

# Выбор мультимodelей в задачах классификации

А. А. Адуенко

Диссертация на соискание ученой степени  
кандидата физико-математических наук

01.01.09 — Дискретная математика и математическая кибернетика

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

Московский физико-технический институт  
(государственный университет)

2017 г.

# Мультимоделирование в задачах классификации

Исследуется проблема аппроксимации выборок, для которых не выполняется гипотеза о простоте. Решается задача повышения качества аппроксимации путем построения мультимоделей.

## Цели исследования

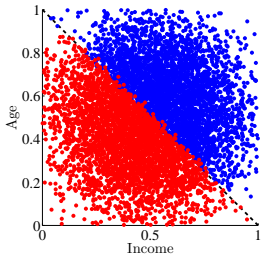
- 1 Разработка статистического подхода к задаче сравнения моделей.
- 2 Построение и теоретическое обоснование свойств функции сходства плотностей апостериорных распределений, решающей задачу сравнения моделей.
- 3 Разработка методов выбора адекватных мультимоделей.

## Задачи

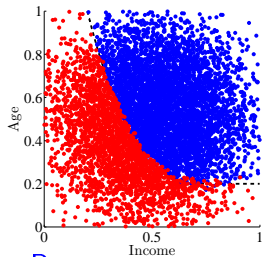
- 1 Разработать алгоритмы построения адекватных мультимоделей. Получить оценки максимального числа моделей в адекватной мультимодели.
- 2 Предложить метод совместного обучения и отбора признаков для смеси моделей.
- 3 Проверить улучшение качества и интерпретируемости мультимоделей, установить границы применимости предлагаемых методов.

# Мультимоделирование в задачах классификации

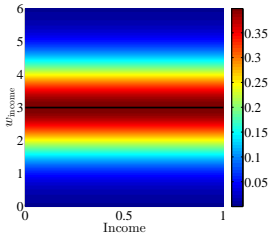
**Проблема:** выборка  $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}, i \in \mathcal{I} = \overline{1, m}, \mathbf{x}_i \in \mathbb{X}, y_i \in \mathbb{Y}$  не соответствует гипотезе порождения данных из одиночной модели.



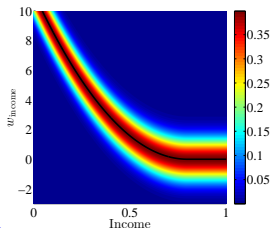
Модель порождения данных



Реальные данные



Модельное апостериорное распределение  $p(\mathbf{w}|\mathbf{X}, \mathbf{y})$



Реальное апостериорное распределение  $p(\mathbf{w}|\mathbf{X}, \mathbf{y})$

# Мультимодели: Смеси моделей и многоуровневые модели

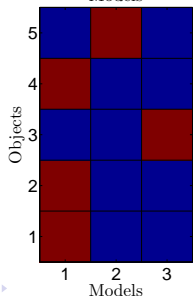
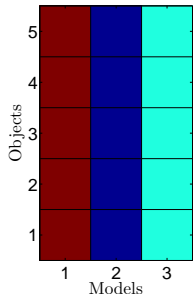
**Определение 1. Смесь регрессионных моделей** — регрессионная модель вида

$$f = \sum_{k=1}^K \pi_k f_k(\mathbf{w}_k), \text{ где}$$

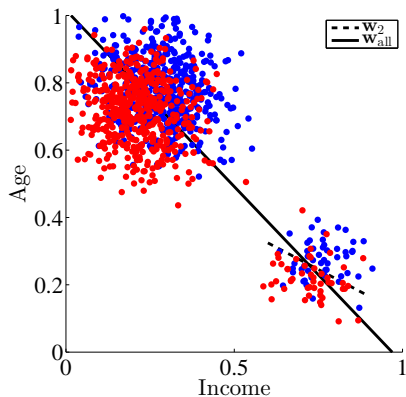
$$\sum_{k=1}^K \pi_k = 1, \pi_k \geq 0.$$

**Определение 2. Многоуровневая регрессионная модель** — набор

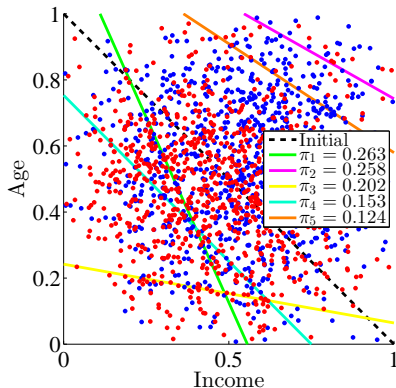
регрессионных моделей  $f_k$ ,  $k = 1, \dots, K$  такой, что при разбиении множества индексов объектов  $\mathcal{I} = \sqcup_{k=1}^K \mathcal{I}_k$  для всех объектов с индексами из  $\mathcal{I}_k$  используется модель  $f_k$ .



