Домашняя работа по курсу «Введение в машинное обучение »

Выполнила Жунусова Асыллия Канатовна, студент 5 курса, ВМК МГУ.

Содержание

Задача 1

Рассматриваем функцию, равную косинусу угла между двумя векторами $x, z \in \mathbb{R}^d$:

$$K(\vec{x}, \vec{z}) = \cos(\vec{x}, \vec{z}) = \frac{\langle \vec{x}, \vec{z} \rangle}{|\vec{x}| \cdot |\vec{z}|} = \left\langle \frac{\vec{x}}{|\vec{x}|}, \frac{\vec{z}}{|\vec{z}|} \right\rangle = \langle \psi(\vec{x}), \psi(\vec{z}) \rangle$$

Значит $K(\vec{x}, \vec{z})$ - ядро.

Задача 2

Рассматриваем следующую функцию на пространстве вещественных чисел:

$$K(x, z) = \frac{1}{1 + e^{-xz}}.$$

Допустим $K(x, z) = \dfrac{1}{1 + e^{-xz}}$ - ядро, т. е. $\exists$ функция $\psi(x)$, такая что $K(x, z) = \psi(x) \cdot \psi(z)$.

Тогда $1 + e^{-xz}$ - тоже ядро, т.к. $1 + e^{-xz} = \widetilde{\psi}(x) \cdot \widetilde{\psi}(z)$, где $\widetilde{\psi}(x) = \dfrac{1}{\psi(x)}$.

Однако в силу свойств степени e^{-xz} невозможно представить виде произведении функции $f(x)$ и $g(z)$.

Следовательно не существует такие функции $\tilde{f}(x)$ и $\widetilde{g}(z)$: $1 + e^{-xz} = \tilde{f}(x) \cdot \widetilde{g}(z)$.

Это означает, что не существует функции $\widetilde{\psi}(x)$, следовательно, не существует функции $\psi(x)$. Получаем противоречие. Значит $K(x, z)$ - не ядро.

Задача 3.

$$\begin{cases} \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l}\left(\xi_i - \nu p\right) \to \min_{w,\xi,\nu}, \xi = \left(\xi_i\right)_{i=1}^{l} \\ \left\langle w, x_i \right\rangle \geq p - \xi_i, \quad \xi_i \geq 0, \quad i = 1,2,...,l \end{cases} \tag{3.1}$$

Очевидно, что функции $g_{1i}(w, \xi_i) = p - \xi_i - \langle w, x_i \rangle$ и $g_{2i}(\xi_i) = -\xi_i$ являются выпуклыми, так как g_{2i} - это прямая, а g_{1i} описывает плоскость.

Рассмотрим функцию $f(w, \xi, \nu) = \dfrac{1}{2}\|w\|^2 + \sum\limits_{i=1}^{l}\left(\xi_i - \nu p\right)$.

Ее вторые производные $\dfrac{d^2 f}{dw^2} = \dfrac{1}{2} > 0, \dfrac{d^2 f}{d\xi^2} = 0$ и $\dfrac{d^2 f}{d\nu^2} = 0$, следовательно,

функция f тоже выпуклая. Это означает, что задача (3.1) является задачей выпуклого программирования, так как все функции выпуклые. Значит, можем использовать теоремой Каруша-Куна-Таккера.

Построим функцию Лагранжа.

$$L = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l}(\xi_i - \nu p) - \sum_{i=1}^{l}\lambda_i(\left\langle w, x_i \right\rangle - p + \xi_i) - \sum_{i=1}^{l}\eta_i\xi_i =$$

$$= \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l}\xi_i(1 - \lambda_i - \eta_i) + \sum_{i=1}^{l}(\lambda_i p - \lambda_i\left\langle w, x_i \right\rangle - \nu p)$$

Необходимые условия седовой точки функции Лагранжа:

$$\frac{dL}{dw} = w - \sum_{i=1}^{l}\lambda_i x_i = 0 \Rightarrow w = \sum_{i=1}^{l}\lambda_i x_i;$$
$$\frac{dL}{dp} = \sum_{i=1}^{l}\lambda_i - \nu l = 0 \Rightarrow \sum_{i=1}^{l}\lambda_i = \nu l;$$
$$\frac{dL}{d\xi_i} = 1 - \lambda_i - \eta_i = 0 \Rightarrow \lambda_i + \eta_i = 1, \quad i = 1,2,...,l;$$
$$\xi_i \geq 0; \qquad \eta_i \geq 0; \qquad \lambda_i \geq 0$$
$$\lambda_i = 0 \text{ либо } \left\langle w, x_i \right\rangle = p - \xi_i$$
$$\eta_i = 0 \text{ либо } \xi_i = 0$$

