляя выражение (18) в (17) и учитывая, что

$$\mathbf{w}^T = \mathbf{y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1},$$

получаем

$$\mathbf{S} = \mathbf{y}^T \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y} = \mathbf{y}^T \mathbf{y} - \mathbf{w}^T \mathbf{X}^T \mathbf{y}. \quad (19)$$

Здесь  $S$  записана как квадратичная форма вектора  $\mathbf{y}$ . В предположении, что

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\mathbf{w}, \sigma^2 \mathbf{I}), \quad (20)$$

математическое ожидание  $\mathbf{S}$  имеет вид [28].

Так как наиболее правдоподобная оценка параметров  $\mathbf{w}$  при предположении (20) имеет вид  $\mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ , и функция ошибки в этом случае задана как  $S(\mathbf{w}) = \mathbf{S}$ , то математическое ожидание суммы квадратов регрессионных остатков имеет вид

$$E(S(\mathbf{w})) = \text{tr} \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \mathbf{I} \sigma^2 + (\mathbf{X} \mathbf{w})^\top \left( \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \mathbf{X} \mathbf{w}. \quad (21)$$

Так как след идемпотентной матрицы (в данном случае это матрица Мура-Пенроуза  $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ ), то

$$\begin{aligned} E(S(\mathbf{w})) &= \text{rank} \left( (\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{I} \right) \sigma^2 = (m - \text{rank}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)) \sigma^2 \\ &= (m - \text{rank}(\mathbf{X})) \sigma^2 = (m - n) \sigma^2, \end{aligned}$$

здесь  $m$  — число элементов выборки и строк матрицы  $\mathbf{X}$ . Если матрица плана  $\mathbf{X}$  не содержит коллинеарных столбцов и её ранг  $\text{rank}(\mathbf{X}) = n$ , то несмещенной оценкой  $\sigma^2$  является оценка

$$\hat{\sigma}^2 = \frac{S(\mathbf{w})}{m - n}.$$

Поэтому величины  $\delta S_{1,2}$  можно определить как дисперсию вектора регрессионных остатков  $S$ .

## 5 Построение мультимодели

**Определение:** Модель — отображение  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , где  $\mathcal{X}$  есть пространство признаков, а  $\mathcal{Y}$  — пространство ответов.

В таком случае для фиксированной модели можно представить задачу регрессии (Рис. 6) следующим образом:

$$\hat{\mathbf{y}} = f(\mathbf{x}, \hat{\mathbf{w}}), \quad \hat{\mathbf{w}} = \arg \min_{\mathbf{w}} S(\mathbf{w} | f(\mathbf{x}, \mathbf{w}), \mathbf{y}), \quad (22)$$

где  $\mathbf{w}$  — параметры модели.

В качестве функции ошибки  $S$  используется *SMAPE* (1) каждой пары объект-ответ  $[\mathbf{x}_i | \mathbf{y}_i]$ , которая затем усредняется по всем парам  $i = 1, \dots, m$  тестовой выборки:

$$S(\mathbf{w} | f(\mathbf{x}, \mathbf{w}), \mathbf{y}) = \frac{r}{m} \sum_{i=1}^m \text{SMAPE}(\mathbf{y}_i, f(\mathbf{x}_i, \mathbf{w})).$$