**Проблема:** большое число близких или совпадающих моделей ведет к неинтерпретируемости и низкому качеству прогноза.



Неадекватная многоуровневая модель



Неадекватная смесь моделей

**Определение 3.** Мультимодель с совместным распределением  $p(\mathbf{y}, \mathbf{w}_1, \dots, \mathbf{w}_K, (\boldsymbol{\pi}) | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, (\mu))$  называется  $(s, \alpha)$ -адекватной, если модели, ее составляющие, являются попарно статистически различимыми с помощью функции сходства  $s$  на уровне значимости  $\alpha$ .

## Обучение мультимодели

$$[\mathbf{w}_1^*, \dots, \mathbf{w}_K^*, (\boldsymbol{\pi}^*)] = \arg \max_{\mathbf{w}_1, \dots, \mathbf{w}_K, (\boldsymbol{\pi})} p(\mathbf{w}_1, \dots, \mathbf{w}_K, (\boldsymbol{\pi}) | \mathbf{X}, \mathbf{y}, \mathbf{A}_1, \dots, \mathbf{A}_K, (\mu)).$$

**Определение 4.** Мультимодель называется оптимальной, если она обладает наибольшей обоснованностью

$$[\mathbf{A}_1^*, \dots, \mathbf{A}_K^*, (\mu^*)] = \arg \max_{\mathbf{A}_1, \dots, \mathbf{A}_K, (\mu)} p(\mathbf{y} | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, (\mu)) =$$

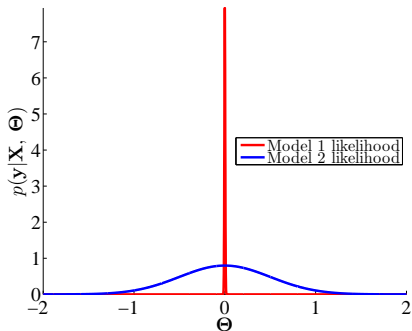
$$\arg \max_{\mathbf{A}_1, \dots, \mathbf{A}_K, (\mu)} \int p(\mathbf{y}, \mathbf{w}_1, \dots, \mathbf{w}_K, (\boldsymbol{\pi}) | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, (\mu)) d\mathbf{w}_1 \dots d\mathbf{w}_K (d\boldsymbol{\pi}),$$

где  $\mathbf{A}_1 \in Q_{\mathbf{A}_1}, \dots, \mathbf{A}_K \in Q_{\mathbf{A}_K}, (\mu \in Q_{\mu})$ .

## Априорные распределения

$p(\boldsymbol{\pi}|\boldsymbol{\mu}) = \text{Dir}(\boldsymbol{\pi}|\boldsymbol{\mu})$ ,  $p(\mathbf{w}_k|\mathbf{A}_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1})$ ,  $k = 1, \dots, K$ .

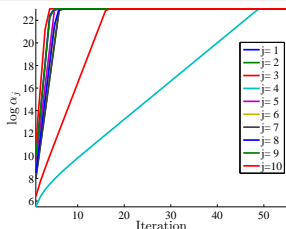
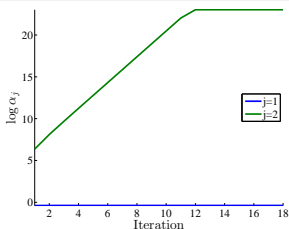
Обозначим  $\Theta = [\mathbf{w}_1, \dots, \mathbf{w}_K, (\boldsymbol{\pi})]$  набор параметров мультимодели, а  $\boldsymbol{\mu} = [\mathbf{A}_1, \dots, \mathbf{A}_K, (\boldsymbol{\mu})]$  набор гиперпараметров.



## Условие оптимальности мультимодели

$$\boldsymbol{\mu}^* = \arg \max_{\boldsymbol{\mu}} p(\mathbf{y}|\mathbf{X}, \boldsymbol{\mu}) = \arg \max_{\boldsymbol{\mu}} \int p(\mathbf{y}|\mathbf{X}, \Theta) p(\Theta|\boldsymbol{\mu}) d\Theta.$$

# Отбор признаков с помощью максимизации обоснованности



## Теорема 1 (Адуенко, 2016)

Пусть  $n = 2$ ,  $K = 1$ ,  $\mathbf{w} = [w_1, w_2]^\top$ ,  $w_1, w_2 \neq 0$ . Обозначим

$$\Sigma = \mathbf{X}^\top \mathbf{R} \mathbf{X} = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} s_1^2 & \kappa s_1 s_2 \\ \kappa s_1 s_2 & s_2^2 \end{bmatrix}.$$

Тогда если при  $m \rightarrow \infty$  выполнено

- $\sigma_1^2, \sigma_2^2 \xrightarrow{\text{п.н.}} \infty$ ,

- $P(\omega : \exists c_\omega > 0, \exists m_\omega : \forall m \geq m_\omega 1 - \rho_m^2 \geq c_\omega) = 1$ ,

то  $s_1^*, s_2^* \xrightarrow{\text{п.н.}} \infty, \kappa^* \xrightarrow{\text{п.н.}} -\text{sign}(w_1 w_2)$ .