В итоге получим систему условии ККТ

$$\begin{cases} w = \sum_{i=1}^{l} \lambda_i x_i; \quad \sum_{i=1}^{l} \lambda_i = \nu l; \quad \left\langle w, x_i \right\rangle \geq p - \xi_i \\ \xi_i \geq 0; \quad \eta_i \geq 0; \quad \lambda_i \geq 0; \quad \eta_i + \lambda_i = 1 \\ \lambda_i = 0 \quad \textbf{либо} \quad \left\langle w, x_i \right\rangle = p - \xi_i \\ \eta_i = 0 \quad \textbf{либо} \quad \xi_i = 0 \end{cases}$$

Построим двойственную задачу через функцию Лагранжа

$$L = \frac{1}{2} \sum_{i,j=1}^{l} \lambda_i \lambda_j \left\langle x_i, x_j \right\rangle - \sum_{i,j=1}^{l} \lambda_i \lambda_j \left\langle x_i, x_j \right\rangle - l\nu p + p \sum_{i,j=1}^{l} \lambda_i = -\frac{1}{2} \sum_{i,j=1}^{l} \lambda_i \lambda_j \left\langle x_i, x_j \right\rangle$$

Двойственная задача	Обобщение двойственной задачи
$L = -\frac{1}{2} \sum_{i,j=1}^{l} \lambda_i \lambda_j \left\langle x_i, x_j \right\rangle \to \max_{\lambda}$ $\sum_{i=1}^{l} \lambda_i = \nu l; \quad 0 \leq \lambda_i \leq 1$	$L = -\frac{1}{2} \sum_{i,j=1}^{l} \lambda_i \lambda_j K(x_i, x_j) \to \max_{\lambda}$ $\sum_{i=1}^{l} \lambda_i = \nu l; \quad 0 \leq \lambda_i \leq 1$

Двойственная задача тоже является задачей выпуклого программирования, так как $-L$ тоже является выпуклой функции (ее вторая производная положительна).

Типизация объектов $x_i, i = 1,2,...,l$

1. $\lambda_i = 0;\ \eta_i = 1;\ \xi_i = 0; \left\langle w, x_i \right\rangle > p;\ a(x_i) = 1$ - периферийный.
2. $\lambda_i \in (0,1);\ \eta_i \in (0,1);\ \xi_i = 0; \left\langle w, x_i \right\rangle = p;\ a(x_i) = 0$ - опорный-граничный
3. $\lambda_i = 1;\ \eta_i = 0;\ \xi_i > 0; \left\langle w, x_i \right\rangle = p - \xi_i < p;\ a(x_i) = -1$ - опорный-нарушитель

Покажем, что $\frac{1}{l} \sum_{i=1}^{l} \left[\left\langle w, x_i \right\rangle < p \right] \leq \nu$, где w - опт. вектор весов.

$\sum_{i=1}^{l} \left[\left\langle w, x_i \right\rangle < p \right]$ -кол-во элементов в множестве $\left\{ p - \xi_i \leq \left\langle w, x_i \right\rangle < p \right\} = A$.

Кол-во элементов в мн-ве A столько же, сколько $\lambda_i = 1$.

$$\sum_{i=1}^{l} \left[\left\langle w, x_i \right\rangle < p \right] = \sum_{i=1}^{l} \left[\lambda_i = 1 \right] \leq \sum_{i=1}^{l} \lambda_i = 1 = \nu l.$$

То есть $\sum_{i=1}^{l} \left[\left\langle w, x_i \right\rangle < p \right] \leq \nu l \iff \frac{1}{l} \sum_{i=1}^{l} \left[\left\langle w, x_i \right\rangle < p \right] \leq \nu.$

Задача 4

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} p_i(\max\{1 - y_i(\langle w, x_i\rangle + b), 0\})^2 \to \min_{w,b}, \quad C > 0 \qquad (4.1)$$

Перепишем задачу безусловной оптимизации (4.1) в виде эквивалентной задачи условной оптимизации

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} p_i\xi_i^2 \to \min_{w,b,\xi} \\ \xi_i \geq 1 - y_i(\langle w, x_i\rangle + b); \\ \xi_i \geq 0, \quad i = 1,2,...,l \end{cases}$$

