Анализ пространства параметров в задачах выбора мультимоделей

А. А. Адуенко, В. В. Стрижов

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Конференция ММРО 2015

20 сентября 2015 года

Цели исследования

Цель исследования: создать метод выбора мультимоделей при построении моделей в задаче двухклассовой классификации.

Мотивация. Логистическая модель является стандартом де-факто в банковском кредитном скоринге. Мультимодель являются ее интерпретируемым обобщением. Она позволяет учитывать неоднородность выборки.

Проблема. Мультимодель может содержать большое число похожих моделей, что ведет к ее неинтерпретируемости и низкому качеству прогноза. Признаковые пространства параметров моделей могут не совпадать, в частности иметь различню размерность.

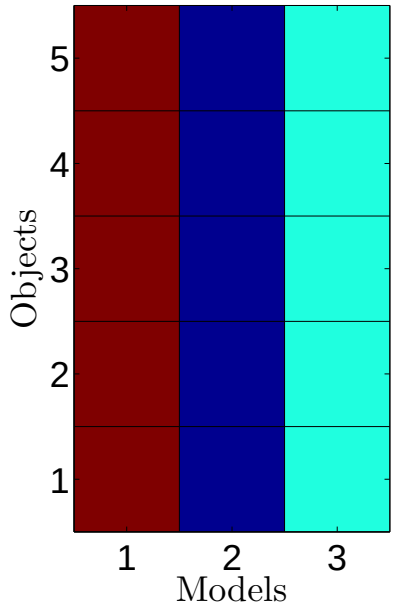
Метод. Анализ пространства параметров мультимодели с помощью введенной функции сходства моделей. Функция сравнивает распределения, которые могут быть определены на несовпадающих носителях.

Мультимодели: Смеси моделей и многоуровневые модели

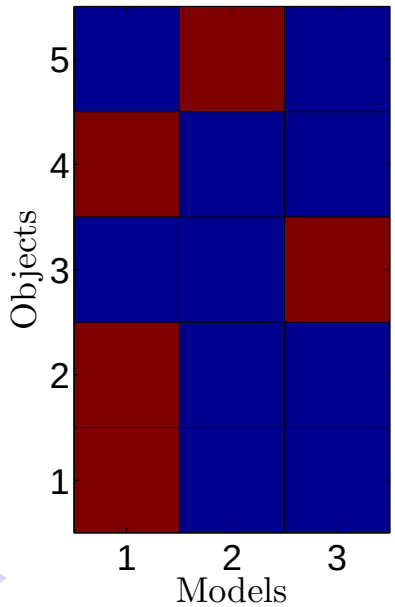
Смесь регрессионных моделей — регрессионная модель вида

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

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



Многоуровневая регрессионная модель — набор регрессионных моделей f_k, $k = 1, \ldots, K$ 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.



Вероятностная модель, соответствующая смеси моделей

Гипотеза порождения данных

- Сэмплируем веса $[\pi_1, \ldots, \pi_K]$т моделей из некоторого априорного распределения, $\boldsymbol{\pi} \sim q(\boldsymbol{\pi}|\alpha)$.
- Сэмплируем параметры $\mathbf{w}_k$ каждой из моделей из некоторого априорного распределения $\mathbf{w}_k \sim p_k(\mathbf{w}_k)$.
- Для каждого объекта $\mathbf{x}_i$ выбираем модель f_{m_i} с идексом m_i, которой он описывается. Считаем, что $p(m_i = k) = \pi_k$.
- Для каждого объекта $\mathbf{x}_i$ определяем значение целевой переменной y_i в соответствии с моделью m_i: $y_i \sim \mathrm{Be}\big(f_{m_i}(\mathbf{x}_i, \mathbf{w}_{m_i})\big)$.

Совместное распределение для мультимодели

$$p(\mathbf{y}, \mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}) =$$
$$q(\boldsymbol{\pi}|\alpha) \prod_{j=1}^{n} p_j(\mathbf{w}_j) \prod_{i=1}^{m} \left(\sum_{k=1}^{K} \pi_k f_k(\mathbf{x}_i, \mathbf{w}_k)^{y_i} \big(1 - f_k(\mathbf{x}_i, \mathbf{w}_k)\big)^{1-y_i} \right).$$

Условие оптимальности и адекватности мультимодели

Определение. Будем называть мультимодель, заданную совместным распределением $p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K,\ \boldsymbol{\pi}|\mathbf{X})$ $(s,\ \alpha)$-**адекватной**, если модели, ее составляющие, являются попарно статистически различимыми с помощью функции сходства s на уровне значимости α.