## Теорема 2 (Адуенко, 2016)

Пусть имеется  $l$  линейно независимых факторов и  $\mathbf{v} \in \mathbb{R}^l$ . Обозначим  $\mathbf{f}_i$  – вектор значений факторов для  $i$ -го объекта. Пусть  $\mathbf{x}_i = \mathbf{G}\mathbf{f}_i + \boldsymbol{\varepsilon}_i$ , где  $\mathbf{G}$  – матрица размера  $n \times l$ ,  $n \geq l$  полного ранга, а  $\mathbb{E}\boldsymbol{\varepsilon}_i = \mathbf{0}$ ,  $\mathbb{E}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top = \boldsymbol{\Sigma} \succ 0$ .

Тогда оптимальная по дисперсии шума  $\mathbb{E}(\mathbf{w}^\top \boldsymbol{\varepsilon}_i)^2$  оценка  $\mathbf{w}$  такая, что  $\mathbb{E}\mathbf{w}^\top \mathbf{x}_i = \mathbf{v}^\top \mathbf{f}_i$ , имеет вид  $\mathbf{w} = \boldsymbol{\Sigma}^{-1}\mathbf{G} \left( \mathbf{G}^\top \boldsymbol{\Sigma}^{-1}\mathbf{G} \right)^{-1} \mathbf{v}$ .

**Пример.** Пусть имеется  $n$  зашумленных копий одного признака, то есть  $l = 1$ ,  $\mathbf{G} = [1, \dots, 1]^\top$ ,  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ .

$$\mathbf{w} = v \left[ \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \dots + \frac{1}{\sigma_n^2}}, \dots, \frac{\frac{1}{\sigma_n^2}}{\frac{1}{\sigma_1^2} + \dots + \frac{1}{\sigma_n^2}} \right]^\top.$$

**Определение 5.** Набор признаков с индексами  $j \in \mathcal{J}$  называется  $\delta$  – мультиколлинеарным для  $\delta > 0$  с вектором весов  $\boldsymbol{\Theta}$ ,  $\theta_j \neq 0 \iff j \in \mathcal{J}$ , если  $\|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}\|_1 < \delta$ .

# Обучение мультимоделей

## Априорные распределения

$$p(\boldsymbol{\pi}|\mu) = \text{Dir}(\boldsymbol{\pi}|\mu), p(\mathbf{w}_k|\mathbf{A}_k) = \mathcal{N}(\mathbf{w}_k|\mathbf{0}, \mathbf{A}_k^{-1}), k = 1, \dots, K.$$

## Обучение многоуровневой модели

$$\mathbf{w}_k^* = \arg \max_{\mathbf{w}_k} \left( p(\mathbf{w}_k|\mathbf{A}_k) \prod_{i \in \mathcal{I}_k} \sigma(y_i \mathbf{w}_k^\top \mathbf{x}_i) \right), \text{ где } \mathcal{I}_k = \{i : \mathbf{x}_i \in \mathbb{X}_k\}.$$

## Обучение смеси моделей

Введем скрытые переменные  $\{z_{ik} \in \{0, 1\}\}$ , где  $z_{ik} = 1$  означает, что объект  $(\mathbf{x}_i, y_i)$  относится к модели  $k$ .

$$p(\mathbf{y}, \mathbf{Z}, \boldsymbol{\pi}, \mathbf{w}_1, \dots, \mathbf{w}_K | \mathbf{X}, \mathbf{A}_1, \dots, \mathbf{A}_K, \mu) = \prod_{k=1}^K p_k(\mathbf{w}_k | \mathbf{0}, \mathbf{A}_k^{-1})$$

$$\Gamma(K\mu) / \Gamma(\mu)^K \prod_{k=1}^K \pi_k^{\alpha-1} \prod_{i=1}^m \prod_{k=1}^K (\pi_k \sigma(y_i \mathbf{w}_k^\top \mathbf{x}_i))^{z_{ik}}.$$

## Е-шаг

$$\gamma_{ik} = \mathbf{E} z_{ik} = \pi_k \sigma(y_i \mathbf{w}_k^\top \mathbf{x}_i) / N_i.$$

## M-шаг

На M-шаге происходит определение весов моделей  $\pi^*$  и векторов параметров моделей  $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ .

$$\pi_k^* = \max(0, \gamma_k + \alpha - 1) / Z_k, \text{ где } \gamma_k = \sum_{i=1}^m \gamma_{ik}, k = 1, \dots, K$$

$$\mathbf{w}_k^* = \arg \max_{\mathbf{w}_k} \left( \log p(\mathbf{w}_k | \mathbf{A}_k) + \sum_{i=1}^m \gamma_{ik} \log \sigma(y_i \mathbf{w}_k^\top \mathbf{x}_i) \right) = \arg \min_{\mathbf{w}_k} \tilde{l}_k(\mathbf{w}_k)$$

$$\frac{\partial^2 \tilde{l}_k(\mathbf{w}_k)}{\partial \mathbf{w}_k^2} = \mathbf{H}_k = \mathbf{X}^\top \mathbf{R}_k \mathbf{X} + \mathbf{A}_k, \mathbf{R}_k = \text{diag}(\gamma_{ik} \sigma(\mathbf{w}_k^\top \mathbf{x}_i) \sigma(-\mathbf{w}_k^\top \mathbf{x}_i)).$$