Построим функцию Лагранжа.

$$L = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l} \xi_i(Cp_i\xi_i - \lambda_i - \eta_i) + \sum_{i=1}^{l} \lambda_i(1 - y_i(\langle w, x_i\rangle + b))$$

Необходимые условия седовой точки функции Лагранжа:

$$0 = \frac{dL}{dw} = w - \sum_{i=1}^{l} \lambda_i y_i x_i \implies w = \sum_{i=1}^{l} \lambda_i y_i x_i$$

$$0 = \frac{dL}{db} = \sum_{i=1}^{l} \lambda_i y_i \implies \sum_{i=1}^{l} \lambda_i y_i = 0$$

$$0 = \frac{dL}{d\xi_i} = 2\xi_i Cp_i - \lambda_i - \eta_i \implies \lambda_i + \eta_i = 2\xi_i Cp_i$$

$$\lambda_i \geq 0; \quad \eta_i \geq 0$$

Построим двойственную задачу через функцию Лагранжа

$$L = \frac{1}{2}\sum_{i,j=1}^{l} \lambda_i\lambda_j y_i y_j\langle x_i, x_j\rangle - \sum_{i=1}^{l} \xi_i Cp_i + \sum_{i=1}^{l} \lambda_i - \sum_{i,j=1}^{l} \lambda_i\lambda_j y_i y_j\langle x_i, x_j\rangle$$

Двойственная задача

$$\begin{cases} L = -\frac{1}{2}\sum_{i,j=1}^{l} \lambda_i\lambda_j y_i y_j K(x_i, x_j) - C\sum_{i=1}^{l} \xi_i p_i + \sum_{i=1}^{l} \lambda_i \to \max_{\lambda,\xi} \\ \sum_{i=1}^{l} \lambda_i y_i = 0; \quad 0 \leq \lambda_i \leq 2Cp_i\xi_i, \quad i = 1,2,...,l \end{cases}$$

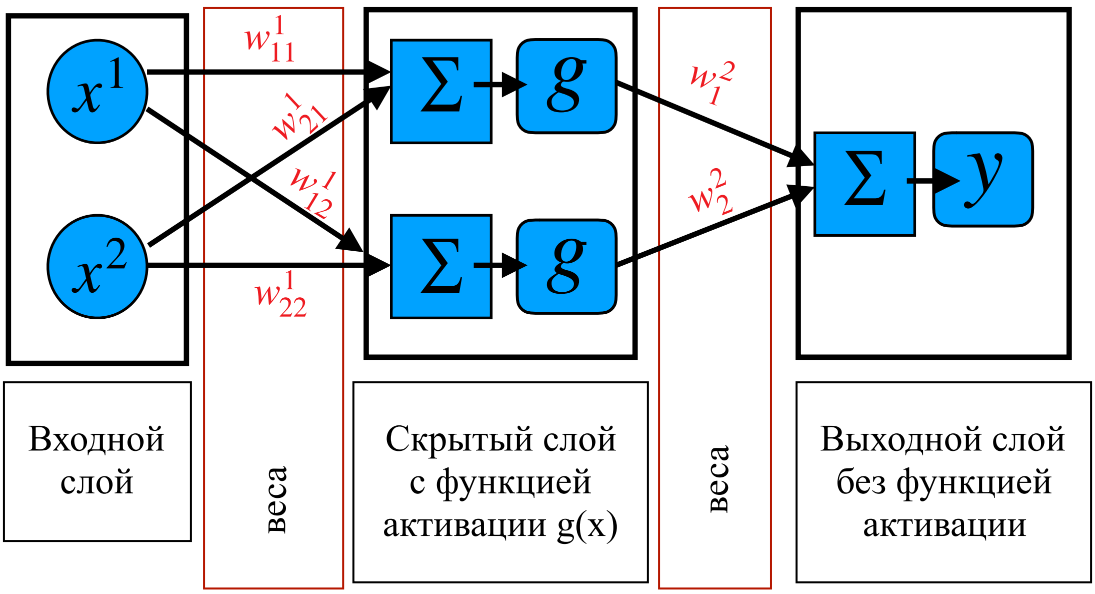
Решение прямой задачи

$$\begin{cases} w = \sum_{i=1}^{l} \lambda_i y_i x_i \\ b = y_i - \langle w, x_i \rangle, \text{ для } \forall i : \lambda_i \in (0, 2Cp_i\xi_i) \end{cases}$$