Множество всех $(s,\ \alpha)$-адекватных моделей обозначим $\mathcal{M}_{s,\,\alpha}$.

Определение 2. Будем называть мультимодель **оптимальной**, если она обладает наибольшей обоснованностью

$$[q(\pi|\alpha),\ p_1(\mathbf{w}_1),\ \ldots,\ p_K(\mathbf{w}_K)] = \underset{q,\,p_1,\,\ldots,\,p_K}{\arg\max}\ p(\mathbf{y}|\mathbf{X}) =$$

$$\underset{q,\,p_1,\,\ldots,\,p_K}{\arg\max} \int p(\mathbf{y},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K,\ \boldsymbol{\pi}|\mathbf{X}) d\mathbf{w}_1 \ldots d\mathbf{w}_K d\boldsymbol{\pi}.$$

Оценка максимума апостериорной вероятности для параметров и весов мультимодели

$$[\boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K] = \underset{\boldsymbol{\pi},\,\mathbf{w}_1,\,\ldots,\,\mathbf{w}_K}{\arg\max}\ p(\boldsymbol{\pi},\ \mathbf{w}_1,\ \ldots,\ \mathbf{w}_K|\mathbf{X},\ \mathbf{y}).$$

EM-алгоритм для смеси логистических моделей

Совместное распределение для смеси моделей

Введем скрытые переменные $\{z_{ik}\}$, такие, что $\sum_{k=1}^{K} z_{ik} = 1$ и $z_{ik} = 1$ означает, что объект $(\mathbf{x}_i,\ y_i)$ относится к модели k.

$$p(\boldsymbol{\pi}|\alpha) = \begin{cases} 0, \ \min\limits_k \pi_k = 0, \\ \dfrac{\Gamma(K\alpha)}{\Gamma^K(\alpha)} \prod\limits_{k=1}^{K} \pi_k^{\alpha-1}, \text{ иначе.} \end{cases}$$

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

$$\frac{\Gamma(K\alpha)}{\Gamma^K(\alpha)} \prod_{k=1}^{K} \pi_k^{\alpha-1} \prod_{i=1}^{m} \prod_{k=1}^{K} \{\pi_k f(\mathbf{x}_i,\ \mathbf{w}_k)^{y_i}(1 - f(\mathbf{x}_i,\ \mathbf{w}_k))^{1-y_i}\}^{z_{ik}}.$$

E-шаг

$$\gamma_{ik} = \mathsf{E} z_{ik} = \frac{\pi_k f(\mathbf{x}_i,\ \mathbf{w}_k)^{y_i}(1 - f(\mathbf{x}_i,\ \mathbf{w}_k))^{1-y_i}}{\sum_{j=1}^{K} \pi_j f(\mathbf{x}_i,\ \mathbf{w}_j)^{y_i}(1 - f(\mathbf{x}_i,\ \mathbf{w}_j))^{1-y_i}}.$$

М-шаг

На М-шаге происходит определение весов моделей $\boldsymbol{\pi}$ и векторов параметров моделей $\mathbf{w}_1, \ldots, \mathbf{w}_K$.

$$\pi_k = \begin{cases} 0, \text{ если } \sum\limits_{i=1}^{m} \gamma_{ik} + \alpha - 1 < 0, \\ \dfrac{\sum_{i=1}^{m} \gamma_{ik} + \alpha - 1}{\sum_{l:\gamma_{il}+\alpha-1>0} \left(\sum_{i=1}^{m} \gamma_{il} + \alpha - 1\right)}, \text{ иначе.} \end{cases}$$

$$\tilde{l}(\mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi} | \mathbf{X}, \mathbf{y}) =$$

$$-\sum_{k=1}^{K} \{\log \pi_k \sum_{i=1}^{m} \gamma_{ik}\} + \sum_{k=1}^{K} \tilde{l}_k(\mathbf{w}_k | \mathbf{X}, \mathbf{y}, \mathbf{A}_k).$$

$$\frac{\partial \tilde{l}_k}{\partial \mathbf{w}_k} = \mathbf{X}^\mathsf{T} \boldsymbol{\Gamma}_k (\mathbf{f} - \mathbf{y}) + \mathbf{A}_k \mathbf{w}_k, \; \mathbf{H}_k = \mathbf{X}^\mathsf{T} \mathbf{R}_k \mathbf{X} + \mathbf{A}_k,$$

$$\mathbf{R}_k = \operatorname{diag}(\gamma_{ik} f(\mathbf{x}_i^\mathsf{T} \mathbf{w}_k) f(-\mathbf{x}_i^\mathsf{T} \mathbf{w}_k)).$$