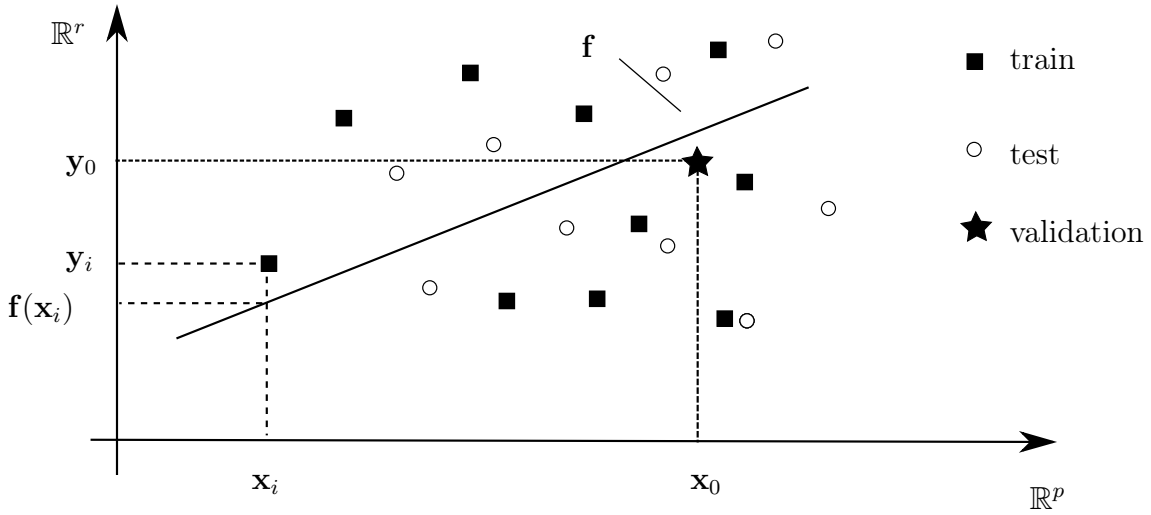


Рис. 6: Задача регрессии.

## 5.1 Смесь моделей

Пусть вектор ответов  $y$  описывается с помощью модели  $f$  с некоторой ошибкой  $\varepsilon$ , распределенной нормально:

$$y = f(\mathbf{x}, \mathbf{w}) + \varepsilon, \quad f(\mathbf{x}, \mathbf{w}), \quad \varepsilon \sim \mathcal{N}(0, \beta) \Rightarrow y \sim \mathcal{N}(f(\mathbf{x}, \mathbf{w}), \beta),$$

В таком случае оценка максимального правдоподобия для  $\mathbf{y}$  имеет вид:

$$\hat{\mathbf{y}} = f(\mathbf{x}, \hat{\mathbf{w}}), \quad \hat{\mathbf{w}} = \operatorname{argmax}_{\mathbf{w}} \frac{1}{2\beta} \sum_{i=1}^m (y_i - f(\mathbf{x}, \mathbf{w}))^2$$

Ввиду наличия множества разнородных временных рядов, предлагается использовать для их описания смесь моделей. Пусть задано  $K$  моделей  $f_k(\mathbf{x}, \mathbf{w}_k)$  и вектор ответов  $y$  порождается одной из них. В вероятностной постановке распределение вектора ответов будет смесью нормальных гауссовских распределений:

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{y}|f(\mathbf{x}, \mathbf{w}_k), \beta) = \quad (23)$$

$$\sum_{k=1}^K \frac{1}{(2\pi\beta_k)^{n/2}} \exp\left(-\frac{1}{2\beta_k} (\mathbf{y} - f(\mathbf{x}, \mathbf{w}_k))^T (\mathbf{y} - f(\mathbf{x}, \mathbf{w}_k))\right).$$

За  $\pi_i$  обозначены веса моделей  $\boldsymbol{\pi} = [\pi_1, \dots, \pi_k]$ , а за  $\boldsymbol{\theta}$  — вектор гиперпараметров

$$\boldsymbol{\theta} = [\mathbf{w}_1, \dots, \mathbf{w}_k, \boldsymbol{\pi}, \beta]^T,$$

где  $\mathbf{B} = \beta \mathbf{I}_m$  — ковариационная матрица вектора ответов.