$$a(x) = \sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b$$

Задача 5

Схематичное представление нейронной сети, на вход подается вектор $x = [x^1, x^2]$, потом на входной вектор умножают справа на матрицу весов $w^1 = \begin{pmatrix} w_{11}^1 & w_{12}^1 \\ w_{21}^1 & w_{22}^1 \end{pmatrix}$, результаты которого проходят через функцию активации $g(x)$, потом умножают на матрицу весов $w^2 = \begin{pmatrix} w_1^2 \\ w_2^2 \end{pmatrix}$, на выходе мы получаем скаляр y.



Определим вектор весов $w^1 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ и $w^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Проверим, что вернет нейронная сеть при x_1, x_2, x_3, x_4.

$y_1 = g(x_1 w^1) w^2 = g([0\ 0] w^1) w^2 = g([0\ 0]) w^2 = [0\ 0] w^2 = 0$

$y_2 = g(x_2 w^1)w^2 = g([0\ 1]w^1)w^2 = g([\text{-}1\ 1])w^2 = [0\ 1]w^2 = 1$

$y_3 = g(x_3 w^1)w^2 = g([1\ 0]w^1)w^2 = g([1\ \text{-}1])w^2 = [1\ 0]w^2 = 1$

$y_4 = g(x_4 w^1)w^2 = g([1\ 1]w^1)w^2 = g([0\ 0])w^2 = [0\ 0]w^2 = 0$

Задача 6

Критерий информативности $H(R) = \min_{c \in \mathbb{Y}} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} L(y_i, c)$.

1. $L(y, c) = (y - c)^2$

$$f(c) = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} (y_i - c)^2$$

$$\frac{df(c)}{dc} = -\frac{2}{R} \sum_{(x_i, y_i) \in R} (y_i - c) = 0 \implies c = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} y_i = \bar{y}$$

$$\frac{d^2 f(c)}{dc^2} = 2 > 0 \implies c = \bar{y} \text{ - точка минимума}$$

$$H(R) = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} (y_i - \bar{y})^2$$

2. $L(y, c) = \sum_{k=1}^{K} (c_k - [y = k])^2$. Причем $\sum_{k=1}^{K} c_k = 1$.

Введем обозначения $p_k = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y_i = k]$.

$$H(R) = \min_{c \in \mathbb{Y}} \frac{1}{|R|} \sum_{(x_i, y_i) \in R} \sum_{k=1}^{K} \left(c_k - [y = k]\right)^2 =$$

$$= \min_{c \in \mathbb{Y}} \sum_{k=1}^{K} \left(c_k^2 - 2c_k \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y = k] + \frac{1}{|R|} \sum_{(x_i, y_i) \in R} [y = k] \right) =$$

$$= \min_{c \in \mathbb{Y}} \sum_{k=1}^{K} \left(c_k^2 + p_k (1 - 2c_k) \right)$$

$$f_k(c_k) = c_k^2 + p_k(1 - 2c_k)$$

$$\frac{df_k(c_k)}{dc_k} = 2c_k - 2p_k = 0 \implies c_k = p_k;$$

$$\frac{d^2f_k(c_k)}{dc_k^2} = 2 > 0 \implies c_k = p_k \text{ - точка минимума}$$

$$H(R) = \sum_{k=1}^{K} (p_k^2 + p_k - 2p_k^2) = \sum_{k=1}^{K} p_k(1 - p_k)$$

3. $L(y, c) = -\sum_{k=1}^{K} [y = k] \log c_k$

$$\begin{cases} H(R) = -\frac{1}{|R|} \sum_{(x_i,y_i) \in R} \sum_{k=1}^{K} [y = k] \log c_k = -\sum_{k=1}^{K} p_k \log c_k \to \min\{c\} \\ \sum_{k=1}^{K} c_k = 1 \end{cases}$$