Алгоритм совместного обучения и оптимизации смесей моделей

$$p(\mathbf{y} | \mathbf{X}, \mu, \mathbf{A}_1, \dots, \mathbf{A}_K) \approx e^{-E_q \log q(\mathbf{Z})} \int p(\boldsymbol{\pi} | \mu) \prod_{k=1}^K \pi_k^{\sum_{i=1}^m E z_{ik}} d\boldsymbol{\pi}$$

$$\prod_{k=1}^K \int \mathcal{N}(\mathbf{w}_k | \mathbf{0}, \mathbf{A}_k^{-1}) \prod_{i=1}^m \sigma(y_i \mathbf{w}_k^\top \mathbf{x}_i)^{E z_{ik}} d\mathbf{w}_k.$$

## Проблема

Несмотря на прореживание мультимодели, она может не являться  $(s, \alpha)$  – адекватной, то есть может содержать похожие модели.

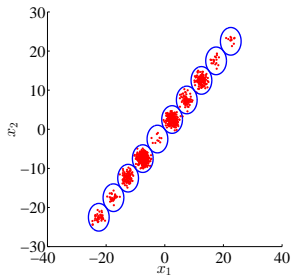
## Дано

- Две модели  $f_1$  и  $f_2$ , векторы параметров моделей  $\mathbf{w}_1, \mathbf{w}_2$ .
- Выборки  $(\mathbf{X}_1, \mathbf{y}_1)$  и  $(\mathbf{X}_2, \mathbf{y}_2)$ ,  
 $y_{1,i} = f_1(\mathbf{x}_{1,i}, \mathbf{w}_1)$ ,  $y_{2,i} = f_2(\mathbf{x}_{2,i}, \mathbf{w}_2)$ .
- Априорные распределения параметров моделей  $\mathbf{w}_1 \sim p_1(\mathbf{w})$ ,  $\mathbf{w}_2 \sim p_2(\mathbf{w})$ .
- Апостериорные распределения  $p(\mathbf{w}_1 | \mathbf{X}_1, \mathbf{y}_1)$  и  $p(\mathbf{w}_2 | \mathbf{X}_2, \mathbf{y}_2)$ , обозначаемые далее  $g_1(\mathbf{w})$  и  $g_2(\mathbf{w})$ .

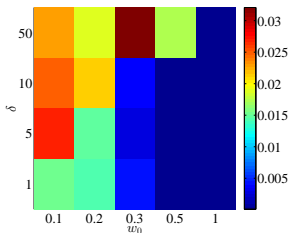
**Требуется:** построить функцию сходства, определенную на паре распределений  $g_1(\mathbf{w})$  и  $g_2(\mathbf{w})$ , удовлетворяющую ряду требований.

# Пример неадекватности многоуровневых моделей.

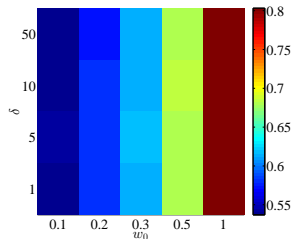
Пусть  $y_i \sim \text{Be}(\sigma(\mathbf{w}^\top \mathbf{x}_i))$ ,  $\mathbf{w} = w_0[1, -1]^\top$ .



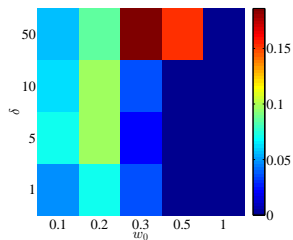
Данные для  $m = 1000$ ,  $\delta_0 = 50$



Повышение AUC



AUC для  $(s, 0.05)$ -адекватной  
мультимодели



Повышение мин. AUC по кластерам

Корректная функция сходства  $s$  должна быть

- 1 определена в случае несовпадения носителей,
- 2  $s(g_1, g_2) \leq s(g_1, g_1)$ ,
- 3  $s \in [0, 1]$ ,
- 4  $s(g_1, g_1) = 1$ ,
- 5 близка к 1, если  $g_2(\mathbf{w})$  — малоинформативное распределение,
- 6 симметрична,  $s(g_1, g_2) = s(g_2, g_1)$ .

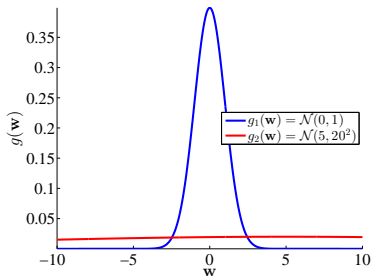
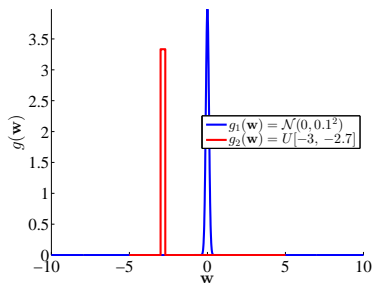
Теорема 3 (Адуенко, 2014)

Функции сходства, порожденные расстояниями Кульбака-Лейблера, Дженсона-Шеннона, Хеллингера, Бхаттачарая, не являются корректными.