**Определение гиперпараметров:** Необходимо найти такие гиперпараметры  $\hat{\boldsymbol{\theta}}$ , чтобы функция правдоподобия достигала своего максимального значения:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}), \quad \ln p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^m \ln \left( \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{y}|f(x_i, \mathbf{w}_k), \beta) \right). \quad (24)$$

Для вычисления оценки максимального правдоподобия (24) вектора гиперпараметров  $\boldsymbol{\theta}$  смеси моделей (23), введем скрытые индикаторные переменные

$$Z = [\mathbf{z}_1, \dots, \mathbf{z}_m], \quad z_{ik} \in \{0, 1\},$$

$$z_{ik} = 1 \Leftrightarrow y_i \sim \mathcal{N}(f(\mathbf{x}_i, \mathbf{w}_k), \beta).$$

Тогда функция правдоподобия  $p(\mathbf{y}|\mathbf{x}, Z, \boldsymbol{\theta})$  примет вид

$$\begin{aligned} p(\mathbf{y}|X, Z, \boldsymbol{\theta}) &= \sum_{i=1}^m \sum_{k=1}^K z_{ik} (\ln \pi_k + \ln \mathcal{N}(y_i|f(x_i, \mathbf{w}_k), \beta)) = \\ &= \sum_{i=1}^m \sum_{k=1}^K z_{ik} \left( \ln \pi_k - \frac{1}{2\beta} (y_i - f(x_i, \mathbf{w}_k))^2 + \frac{n \ln \beta}{2} + \text{const} \right). \end{aligned}$$

Ввиду того, что в  $p(\mathbf{y}|\mathbf{x}, Z, \boldsymbol{\theta})$  появилась зависимость от случайных величин  $z_{ik}$ , для нахождения оптимальных гиперпараметров можно максимизировать матожидание правдоподобия по  $Z$

$$\begin{aligned} \mathbb{E}_Z[p(\mathbf{y}, Z|\mathbf{x}, \boldsymbol{\theta})] &= \\ &= \sum_{i=1}^m \sum_{k=1}^K \gamma_{ik} \left( \ln \pi_k - \frac{1}{2\beta} (y_i - f(x_i, \mathbf{w}_k))^2 + \frac{n \ln \beta}{2} \right), \end{aligned}$$

где  $\gamma_{ik} = \mathbb{E}[z_{ik}|\mathbf{y}, \mathbf{x}]$ .

Для нахождения оптимальной оценки вектора гиперпараметров  $\hat{\boldsymbol{\theta}}$ , максимизирующей  $\mathbb{E}_Z[p(\mathbf{y}, Z|\mathbf{x}, \boldsymbol{\theta})]$  предлагается использовать итеративный двухшаговый алгоритм Expectation-Maximization.

**Е-шаг:** Сгруппируем величины  $\gamma_{ik}$  в виде матрицы  $\Gamma = [\gamma_{ik}]$ . Элемент этой матрицы стоящий на пересечении строки  $i$  и столбца  $j$  есть апостериорная вероятность того, что  $i$ -ая пара объект-ответ описывается  $j$ -ой моделью. Тогда новая оценка  $\gamma_{ij}^{r+1}$  имеет вид

$$\begin{aligned} \gamma_{ik}^{(r+1)} &= \mathbf{E}(z_{ik}) = p(k|\mathbf{x}_i, \boldsymbol{\theta}^{(r)}) = \\ &= \frac{\pi_k \mathcal{N}(y_i | f(\mathbf{x}_i, \mathbf{w}_k^{(r)}), \beta^{(r)})}{\sum_{k'=1}^K \pi_{k'} \mathcal{N}(y_i | f(\mathbf{x}_i, \mathbf{w}_{k'}^{(r)}), \beta^{(r)})}. \end{aligned} \quad (25)$$

Обозначим функцию правдоподобия за  $Q^{(r)}(\boldsymbol{\theta})$

$$\begin{aligned} Q^{(r)}(\boldsymbol{\theta}) &= \mathbf{E}_Z(\ln p(\mathbf{y}, Z|\boldsymbol{\theta})) = \sum_{i=1}^m \sum_{k=1}^K \gamma_{ik}^{(r+1)} \\ &\left( \ln \pi_k^{(r)} + \ln \mathcal{N}(y_i | f(\mathbf{x}_i, \mathbf{w}_k^{(r)}), \beta^{(r)}) \right). \end{aligned} \quad (26)$$