Построим функцию Лагранжа

$$L = -\sum_{k=1}^{K} p_k \log c_k + \lambda \Big(\sum_{k=1}^{K} c_k - 1 \Big)$$

$$\frac{dL}{dc_k} = -\frac{p_k}{c_k} + \lambda = 0 \implies c_k = \frac{p_k}{\lambda} \text{ - точка минимума}$$

$$1 = \sum_{k=1}^{K} c_k = \sum_{k=1}^{K} \frac{p_k}{\lambda} = \frac{1}{\lambda} \sum_{k=1}^{K} p_k = \frac{1}{\lambda} \implies \lambda = 1$$

$$H(R) = -\sum_{k=1}^{K} p_k \log p_k$$

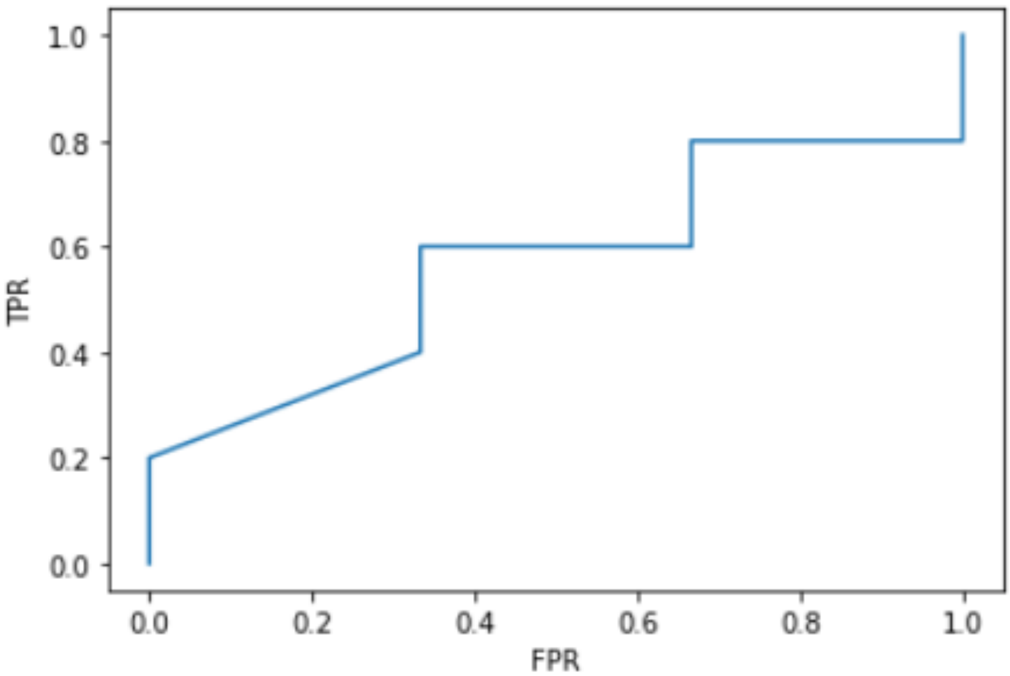
Задача 7

$b(x_1) \le b(x_3) \le b(x_6) \le b(x_7) \le b(x_2) \le b(x_4) \le b(x_5) \le b(x_8)$

0.1	0.2	0.3	0.6	0.8	0.9	0.9	0.95
+1	− 1	+ 1	− 1	+ 1	− 1	+ 1	+ 1

$a(x) = sign(b(x) - t)$

t=0.95 FPR=0 TPR=0	t=0.9 FPR=0 TPR=1/5=0.2	t=0.8 FPR=1/3=0.3(3) TPR=2/5=0.4	t=0.6 FPR=1/3=0.3(3) TPR=3/5=0.6
t=0.3 FPR=2/3=0.6(6) TPR=3/5=0.6	t=0.2 FPR=2/3=0.6(6) TPR=4/5=0.8	t=0.1 FPR=1 TPR=4/5=0.8	t=0 FPR=1 TPR=1



Roc-кривая

$$AUC - ROC = \frac{1}{3} \cdot \frac{0.2 + 0.4}{2} + \frac{1}{3} \cdot 0.6 + \frac{1}{3} \cdot 0.8 = \frac{17}{30} \approx 0.56667$$

Задача 8

Разберем процедуру построение AUC-ROC при

$$b(x_{(1)}) < b(x_{(2)}) < \ldots < b(x_{(n)})$$

Сначала все объекты сортируются по оценке классификатора. Стартуем из точки (0,0) - она соответствует порогу $y_{(l)}$. Начинаем идти от большей оценки к меньшей.

Если текущий объект имеет класс «1», то у алгоритма увеличивается TPR; значит, ROC-кривая сдвигается вверх на $\frac{1}{l_+}$ (l_+-число объектов в положительного класса).