Проблема

Несмотря на прореживание мультимодели, она может не являться (s, α) – адекватной, то есть может содержать похожие

Отбор признаков с помощью максимизации обоснованности

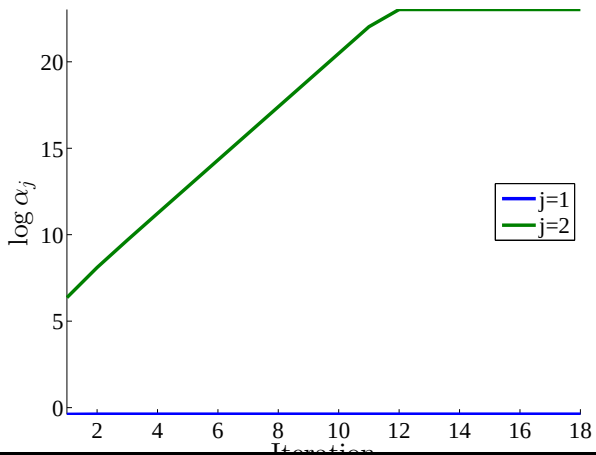
Свойства оптимизируемой функции

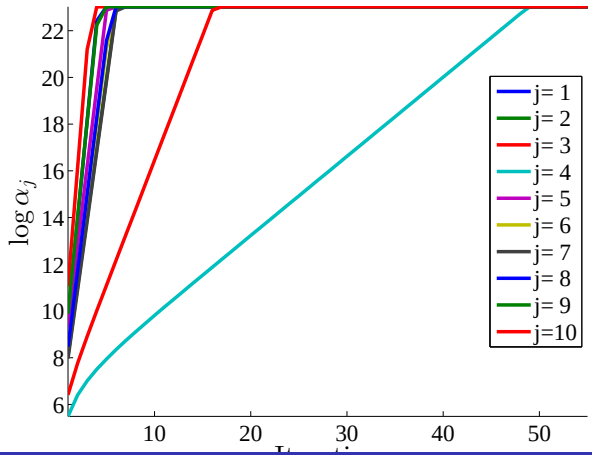
$\tilde{l}_k(\mathbf{y},\ \mathbf{w}_k|\mathbf{X},\ \mathbf{A}_k,\ \mathbf{\Gamma}_k)$ при фиксированных весах объектов $\mathbf{\Gamma}_k$ есть логарифм совместного правдоподобия для логистической модели со взвешенными объектами.

Предлагаемый способ отбора признаков

$$\mathbf{A}_k = \arg\max_{\mathbf{A}\in\mathcal{M}} \tilde{p}(\mathbf{y}|\mathbf{X}, \mathbf{A},\ \mathbf{\Gamma}_k), \text{ где}$$

$$\tilde{p}(\mathbf{y}|\mathbf{X},\ \mathbf{A},\ \mathbf{\Gamma}_k) = \int \tilde{l}_k(\mathbf{y},\ \mathbf{w}_k|\mathbf{X},\ \mathbf{A},\ \mathbf{\Gamma}_k)d\mathbf{w}_k.$$





Постановка задачи сравнения моделей

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 $(\mathbf{X}_2,\ \mathbf{y}_2)$,
  $y_{1,i} = f_1(\mathbf{x}_{1,i},\ \mathbf{w}_1), \quad y_{2,i} = f_2(\mathbf{x}_{2,i},\ \mathbf{w}_2).$
- 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 $g_2(\mathbf{w})$. Она должна удовлетворять ряду требований.

Требования к функции сходства s

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, $s(g_1,\, g_2) = s(g_2,\, g_1)$.