**М-шаг:** Вычисляется новая оценка вектора гиперпараметров  $\boldsymbol{\theta}$  путем максимизации функции  $Q^{(r)}(\boldsymbol{\theta})$  при фиксированных значениях  $\gamma_{ik}^{r+1}$ . В первую очередь пересчитываются веса моделей  $\pi_k$

$$\pi_k^{(r+1)} = \frac{1}{n} \sum_{i=1}^m \gamma_{ik}^{(r+1)}.$$

Здесь учитывается тот факт, что на сумму весов всех моделей наложено ограничение  $\sum_{k=1}^K \pi_k = 1$ .

Теперь зафиксируем  $\pi_k$  и будем максимизировать  $Q^{(r)}$  по параметрам модели  $\mathbf{w}_k$ . Максимизацию выражения (26) с фиксированными весами моделей можно представить в виде

$$\mathbf{w}_k^{(r+1)} = \operatorname{argmax}_{\mathbf{w}_k} \sum_{i=1}^m -\gamma_{ik}^{(r+1)} (y_i - f(\mathbf{x}_i, \mathbf{w}_k))^2,$$

С учетом найденной оценок вектора параметров модели  $w$ , можно вычислить новую оценку вектора  $\beta$

$$\beta_k^{(r)} = \operatorname{argmax}_{\beta} \sum_{i=1}^m \gamma_{ik}^{(r+1)} \left( n \ln \beta - \frac{1}{\beta} (y_i - f(\mathbf{x}_i, \mathbf{w}_k^{(r+1)}))^2 \right).$$

**Критерий останова:** Алгоритм останавливается, если вектор гиперпараметров перестает значительно изменяться от итерации к итерации.

## 5.2 Смесь экспертов

Пусть каждая модель  $f_k(x, w_k)$  описывает пару объект-ответ  $(\mathbf{x}, \mathbf{y})$  с некоторой вероятностью  $p(k|f_k(\mathbf{x}, \mathbf{w}_k))$

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^K p(\mathbf{y}, k|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^K p(k|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{y}|k, \mathbf{x}, \boldsymbol{\theta})$$

Полученное выражение аналогично (23) с той лишь разницей, что веса моделей, в роли которых выступают вероятности  $p(k|\mathbf{x}, \boldsymbol{\theta})$ , зависят от объектов. В рамках принятых обозначений,  $\pi_k(x) = p(k|\mathbf{x}, \boldsymbol{\theta})$ . На Рис. ?? приведен пример использования смеси двух экспертов.

Полагая ошибки каждой модели  $f_k(\mathbf{x}, \mathbf{w}_k)$  распределенными нормально, то есть  $p(\mathbf{y}|f(\mathbf{x}, \mathbf{w}_k)) = \mathcal{N}(\mathbf{y}|f(\mathbf{x}, \mathbf{w}_k), \beta)$ , распределение вектора ответов можно представить как смесь распределений

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{v}_k) \mathcal{N}(\mathbf{y}|f_k(\mathbf{x}, \mathbf{w}_k), \beta), \quad (27)$$

где

$$\pi_k(\mathbf{x}, \mathbf{v}_k) = \frac{\exp(\mathbf{v}_k^\top \mathbf{x})}{\sum_{k'=1}^K \exp(\mathbf{v}_{k'}^\top \mathbf{x})}.$$

Ввиду того, что смесь экспертов (27) отличается от смеси моделей (23) лишь наличием зависимости  $\pi_k(x)$ , EM-алгоритмы для решения данных задач будут отличаться лишь на M-шаге при пересчете величин  $\gamma_{ik}$ .