Если у текущего объекта класс «-1», то у алгоритма увеличивается FPR; значит, ROC-кривая сдвигается вправо на $\frac{1}{l_-}$ (l_--число объектов отрицательного класса), а к AUC надо прибавить $\sum_{j=i+1}^{l} [y_{(j)} = +1] \cdot \frac{1}{l_+ l_-}$.

Получаем $AUC = \frac{1}{l_+ l_-} \sum_{i=1}^{l} [y_{(i)} = -1] \sum_{j=i+1}^{l} [y_{(j)} = +1] = \frac{1}{l_+ l_-} \sum_{i<j} [y_{(i)} < y_{(j)}]$.

Однако в случае $\exists k < l : b(x_{(k)}) = b(x_{(k+1)})$ и $y_{(k)} < y_{(k+1)}$, такое утверждение является не верным.

Приведу контр-пример.

$$b(x_1) = 0.5 \quad y_1 = -1$$
$$b(x_2) = 0.5 \quad y_2 = +1$$

$$\frac{1}{2} = AUC < \frac{1}{l_+ l_-} \sum_{i<j} [y_{(i)} < y_{(j)}] = 1$$

Задача 9

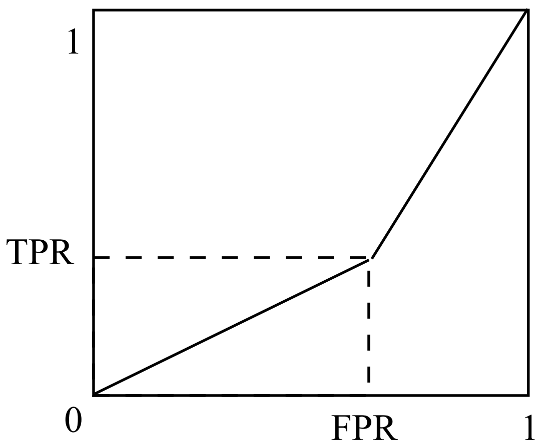
1. Т. к. классификатор b(x) возвращает только 2 значения, то, отсортировав его, получим:

$$0 = b(x_{(1)}) = b(x_{(2)}) = \ldots = b(x_{(l_-)}) < b(x_{(l_-+1)}) = b(x_{(l_-+2)}) = \ldots = b(x_{(l)}) = 1$$

с истинными ответами $y_{(1)}, y_{(2)}, \ldots, y_{(l)}$ соотвественно.

В этом случае у ROC-кривой будут 3 точки:

t=1 FPR=0 TPR=0	t=0 FPR=$\frac{1}{l_-}\sum\limits_{i=1}^{l_+}[y_{(l_-+i)}=-1]$ TPR=$\frac{1}{l_+}\sum\limits_{i=1}^{l_+}[y_{(l_-+i)}=+1]$	t<0 FPR=1 TPR=1



2. $$AUC = \frac{TPR \cdot FPR}{2} + \frac{(1+TPR)\cdot(1-FPR)}{2} =$$

$$= \frac{1 - FPR + TPR}{2} = \frac{TNR + TPR}{2}$$

$$TNR = \frac{1}{l_-}\sum_{i=1}^{l_-}[y_{(i)} = -1]$$

$$AUC = \frac{1}{2l_-}\sum_{i=1}^{l_-}[y_{(i)} = -1] + \frac{1}{2l_+}\sum_{i=1}^{l_+}[y_{(l_-+i)} = +1]$$

3. Если $l_- = l_+$, то

$$AUC = \frac{1}{2l_+}\left(\sum_{i=1}^{l_-}[y_{(i)} = -1] + \sum_{i=1}^{l_+}[y_{(l_-+i)} = +1]\right) = \frac{TP + TN}{2} = accuracy.$$

Задача 10

Сначала посчитаем мат. ожидание и дисперсию для отрицательного биномиального распределения.

$$P(y=k)=C_{k+r-1}^{k}p^{r}q^{k}.$$

Обозначим мат ожидание A и вычислим ее

$$A=Ey=\sum_{k=0}^{+\infty}k\cdot C_{k+r-1}^{k}p^{r}q^{k}=\sum_{k=1}^{+\infty}k\cdot\frac{(k+r-1)!}{k!(r-1)!}p^{r}q^{k}=$$

$$=q\sum_{k=0}^{+\infty}k\cdot\frac{(k+r-1)!(k+r)}{k!(r-1)!}p^{r}q^{k}=q\sum_{k=0}^{+\infty}k\cdot C_{k+r-1}^{k}p^{r}q^{k}+$$