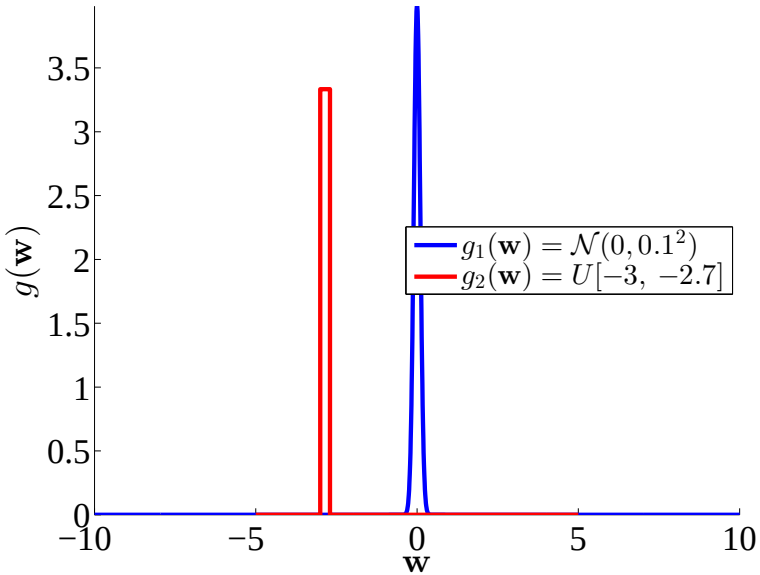
Теорема

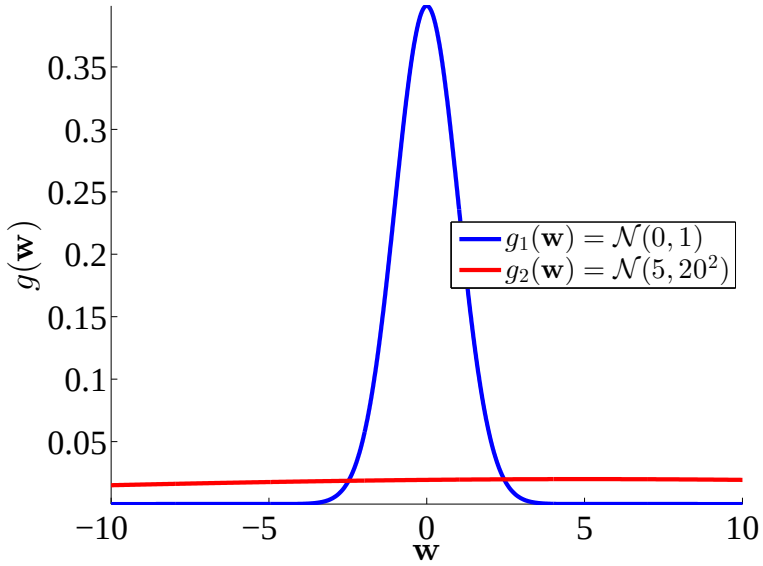
Расстояния Кульбака-Лейблера, Дженсона-Шеннона, Хеллингера, Бхаттачарайа не удовлетворяют требованиям к функции сходства распределений.

Иллюстрация требований к функции сходства

Важно, чтобы значение функции s

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





Теорема

Расстояния Кульбака-Лейблера, Дженсона-Шеннона, Хеллингера, Бхаттачарайа не удовлетворяют всем перечисленным требованиям.

Случай 1, 2

Расстояние Кульбака-Лейблера и Дженсона-Шеннона

$$D_{KL}(g_1,\, g_2) = \int g_1(\mathbf{w}) \log \frac{g_1(\mathbf{w})}{g_2(\mathbf{w})} d\mathbf{w}$$

$D_{JS}(g_1,\, g_2) = \frac{1}{2} D_{KL}(g_1,\, \frac{1}{2}(g_1 + g_2)) + \frac{1}{2} D_{KL}(g_2,\, \frac{1}{2}(g_1 + g_2))$ не удовлетворяют требованиям к функции сходства.

Доказательство

1. $D_{KL} = \infty$, если $g_1(x) \neq 0,\ g_2(x) = 0$ на множестве ненулевой меры относительно g_1.
2. $D_{KL}(g_1,\, g_2) \neq D_{KL}(g_2,\, g_1)$.
3. $D_{KL} \to \infty$, например, для пары нормальных распределений $\mathcal{N}(0,\, 1)$ и $\mathcal{N}(0,\, \sigma^2)$ при $\sigma^2 \to \infty$.
4. $D_{JS} \not\to 0$, например, для пары нормальных распределений $\mathcal{N}(0,\, 1)$ и $\mathcal{N}(0,\, \sigma^2)$ при $\sigma^2 \to \infty$.