---

**Algorithm 1:** EM-алгоритм для смеси экспертов.

---

**begin**

**Data:**  $(\mathbf{x}_i, y_i)$ ,  $i = 1, \dots, m$ . Parameters: number of experts  $K$ .

**Result:** Parameters  $\boldsymbol{\theta}$  of the model (27).

Initialize  $[\mathbf{w}, \beta, \mathbf{v}] \equiv \boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$ ,  $r = 0$

**while**  $\boldsymbol{\theta}$  keeps changing **do**

**E step:** compute hidden variables  $\gamma_{ik}^{(r+1)}$ , the expectation of the indicator variables, using (25)

**M step:** find new parameter estimates  $\mathbf{v}_k^{(r+1)} = \operatorname{argmax}_{\mathbf{v}} Q_k^{(r), \mathbf{v}}(\mathbf{v})$ , where

$$Q_k^{(r), \mathbf{v}}(\mathbf{v}) = \sum_{i=1}^m \gamma_{ik}^{(r+1)} \ln \pi_k(\mathbf{x}_i, \mathbf{v})$$

$\mathbf{w}_k^{(r+1)} = \operatorname{argmax}_{\mathbf{w}_k} Q_k^{(r), \mathbf{w}}(\mathbf{w}_k)$ , where

$$Q_k^{(r), \mathbf{w}}(\mathbf{w}_k) = \sum_{i=1}^m \gamma_{ik}^{(r+1)} (y_i - f(\mathbf{x}_i, \mathbf{w}_k))^2,$$

$\beta_k^{(r+1)} = \operatorname{argmax}_{\beta} Q_k^{(r), \beta}(\beta)$ , where

$$Q_k^{(r), \beta}(\beta) = \left( n \ln \beta - \frac{1}{\beta} (y_i - f(\mathbf{x}_i, \mathbf{w}_k^{(r+1)}))^2 \right)$$

---

## 6 Вычислительный эксперимент

### 6.1 Описание данных

Вычислительный эксперимент проводился на данных о потреблении электроэнергии в Польше за 1999-2004 года (Energy-Weather) [1]. Они представляют собой временные ряды различной природы и частоты дискретизации, что удовлетворяет предположениям о разнородности используемых временных рядов. Данные включают в себя

- Почасовые значения потребления электроэнергии (52512 наблюдений)
- Шесть временных рядов, фиксирующих природные показатели: максимальная и минимальная температуры, влажность, количество осадков, сила ветра, облачность (2188 наблюдений).

На Рис.7 приведена матрица объект-признак  $\mathbf{X}$  для данных об электроэнергии.

### 6.2 Описание эксперимента

Для решения задачи регрессии использовались следующие модели:

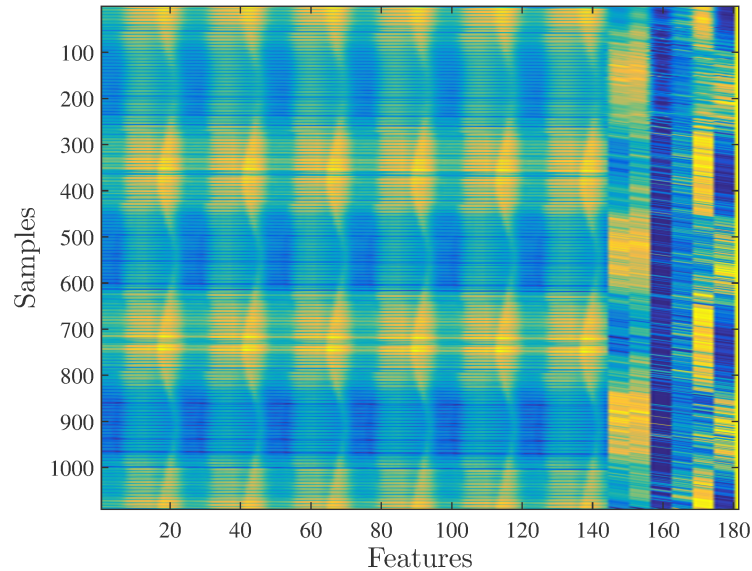


Рис. 7: Матрица объект-признак для данных об электроэнергии.