$$+qr\sum_{k=0}^{+\infty}C_{k+r-1}^{k}p^{r}q^{k}=Aq+qr$$

$$A=\frac{rq}{1-q}=\frac{rq}{p};\qquad \text{то есть } Ey=\frac{rq}{p}.$$

Обозначим второй начальный момент B и вычислим ее

$$B=Ey^{2}=\sum_{k=0}^{+\infty}k^{2}\cdot C_{k+r-1}^{k}p^{r}q^{k}=\sum_{k=0}^{+\infty}(k^{2}-k+k)\cdot C_{k+r-1}^{k}p^{r}q^{k}=$$

$$=\sum_{k=2}^{+\infty}k(k-1)\cdot\frac{(k+r-1)!}{k!(r-1)!}p^{r}q^{k}+\sum_{k=1}^{+\infty}k\cdot\frac{(k+r-1)!}{k!(r-1)!}p^{r}q^{k}=$$

$$=q^{2}\sum_{k=0}^{+\infty}(k+r+1)(k+r)\cdot\frac{(k+r-1)!}{k!(r-1)!}p^{r}q^{k}+A=$$

$$=q^{2}\sum_{k=0}^{+\infty}k^{2}\cdot C_{k+r-1}^{k}p^{r}q^{k}+q^{2}(2r+1)\sum_{k=0}^{+\infty}k\cdot C_{k+r-1}^{k}p^{r}q^{k}+$$

$$+q^{2}(r+r^{2})\sum_{k=0}^{+\infty}C_{k+r-1}^{k}p^{r}q^{k}+\frac{rq}{p}=$$

$$=q^{2}B+\frac{q^{3}(2r+1)r}{p}+q^{2}(r+1)r+\frac{rq}{p}$$

$$B(1-q^{2})=\frac{1}{p}rq(q+1)(qr+1)$$

$$B = \frac{r^2q^2}{p^2} + \frac{rq}{p^2}$$

Дисперсия по формуле вычисляется

$$Dy = Ey^2 - (Ey)^2 = B - A^2 = \frac{rq}{p^2}$$

Построим обобщенную линейную модель:

$$p(y\,|\,\theta_i, \phi_i) = \exp\left(\frac{y_i\theta_i - c(\theta_i)}{\phi_i} + h(y_i, \phi_i)\right)$$

Известны следующие свойства :

1) $Ey_i = c'(\theta_i)$;
2) $Dy_i = \phi_i c''(\theta_i)$;

$$P(y_i\,|\,\mu(x_i)) = C_{y_i+r-1}^{y_i}\mu(x_i)^r(1-\mu(x_i))^{y_i} = \exp\left(y_i\ln(1-\mu(x_i)) + r\ln\mu(x_i) + \ln C_{y_i+r-1}^{y_i}\right)$$

$$\theta_i = \ln(1-\mu(x_i)) \implies \mu(x_i) = 1 - e^{\theta_i}$$

$$Ey_i = \frac{r(1-\mu(x_i))}{\mu(x_i)} = \frac{r}{e^{-\theta_i} - 1} = c'(\theta_i) \implies c(\theta_i) = -\,r\ln(e^{\theta_i} - 1)$$

$$Dy_i = \frac{r(1-\mu(x_i))}{\mu^2(x_i)} = \frac{re^{\theta_i}}{(1-e^{\theta_i})^2} = c''(\theta_i)\cdot\phi_i$$

$$c''(\theta_i) = -\frac{re^{\theta_i}}{(e^{\theta_i}-1)^2} \implies \phi_i = \frac{Dy_i}{c''(\theta_i)} = -\,1$$

То есть

$$P(y_i\,|\,\mu(x_i)) = \exp\left(y_i\ln(1-\mu(x_i)) + r\ln\mu(x_i) + \ln C_{y_i+r-1}^{y_i}\right) =$$
$$= \exp(y_i\theta_i - c(\theta_i) + h(y_i, \phi_i)) = p(y_i\,|\,\theta_i, \phi_i)$$

Метод максимального правдоподобия

$$L(w) = \prod_{i=1}^{l} p(y_i\,|\,\theta_i, \phi_i) \to \max_w$$

но удобнее прологарифмировать и решать следующую оптимизационную задачу

$$\ln L(w) = \sum_{i=1}^{l} (y_i\theta_i - c(\theta_i) + h(y_i, \phi_i)) \to \max_{w},$$

где $\theta_i = \ln(1 - \sigma(\langle w, x_i \rangle))$,

$c(\theta_i) = -r\ln(e^{\theta_i} - 1)$,

$\phi_i = 1, h(y_i, \phi_i) = \ln C_{y_i+r-1}^{y_i}$.