Случай 3, 4

Расстояние Хеллингера и Бхаттачарайа

$$D_H(g_1,\, g_2) = 1 - \int \sqrt{g_1(\mathbf{w})g_2(\mathbf{w})}d\mathbf{w},$$

$D_B(g_1,\, g_2) = -\log \int \sqrt{g_1(\mathbf{w})g_2(\mathbf{w})}d\mathbf{w}$ не удовлетворяют требованиям к функции сходства.

Доказательство

Обе меры не имеют требуемого свойства для малоинформативных распределений:

$$D_H(g_1,\, g_2) \to 1,\; D_B(g_1,\, g_2) \to \infty.$$

Предлагаемая функция сходства

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 1 (Адуенко, 2014). Предлагаемая функция сходства удовлетворяет всем требованиям к функции сходства.

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

Теорема 2 (Адуенко, 2014).
Пусть $g_1 = \mathcal{N}(\mathbf{v}_1,\ \mathbf{\Sigma}_1)$, $g_2 = \mathcal{N}(\mathbf{v}_2,\ \mathbf{\Sigma}_2)$. Тогда выражение для $s(g_1,\ g_2)$ имеет вид

$$s(g_1,\ g_2) =$$
$$\exp\left[\frac{1}{2}(\mathbf{\Sigma}_1^{-1}\mathbf{v}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{v}_2)^{\mathsf{T}}(\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\mathbf{v}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{v}_2) - \right.$$
$$\left. - \frac{1}{2}\mathbf{v}_1^{\mathsf{T}}\mathbf{\Sigma}_1^{-1}\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2^{\mathsf{T}}\mathbf{\Sigma}_2^{-1}\mathbf{v}_2\right].$$

Следствие 1. В случае $\mathbf{\Sigma}_2 = \mathbf{0}$ выражение для s-score

$$s(g_1,\ g_2) = \exp\left[-\frac{1}{2}(\mathbf{v}_2 - \mathbf{v}_1)^{\mathsf{T}}\mathbf{\Sigma}_1^{-1}(\mathbf{v}_2 - \mathbf{v}_1)\right].$$

Следствие 2 (упрощение выражения для s-score).
Для случае пары нормальных распределений параметров выражение для s-score имеет следующий вид

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

Распределение s-score в условии истинности гипотезы о совпадении моделей

Теорема 3 (Адуенко, 2014). Пусть

- Модели f_1 и f_2 совпадают, то есть $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$.
- Апостериорное распределение $\mathbf{w}_1$, полученное по $(\mathbf{X}_1,\ \mathbf{y}_1)$, есть $\hat{\mathbf{w}}_1 \sim \mathcal{N}(\mathbf{w}_1|\mathbf{w},\ \mathbf{\Sigma}_1)$.
- $m_2 = \infty$, откуда $\hat{\mathbf{w}}_2 = \mathbf{w}$.

Тогда выражение для s-score двух моделей имеет вид

$$s(g_1,\ g_2) = \exp\left[-1/2(\hat{\mathbf{w}}_1 - \mathbf{w})^{\mathsf{T}}\mathbf{\Sigma}_1^{-1}(\hat{\mathbf{w}}_1 - \mathbf{w})\right],$$

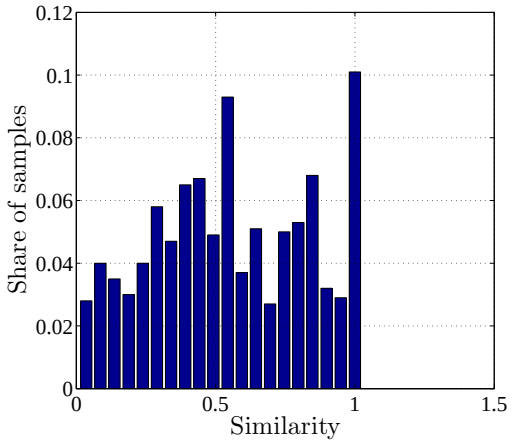
причем $s \sim \exp\left[-1/2\xi\right]$, где $\xi \sim \chi^2(n)$, n – число признаков.