- MLR — многомерная линейная регрессия с  $l_2$  регуляризацией.
- MSVR — метод опорных векторов для задачи регрессии.
- ANN — Искусственные нейронные сети (один скрытый слой, 25 нейронов в скрытом слое)
- RF — Случайный лес, представляющий собой простой случай смеси моделей.

В ходе эксперимента производилось прогнозирование каждого из рядов на основе всех рядов. При настройке параметров моделей для каждого ряда учитывалась только ошибка прогноза целевого ряда. На Рис. 8 приведены ответы для объектов матрицы объект-признак.

### 6.3 Результаты

В Табл. 2 приведены результаты прогноирования каждого из рядов с применением различных моделей.

На Рис. 9 приведен пример прогноза потребления электроэнергии на сутки вперед. На Рис. ?? приведен долгосрочный прогноз, полученный с помощью случайного леса (Random Forest).

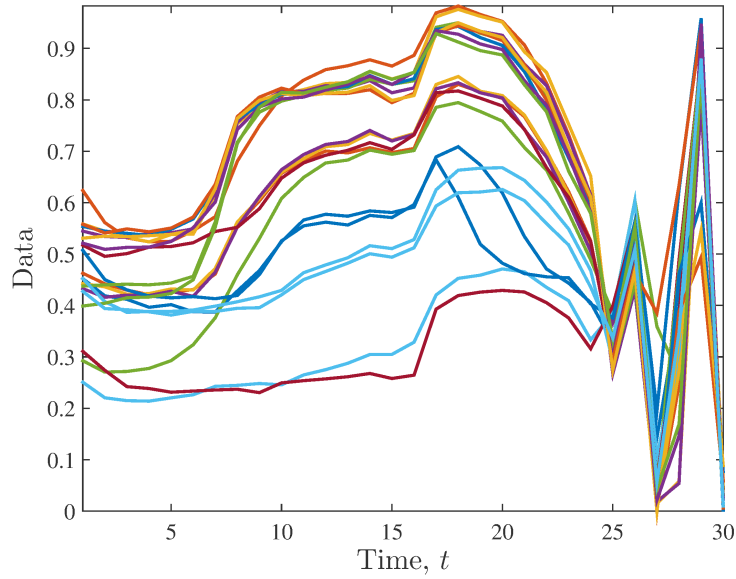


Рис. 8: Векторы ответов для всех рядов.

SMAPE	Data	Energy		Max T.		Min T.		Precipitation		Wind		Humidity		Solar	
Features	Models	test	train	test	train	test	train	test	train	test	train	test	train	test	train
History	Baseline	0.1948	0.2095	0.1040	0.1351	0.1047	0.1141	1.2034	1.2908	0.4581	0.4600	0.1803	0.1918	0.4641	0.5184
4*History	MLR	0.130	0.143	0.090	1.265	0.395	0.239	0.673	0.108	0.079	0.057	1.164	0.358	0.168	1.247
	MSVR	0.280	0.377	0.227	1.243	0.415	0.369	0.780	0.025	0.055	0.033	1.013	0.114	0.059	0.318
	RF	0.137	0.180	0.111	1.306	0.417	0.257	0.473	0.047	0.043	0.030	1.031	0.214	0.102	0.283
	ANN	0.157	0.180	0.102	2.050	0.596	0.281	1.526	0.089	0.102	0.076	1.418	0.358	0.146	0.661

Таблица 2: SMAPE прогнозов для различных временных рядов.

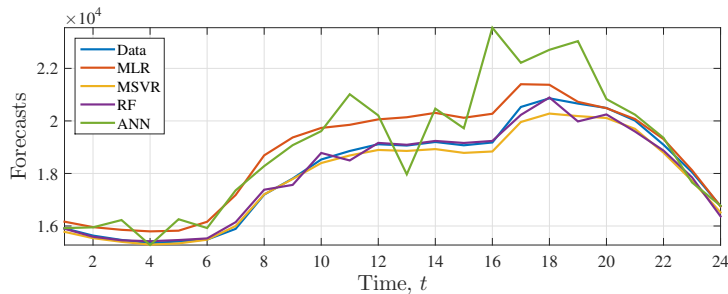


Рис. 9: Прогноз потребления электроэнергии на сутки вперед.