# Иллюстрация требований к функции сходства

Важно, чтобы значение функции  $s$

было близко к 1, если  $g_2(\mathbf{w})$  — малоинформативное распределение.



## Теорема 4 (Адуенко, 2014)

Функции сходства, порожденные дивергенциями Брегмана, симметризованными дивергенциями Брегмана и f-дивергенциями, не являются корректными.

В качестве меры сходства распределения предлагается мера сходства  $s$ -score:

$$s(g_1, g_2) = \frac{\int_{\mathbf{w}} g_1(\mathbf{w})g_2(\mathbf{w})d\mathbf{w}}{\max_{\mathbf{b}} \int_{\mathbf{w}} g_1(\mathbf{w} - \mathbf{b})g_2(\mathbf{w})d\mathbf{w}}.$$

**Теорема 5 (Адуенко, 2014).** Предлагаемая функция сходства является корректной.

Примеры:

$g_1(\mathbf{w})$	$g_2(\mathbf{w})$	$s(g_1, g_2)$
$U[0, 1]$	$U[0.5, 1.5]$	0.5
$U[0, 1]$	$U[0, 1]$	1
$\mathcal{N}(0, 1)$	$\mathcal{N}(10, 10^{10})$	1



## Выражение для $s(g_1, g_2)$ для пары нормальных распределений

**Определение 6.** *Обобщенно-линейной моделью с натуральной функцией связи и априорным распределением на вектор параметров  $p(\mathbf{w}|\mathbf{A})$  называется вероятностная модель с совместным правдоподобием*

$$p(\mathbf{y}, \mathbf{w}|\mathbf{X}, \mathbf{A}) = p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\mathbf{A}), \text{ где } p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{i=1}^m p(y_i|\mathbf{x}_i, \mathbf{w}),$$
$$p(y_i|\mathbf{x}_i, \mathbf{w}) = c(y_i) \exp(\theta_i y_i - b(\theta_i)), \text{ где } \theta_i = \mathbf{w}^\top \mathbf{x}_i.$$

**Теорема 6 (Адуенко, 2014).**

Пусть  $g_1 = \mathcal{N}(\mathbf{v}_1, \Sigma_1)$ ,  $g_2 = \mathcal{N}(\mathbf{v}_2, \Sigma_2)$ . Тогда выражение для  $s(g_1, g_2)$  имеет вид

$$s(g_1, g_2) = \exp\left(-\frac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2)^\top (\Sigma_1 + \Sigma_2)^{-1}(\mathbf{v}_1 - \mathbf{v}_2)\right).$$

**Следствие 1.** В случае  $\Sigma_2 = \mathbf{0}$  выражение для s-score

$$s(g_1, g_2) = \exp\left(-\frac{1}{2}(\mathbf{v}_2 - \mathbf{v}_1)^\top \Sigma_1^{-1}(\mathbf{v}_2 - \mathbf{v}_1)\right).$$

# Распределение s-score в условии истинности гипотезы о совпадении моделей

Рассматриваем далее пару обобщенно-линейных моделей с натуральной функцией связи. Обозначим  $\mathbf{H}_m(\mathbf{w}) = \mathbf{H}_m$  и введем  $O_m^\delta(\mathbf{w}) = \{\mathbf{v} : \|\mathbf{H}_m^{T/2}(\mathbf{v} - \mathbf{w})\| \leq \delta\}$ .

**Теорема 7 (Адуенко, 2016).** Пусть

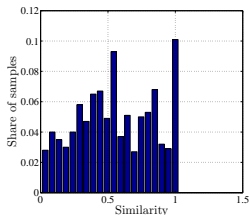
- Модели  $f_1$  и  $f_2$  совпадают:  $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$ ,  $\mathbf{w}_2$  известно;
- Априорное распределение:  $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{A}_{m^k}^{-1})$ ,  $k = 1, 2$ ;
- $\exists c_0^k < \infty : \|\mathbf{A}_{m^k}^k\| < c_0^k \forall m^k, k = 1, 2$ ;
- $\sum_{i=1}^{m^1} \mathbf{x}_i \mathbf{x}_i^\top$  имеет полный ранг для  $m^1 \geq m_0$ ;
- $\lambda_{\min}(\mathbf{H}_{m^1}) \rightarrow \infty$  при  $m^1 \rightarrow \infty$ ;
- $\forall \delta > 0 \max_{\mathbf{v} \in O_{m^1}^\delta(\mathbf{w})} \|\mathbf{H}_{m^1}^{-\frac{1}{2}} \mathbf{H}_{m^1}(\mathbf{v}) \mathbf{H}_{m^1}^{-\frac{1}{2}} - \mathbf{I}\| \rightarrow 0$  при  $m^1 \rightarrow \infty$ .

Тогда  $-2 \log s\text{-score} = (\hat{\mathbf{w}} - \mathbf{w})^\top \tilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}})(\hat{\mathbf{w}} - \mathbf{w}) \xrightarrow{d} \chi^2(n)$ .

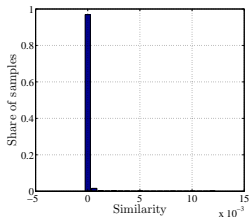
**Следствие 1.** Для случая  $n = 2$  s-score имеет асимптотически равномерное распределение на отрезке  $[0, 1]$ .