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

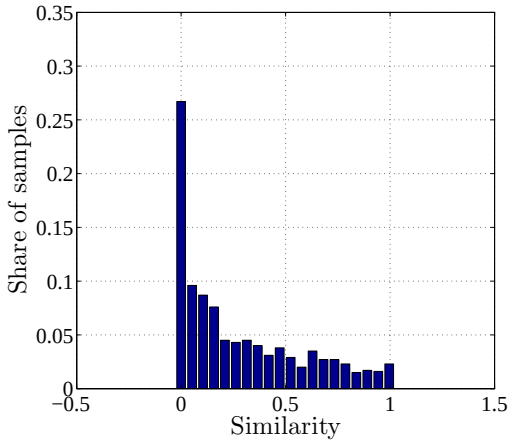
Иллюстрация применения s-score для сравнения двух моделей, $\rho = 0.9$

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

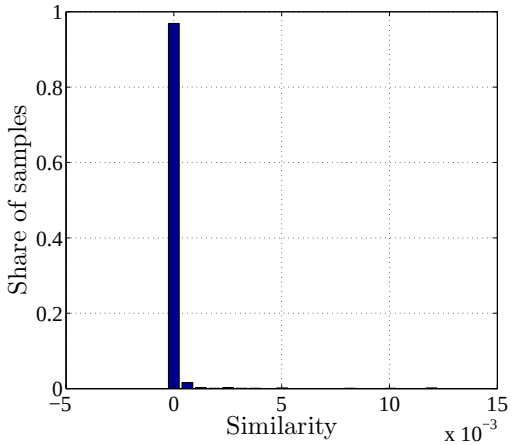
$N_1 = 10000,\ N_2 = 10$



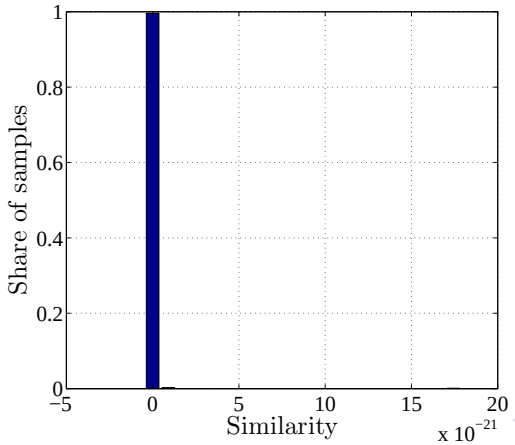
$N_1 = 10000,\ N_2 = 100$



$N_1 = 10000,\ N_2 = 1000$



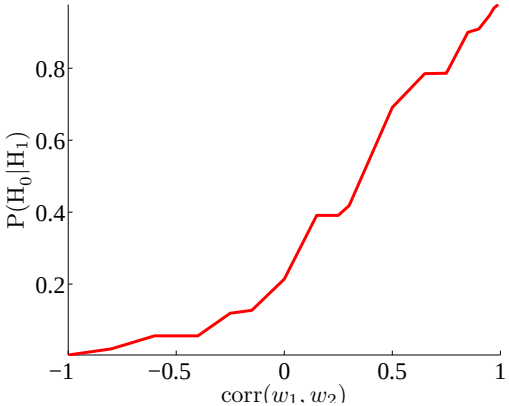
$N_1 = 10000,\ N_2 = 10000$



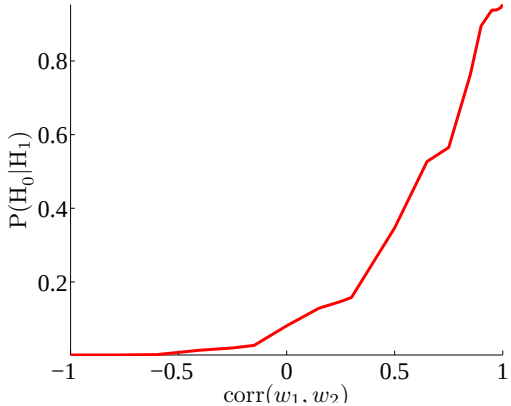
Зависимость $\mathrm{P}(H_0|H_1)$ от корреляции между истинными параметрами двух моделей.