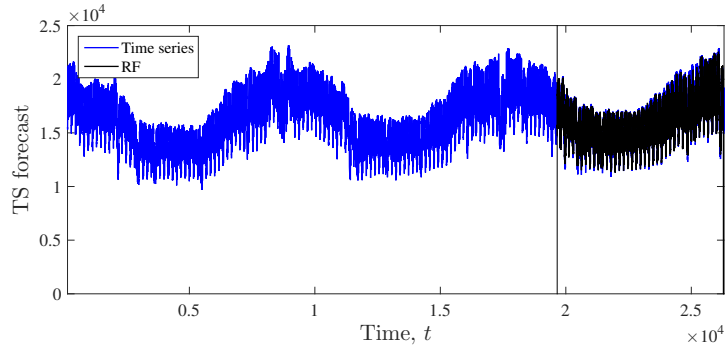


Рис. 10: Прогноз потребления электроэнергии на сутки вперед с помощью смеси моделей.

Полученные результаты согласуются с предположением о том, что смесь моделей способна качественно описывать сложные зависимости в данных.

## 7 Заключение

В данной работе предложен комплексный подход к прогнозированию временных рядов. Задача прогнозирования формулируется как задача регрессии. Для построения матрицы объект-признак используется операция семплирования, что позволяет работать с временными рядами с различной частотой дискретизации. В случае работы с временными рядами различной природы, возможно наличие сложных зависимостей между ними. Их описание с помощью единой модели затруднительно. Для достижения лучших результатов предлагается использование смеси моделей и смеси экспертов. В ходе вычислительного эксперимента произведена практическая проверка данного подхода на реальных данных в ходе вычислительного эксперимента. Результаты данной работы были использованы при подготовке доклада для конференции ICDM 2016.



## Список литературы

- [1] <http://gdudek.el.pcz.pl/varia/stlf-data>.
- [2] Madalena D. Costa, Chung-Kang Peng, and Ary L. Goldberger. Multiscale analysis of heart rate dynamics: Entropy and time irreversibility measures. *Cardiovascular Engineering*, 8(2):88–93, 2008.
- [3] M. U. Ahmed, N. Rehman, D. Looney, T. M. Rutkowski, P. Kidmose, and D. P. Mandic. Multivariate entropy analysis with data-driven scales. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3901 – 3904, 2012.
- [4] Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa. Multi-scale internet traffic forecasting using neural networks and time series methods. *Expert Systems*, 29(2):143–155, 2012.
- [5] Marco A. R. Ferreira, David M. Higdon, Herbert K. H. Lee, and Mike West. Multi-scale and hidden resolution time series models. *Bayesian Analysis*, 1(4):947–967, 2006.
- [6] Zhicheng Cui, Wenlin Chen, and Yixin Chen. Multi-scale convolutional neural networks for time series classification. *Computer Vision and Pattern Recognition*, 2016.
- [7] Chris Aldrich and Lidia Auret. *Process Monitoring and Fault Diagnosis with Machine Learning Methods (Advances in Computer Vision and Pattern Recognition)*, chapter Process Monitoring Using Multiscale Methods, pages 341–369. Springer London, 2013.
- [8] Shuen-De Wu, Chiu-Wen Wu, Shiou-Gwo Lin, Chun-Chieh Wang, and Kung-Yen Lee. Time series analysis using composite multiscale entropy. *Entropy*, 15(3):1069–1084, 2013.
- [9] Ying Jiang, C.-K. Peng, and Yuesheng Xu. Hierarchical entropy analysis for biological signals. *Journal of Computational and Applied Mathematics*, 236:728 [U+FFFF] 742, 2011.