# Иллюстрация применения s-score для сравнения двух моделей, $\rho = 0.9$

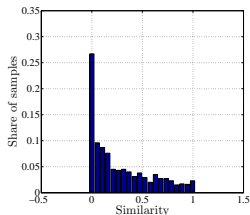
Рассмотрим две близкие в терминах  $\|\mathbf{w}_1 - \mathbf{w}_2\|$  модели,  
 $\|\mathbf{w}_1\| = \|\mathbf{w}_2\| = 1$ ,  $\text{corr}(\mathbf{w}_1, \mathbf{w}_2) = \rho$ .



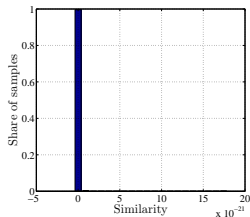
$N_1 = 10000, N_2 = 10$



$N_1 = 10000, N_2 = 1000$



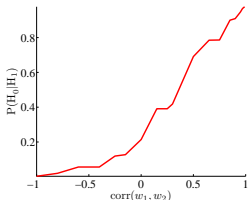
$N_1 = 10000, N_2 = 100$



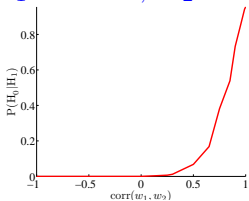
$N_1 = 10000, N_2 = 10000$

# Зависимость $P(H_0|H_1)$ от корреляции между истинными параметрами двух моделей.

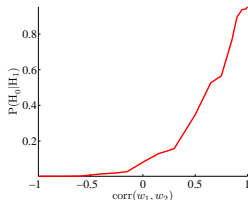
Рассмотрим две близкие в терминах  $\|\mathbf{w}_1 - \mathbf{w}_2\|$  модели,  
 $\|\mathbf{w}_1\| = \|\mathbf{w}_2\| = 1$ ,  $\text{corr}(\mathbf{w}_1, \mathbf{w}_2) = \rho$ .



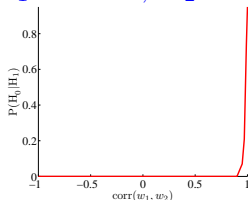
$N_1 = 10000$ ,  $N_2 = 30$



$N_1 = 10000$ ,  $N_2 = 100$



$N_1 = 10000$ ,  $N_2 = 50$



$N_1 = 10000$ ,  $N_2 = 1000$

# Обобщение теоремы о распределении s-score на случай двух конечных выборок

**Теорема 8 (Адуенко, 2016).** Пусть

- Модели  $f_1$  и  $f_2$  совпадают, то есть  $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$ ;
- Априорное распределение:  $\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{A}_{m^k}^{-1})$ ,  $k = 1, 2$ ;
- $\sum_{i=1}^{m^k} \mathbf{x}_i \mathbf{x}_i^\top$  имеет полный ранг для  $m^k \geq m_0$ ,  $k = 1, 2$ ;
- $\lambda_{\min}(\mathbf{H}_{m^k}(\mathbf{w})) \rightarrow \infty$  при  $m^k \rightarrow \infty$ ,  $k = 1, 2$ ;
- $\forall \delta > 0 \quad \max_{\mathbf{v} \in O_{m^k}^\delta(\mathbf{w})} \left\| \mathbf{H}_{m^k}^{-\frac{1}{2}} \mathbf{H}_{m^k}(\mathbf{v}) \mathbf{H}_{m^k}^{-\frac{1}{2}} - \mathbf{I} \right\| \rightarrow 0$   
при  $m^k \rightarrow \infty$ ,  $k = 1, 2$ ;
- $\|\tilde{\mathbf{H}}_{m^1}(\hat{\mathbf{w}}_1)\| \|\tilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2)\| \xrightarrow{P} 0$  при  $m = \min(m^1, m^2) \rightarrow \infty$ .

Тогда  $-2 \log s\text{-score} =$

$$(\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1)^\top (\tilde{\mathbf{H}}_{m^1}^{-1}(\hat{\mathbf{w}}_1) + \tilde{\mathbf{H}}_{m^2}^{-1}(\hat{\mathbf{w}}_2))^{-1} (\hat{\mathbf{w}}_2 - \hat{\mathbf{w}}_1) \xrightarrow{d} \chi^2(n).$$

**Теорема 9 (Адуенко, 2014).** Пусть модели, задаваемые  $(\mathbf{v}_1, \Sigma_1)$  и  $(\mathbf{v}_2, \Sigma_2)$  считаются разными, если

$$s(\mathcal{N}(\mathbf{v}_1, \Sigma_1), \mathcal{N}(\mathbf{v}_2, \Sigma_2)) \leq C \in [0, 1].$$

Тогда, если указанные модели разные по приведенному критерию, то

- модели, задаваемые  $(\mathbf{v}_1, \Sigma_1)$  и  $(\mathbf{v}_2, \mathbf{O})$  будут разными согласно приведенному критерию,
- модели, задаваемые  $(\mathbf{v}_1, \Sigma_1)$  и  $(\mathbf{v}_2, \lambda \Sigma_2)$ ,  $\lambda \in [0, 1]$  будут разными согласно приведенному критерию.