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

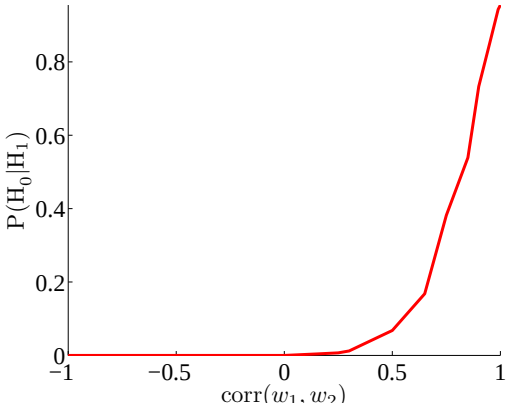
$N_1 = 10000,\ N_2 = 30$



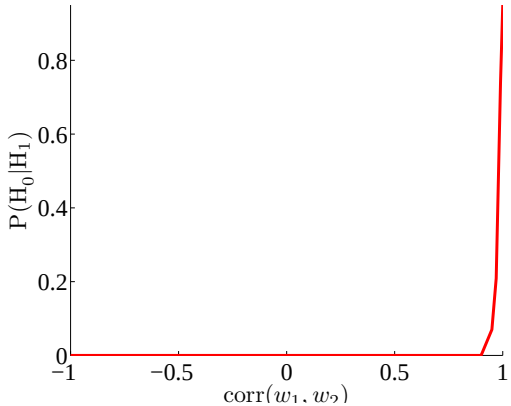
$N_1 = 10000,\ N_2 = 50$



$N_1 = 10000,\ N_2 = 100$



$N_1 = 10000,\ N_2 = 1000$



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

Теорема 4 (Адуенко, 2014). Пусть для ковариационных матриц параметров моделей имеем $\mathbf{\Sigma}_2 \ll \mathbf{\Sigma}_1$, то есть $\|\mathbf{\Sigma}_2\| \cdot \|\mathbf{\Sigma}_1^{-1}\| \ll 1$. Тогда

$$2\log(s(g_1,\, g_2)) \xrightarrow{\text{с.к.}} -(\mathbf{v}_1 - \mathbf{w})^{\mathsf{T}} \mathbf{\Sigma}_1^{-1} (\mathbf{v}_1 - \mathbf{w}) \;\text{ при }\; \|\mathbf{\Sigma}_2\| \to 0,$$

$$s(g_1,\, g_2) \xrightarrow{p} \exp(-\tfrac{1}{2}(\mathbf{v}_1 - \mathbf{w})^{\mathsf{T}} \mathbf{\Sigma}_1^{-1} (\mathbf{v}_1 - \mathbf{w})) \;\text{ при }\; \|\mathbf{\Sigma}_2\| \to 0.$$

Следствие 1. При выполнении условий теоремы выполнено

$$2\log(s(g_1,\, g_2)) \xrightarrow{p} -(\mathbf{v}_1 - \mathbf{w})^{\mathsf{T}} \mathbf{\Sigma}_1^{-1} (\mathbf{v}_1 - \mathbf{w}) \;\text{ при }\; \|\mathbf{\Sigma}_2\| \to 0.$$

Следствие 2. При выполнении условий теоремы выполнено

$$-2\log(s(g_1,\, g_2)) \xrightarrow{d} \chi^2(n) \;\text{ при }\; \|\mathbf{\Sigma}_2\| \to 0, \text{ где}$$

n – число признаков в модели.

Свойства s-score для нормальных распределений

Теорема 5 (Адуенко, 2014). Пусть модели, задаваемые $(\mathbf{v}_1,\, \mathbf{\Sigma}_1)$ и $(\mathbf{v}_2,\, \mathbf{\Sigma}_2)$ считаются разными, если

$$s\left(\mathcal{N}(\mathbf{v}_1,\, \mathbf{\Sigma}_1),\, \mathcal{N}(\mathbf{v}_2,\, \mathbf{\Sigma}_2)\right) \leq C \in (0,\, 1).$$

Тогда, если указанные модели разные по приведенному критерию, то

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

Теорема 6 (Адуенко, 2014). Пусть рассматриваются K моделей с $\|\mathbf{v}_1\| = \ldots = \|\mathbf{v}_K\| = \lambda_1 > 0$ и $\mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_K = \lambda_2 \mathbf{I}$. В качестве критерия отличимости моделей рассматривается следующий: модели с номерами $i \neq j$ разные, если

$$s\left(\mathcal{N}(\mathbf{v}_i,\, \mathbf{\Sigma}_i), \mathcal{N}(\mathbf{v}_j,\, \mathbf{\Sigma}_j)\right) \leq C \in (0,\, 1).$$