- [10] Hongmei Chen, Brani Vidakovic, , and Dimitri Mavris. Multiscale forecasting method using armax models. Technical report, Georgia Institute of Technology, 2004.
- [11] Ugo Vespier, Arno Knobbe, Siegfried Nijssen, and Joaquin Vanschoren. *Machine Learning and Knowledge Discovery in Databases*, volume 7524 of *Lecture Notes in Computer Science*, chapter MDL-Based Analysis of Time Series at Multiple Time-Scales, pages 371–386. 2012.
- [12] Yukun Bao, Tao Xiong, and Zhongyi Hu. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, 129:482–493, 2014.
- [13] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Ji-Wen Yang, and Fan-Zhang Li. Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, 99(1):411 [U+FFFD]422, 2013.
- [14] Ruiyun Yu, Yu Yang, Leyou Yang, Guangjie Han, and Oguti Ann Move. Raq[U+FFFD]a random forest approach for predicting air quality in urban sensing systems sensors 2016. *Sensors*, 16(1):86, 2016.
- [15] Michael J. Kane, Natalie Price, Matthew Scotch, and Peter Rabinowitz. Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. *BMC Bioinformatics*, 15(276), 2014.
- [16] Theodore B. Trafalis and Huseyin Ince. Support vector machine for regression and applications to financial forecasting. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN2000)*, pages 348–353, 2000.
- [17] Raymundo Navarrete and Divakar Viswanath. Support vector regression, smooth splines, and time series prediction, 2015.
- [18] Wei Hao and Songnian Yu. *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management. Proceedings of PROLAMAT 2006, IFIP TC5 International Conference, June 15 [U+FFFD]17, 2006, Shanghai, China*, chapter Support Vector Regression for Financial Time Series Forecasting, pages 825–830. 2006.

- [19] Enzo Busseti, Ian Osband, and Scott Wong. Compared kernalized regression and 3 types of nn using the data from kaggle competition [U+FFFD] global energy forecasting competition 2012 - load forecasting [U+FFFD]. Technical report, Stanford University, 2012.
- [20] Graham W. Taylor and Geoffrey E. Hinton. Factored conditional restricted boltzmann machines for modeling motion style. *ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1025–1032, 2009.
- [21] A. Criminisi, J. Shotton, and E. Konukoglu. *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*, volume 7 of *Foundations and Trends in Computer Graphics and Vision*, chapter Regression Forests, pages 131–148. 2011.
- [22] F. Pérez-Cruz, G. Camps-Valls, E. Soria-Olivas, J. Pérez-Ruixo, A. Figueiras-Vidal, and A. Artés-Rodríguez. Multi-dimensional function approximation and regression estimation. *Artificial Neural Networks –ICANN*, pages 796–796, 2002.
- [23] Xueheng Qiu, Nanyang, Le Zhang, Ye Ren, P. N. Suganthan, and Gehan Amaratunga. Browse conference publications > computational intelligence in ... help working with abstracts ensemble deep learning for regression and time series forecasting. In *Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on*, 2014.
- [24] Aditya Grover, Ashish Kapoor, and Eric Horvitz. A deep hybrid model for weather forecasting. In *KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 379–386, 2015.
- [25] Bangzhu Zhu. A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network. *Energies*, 5:355–370, 2012.
- [26] Yun Bai, Zhiqiang Chen, Jingjing Xie, and Chuan Li. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology*, 532:193–206, 2015.

- [27] Stefano Ferrari, Francesco Bellocchio, Vincenzo Piuri, and N. Alberto Borghese. Hierarchical approach for multiscale support vector regression. *IEEE Transactions on Neural Networks Learning Systems*, 23(9):1448–1460, 2012.
- [28] F. Kachapova I. Kachapov. Orthogonal projection in teaching regression and financial mathematics. *Journal of Statistics Education*, 18(1):1–18, 2010.