**Теорема 10 (Адуенко, 2014).** Пусть рассматриваются  $K$  моделей с  $\|\mathbf{v}_1\| = \dots = \|\mathbf{v}_K\| = \lambda_1 > 0$  и  $\Sigma_1 = \dots = \Sigma_K = \lambda_2 \mathbf{I}$ . В качестве критерия отличимости моделей рассматривается следующий: модели с номерами  $i \neq j$  разные, если

$$s(\mathcal{N}(\mathbf{v}_i, \Sigma_i), \mathcal{N}(\mathbf{v}_j, \Sigma_j)) \leq C \in [0, 1].$$

Тогда максимальное число попарно различных моделей, которое может быть в наборе, есть

$$K_{\max} = \left\lfloor \sqrt{\pi} \frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2} + 1)} \frac{1}{\int_0^{\theta/2} \sin^{n-2} \varphi d\varphi} \right\rfloor.$$

Здесь  $\theta \in [0, \pi]$ ,  $\cos \theta = \rho = \max(-1, 1 + 2\lambda_2/\lambda_1^2 \ln C)$ ,  $n$  – размерность признакового пространства. При этом можно построить  $K_{\min}$  попарно различных моделей, где

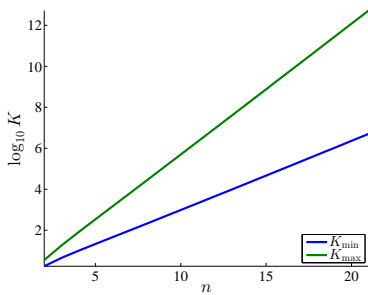
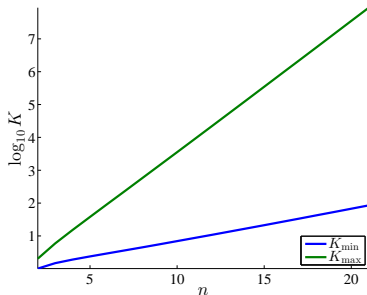
$$K_{\min} = \left\lfloor \sqrt{\pi} \frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2} + 1)} \frac{1}{\int_0^{\theta} \sin^{n-2} \varphi d\varphi} \right\rfloor.$$

# Оценки на число моделей (продолжение)

**Теорема 10 (продолжение).** При  $C$ , близком к 1, имеем

$$K_{\max} \approx \left[ \sqrt{\pi} \frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2}+1)} \frac{2^{\frac{n-1}{2}}}{(1-\rho)^{\frac{n-1}{2}}} \right],$$

$$K_{\min} \approx \left[ \sqrt{\pi} \frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2}+1)} \frac{1}{2^{\frac{n-1}{2}} (1-\rho)^{\frac{n-1}{2}}} \right].$$





# Методы прореживания мультимоделей

Обозначим матрицу парных сходств  $\mathbf{S} = [s_{kl}(g_k(\mathbf{w}_k), g_l(\mathbf{w}_l))]$ ,

а матрицу достигаемых уровней значимости

$\mathbf{T} = [P(s(g_k(\mathbf{w}_k), g_l(\mathbf{w}_l)) < s_{kl} | \mathbf{w}_k = \mathbf{w}_l)]$ ,  $k, l = 1, \dots, K$ .

1 Находим  $[k^*, l^*] = \arg \max_{k < l} t_{kl}$ .

2 Если  $t_{k^*l^*} < \alpha$ , останавливаемся. Иначе на шаг 3.

3 ■ Для многоуровневых моделей:

Объединяем модели  $k^*$ ,  $l^*$  и пересчитываем  $g_{k^*}(\mathbf{w}_{k^*})$ .

$\mathcal{I}_{k^*} \sqcup \mathcal{I}_{l^*} \rightarrow \mathcal{I}_{k^*}$ ,  $\mathbf{A}_{k^*}^* = \arg \max_{\mathbf{A}_{k^*}} p(\mathbf{y}_{\mathcal{I}_{k^*}} | \mathbf{X}_{\mathcal{I}_{k^*}}, \mathbf{A}_{k^*})$ ;

$g_{k^*}(\mathbf{w}_{k^*}) = \mathcal{N}(\mathbf{w}_{k^*} | \mathbf{w}_{k^*}^*, \Sigma_{k^*}^*)$

■ Для смесей моделей:

Объединяем модели  $k^*$ ,  $l^*$  и перенастраиваем смесь моделей. Начальное приближение:

$\pi_{k^*} + \pi_{l^*} \rightarrow \pi_{k^*}, 0 \rightarrow \pi_{l^*}$ ,  $\frac{\mathbf{w}_{k^*} + \mathbf{w}_{l^*}}{2} \rightarrow \mathbf{w}_{k^*}, \mathbf{w}_k \rightarrow \mathbf{w}_k, k \neq k^*, l^*$ .