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

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

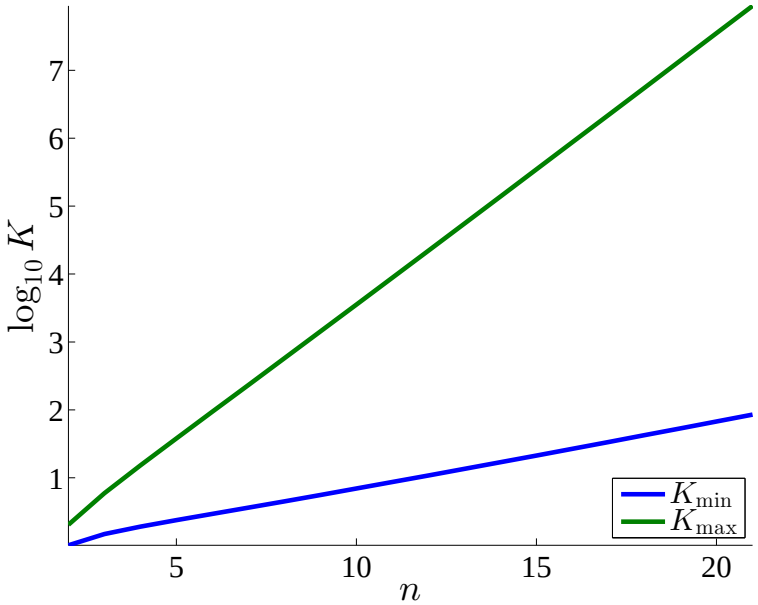
Здесь $\theta \in [0,\, \pi]$, $\cos\theta = \rho = \max(-1,\, 1 + 2\lambda_2/\lambda_1^2 \ln C)$, n – размерность признакового пространства. При этом можно построить K_{min} попарно различимых моделей, где

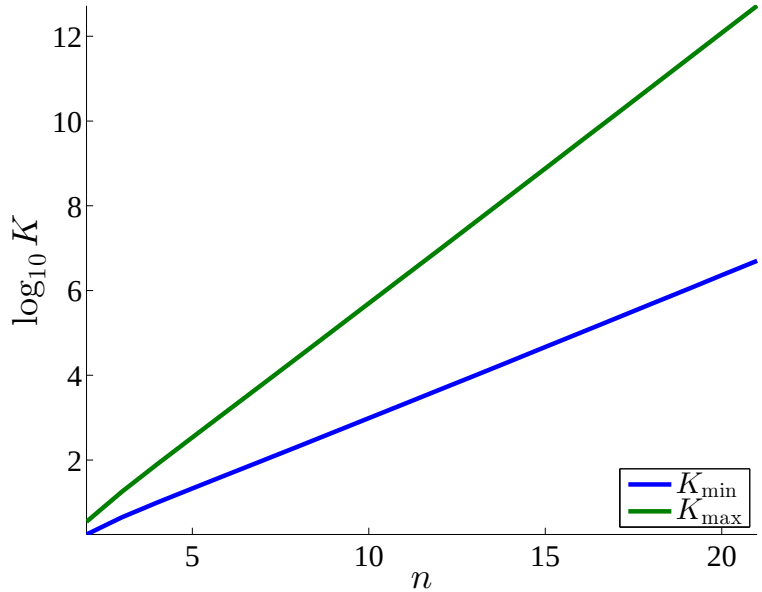
$$K_{\min} = \sqrt{\pi} \frac{n\Gamma(\frac{n+1}{2})}{(n-1)\Gamma(\frac{n}{2}+1)} \frac{1}{\int_0^{\theta} \sin^{n-2}\varphi d\varphi}.$$

Теорема 6 (продолжение). При C, близком к 1, имеем

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

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





Заключение

- Предложен алгоритм отбора признаков, основанный на оценке ковариационной матрицы параметров с помощью максимизации обоснованности модели.
- Построена функция s-score, которая позволяет оценить сходство двух моделей. Доказаны асимптотические свойства сходимости $s(g_1, g_2)$ и $\log(s(g_1, g_2))$.
- Построен метод статистического сравнения моделей на основании свойств введенной функции сходства.
- С помощью введенной s-score получены верхняя и нижняя оценка на число попарно различимых моделей.

Дальнейшие планы

- Рассмотреть структурные ограничения на ковариационную матрицу в алгоритме отбора признаков.
- Обобщить отбор признаков на многоклассовый случай
- Доказать свойства асимптотической сходимости при условии неизвестных $\mathbf{\Sigma}_1$ и $\mathbf{\Sigma}_2$.
- Получить оценки на число моделей в случае недиагональных ковариационных матриц моделей.