4 Удаляем  $l^*$ -й столбец матриц  $\mathbf{S}$  и  $\mathbf{T}$  и пересчитываем  $s_{k^*l}$  и  $t_{k^*l}$  для  $l \neq k^*$ . Переходим на шаг 1.

# Вычислительный эксперимент на реальных данных

Результаты построения многоуровневых моделей и  $(s, \alpha)$  – адекватных многоуровневых моделей для  $K = 20$ .

Выборки	$AUC_m$	$AUC_m^{adeq}$	$t_m$	$K_m$	$\overline{K}_m^{adeq}$
Кредиты	0.6789	<b>0.7604</b>	16.7	20	2.48
Локализация белков	0.6721	<b>0.6959</b>	8.83	20	2.56
Качество вина	<b>0.8185</b>	<b>0.8189</b>	0.57	20	9.08
Заболевания сердца	0.6757	<b>0.7392</b>	11.56	20	2.62
Цены домов	0.897	<b>0.9427</b>	11.34	20	2.56

Результаты построения смесей моделей и  $(s, \alpha)$  – адекватных смесей моделей с автоматическим выбором числа моделей.

Выборки	$AUC_{sp}$	$AUC_{sp}^{adeq}$	$t_{sp}$	$\overline{K}_{sp}$	$\overline{K}_{sp}^{adeq}$
Кредиты	0.7693	<b>0.7784</b>	4.53	15.5	2.66
Локализация белков	<b>0.6988</b>	<b>0.7001</b>	1.67	17.4	2.11
Качество вина	0.8038	<b>0.8125</b>	14.01	31.5	6.48
Заболевания сердца	0.7644	<b>0.7705</b>	3.03	9.66	2.60
Цены домов	0.9411	<b>0.9461</b>	2.46	7.66	2.10

## Выносятся на защиту

- Предложена функция сходства плотностей апостериорных распределений параметров моделей, удовлетворяющая требованиям к функции сходства для решения задачи сравнения моделей.
- Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимодели.
- Разработаны методы выбора адекватных оптимальных обученных мультимоделей. Предложен метод совместного обучения и отбора признаков для смеси моделей.
- Предложен метод комбинирования мультиколлинеарных признаков. Доказана асимптотическая вырожденность недиагональной оценки максимума обоснованности ковариационной матрицы параметров логистической модели.
- Разработан программный комплекс для построения адекватных оптимальных обученных мультимоделей в задачах двухклассовой классификации и комбинирования признаков. Проведены вычислительные эксперименты, подтверждающие адекватность методов.

# Список работ автора по теме диссертации

## Публикации ВАК

- 1 А. А. Адуенко, А. А. Кузьмин, В. В. Стрижов Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия ТулГУ, 2012. № 3. С. 119–131.
- 2 А. А. Адуенко, В. В. Стрижов Алгоритм оптимального расположения названий коллекции документов // Программная инженерия, 2013. № 3. С. 21–25.
- 3 А. В. Иванова, А. А. Адуенко, В. В. Стрижов Алгоритм построения логических правил при разметке текстов // Программная инженерия, 2013. № 6. С. 41–47.
- 4 А. А. Адуенко, В. В. Стрижов Совместный выбор объектов и признаков в задачах многоклассовой классификации коллекции документов // Инфокоммуникационные технологии, 2014. № 1. С. 47–53.
- 5 А. А. Кузьмин, А. А. Адуенко, В. В. Стрижов Тематическая классификация тезисов крупной конференции с использованием экспертной модели // Информационные технологии, 2014. № 6. С. 22–26.
- 6 А. А. Адуенко, А. С. Василейский, А. И. Карелов, И. А. Рейер, К. В. Рудаков, В. В. Стрижов Алгоритмы выделения и совмещения устойчивых отражателей на спутниковых снимках // Компьютерная оптика, 2015. Т. 39. № 4. С. 622–630.
- 7 А. А. Адуенко, Н. И. Амелькин О предельных движениях волчка с внутренней диссипацией в однородном поле тяжести // Труды МФТИ, 2013. Т. 18. № 2. С. 126–133.
- 8 А. А. Адуенко, Н. И. Амелькин Асимптотические свойства движений тяжелого волчка с внутренней диссипацией // ПММ, 2014. Т. 78. № 1. С. 13–28.
- 9 А. А. Адуенко, Н. И. Амелькин О резонансных вращениях маятника с вибрирующим подвесом // ПММ, 2015. Т. 79. № 6. С. 756–767.

## Выступления с докладом

- 10 Международная конференция «Интеллектуализация обработки информации», 2016. Барселона, Испания. Анализ пространства параметров в задачах выбора мультимodelей.
- 11 Математические методы распознавания образов ММО-17, 2015. Светлогорск, Россия. Анализ пространства параметров в задачах выбора мультимodelей.
- 12 27th European Conference for Operational Research, 2015. Glasgow, Scotland. Multimodelling and Model Selection in Bank Credit Scoring.
- 13 57я научная конференция МФТИ, 2014. Долгопрудный, Россия. Топологический анализ пространства параметров в задаче выбора мультимodelей.
- 14 20th Conference of the International Federation of Operational Research Societies, 2014. Barcelona, Spain. Multimodelling and Object Selection for Banking Credit Scoring.