Итоговый алгоритм прогнозирования $y(x) = \dfrac{r(1 - \sigma(\langle w, x \rangle))}{\sigma(\langle w, x \rangle)}$

Задача 11

Вероятность события $x \in \sigma_i$ равно h_i,

а вероятность события $x \notin \sigma_i$ равно $1 - h_i$.

Запишем плотность распределения того, что случайная величина x попадет в область σ_i.

$$f(x) = h_i^{[x\in\sigma_i]} \cdot (1 - h_i)^{[x\notin\sigma_i]}.$$

$$L = \prod_{j=1}^{l} f(x_j) = h_i^{\sum_{j=1}^{l} [x_j\in\sigma_i]} \cdot (1 - h_i)^{\sum_{j=1}^{l} (1-[x_j\in\sigma_i])}$$

$$l = \ln L = \sum_{j=1}^{l} [x_j \in \sigma_i]\ln h_i + \sum_{j=1}^{l} (1 - [x_j \in \sigma_i])\ln(1 - h_i)$$

$$\frac{dl}{dh_i} = \sum_{j=1}^{l} \left(\frac{[x_j \in \sigma_i]}{h_i} - \frac{1 - [x_j \in \sigma_i]}{1 - h_i} \right) = 0$$

$$\sum_{j=1}^{l} ([x_j \in \sigma_i] - h_i) = \sum_{j=1}^{l} [x_j \in \sigma_i] - lh_i = 0$$

$h_i = \dfrac{1}{l} \displaystyle\sum_{j=1}^{l} [x_j \in \sigma_i]$ - оптимальное значение h_i полученное с помощью

метода максимального правдоподобия.

Задача 12

Пусть $m = |\mathbb{Y}|$ и $y_i \in \{1, \ldots, m\}$.

Тогда

$$\sum_{i=1}^{l} \log \pi_{y_i} p(x_i | y_i) + \sum_{i=l+1}^{l+k} \log \sum_{j=1}^{m} \pi_j p(x_i | j) \to \max_{\pi,\theta}$$

$$\sum_{y \in \mathbb{Y}} \pi_y = \sum_{j=1}^{m} \pi_j = 1 \tag{12.1}$$

Причем $\pi_j = p(j) \in (0,1)$ и $p(x_i) = \sum_{j=1}^{m} p(x_i | j) p(j) = \sum_{j=1}^{m} \pi_j p(x_i | j)$.

1) Е-шаг

Построим матрицу g_{ij}:

при $1 \le i \le l$ $\qquad g_{ij} = p(j | x_i) = [y_i = j]$

при $l + 1 \le i \le l + k$ $\qquad g_{ij} = p(j | x_i) = \dfrac{p(j) p(x_i | j)}{p(x_i)} = \dfrac{\pi_j p(x_i | j)}{p(x_i)}$

2) **М - шаг**

Запишем Лагранжиан

$$L(\pi, \theta) = \sum_{i=1}^{l} \log \pi_{y_i} p(x_i | y_i) + \sum_{i=l+1}^{l+k} \log \sum_{j=1}^{m} \pi_j p(x_i | j) - \lambda \Big(\sum_{j=1}^{m} \pi_j - 1 \Big) \to \max_{\pi,\theta}$$

Необходимые условия седовой точки функции Лагранжа:

a) $\dfrac{dL}{d\pi_a} = 0$ $\qquad$ b) $\dfrac{dL}{d\theta_a} = 0$

<u>Рассмотрим пункт а</u>

$$\frac{dL}{d\pi_a} = \sum_{i=1}^{l} \frac{p(x_i | y_i)[y_i = a]}{\pi_{y_i} p(x_i | y_i)} + \sum_{i=l+1}^{l+k} \frac{p(x_i | a)}{p(x_i)} - \lambda =$$

$$= \frac{1}{\pi_a} \sum_{i=1}^{l} [y_i = a] + \frac{1}{\pi_a} \sum_{i=l+1}^{l+k} g_{ia} - \lambda = \frac{1}{\pi_a} \sum_{i=1}^{l+k} g_{ia} - \lambda = 0$$

В итоге получим, что $\pi_a = \frac{1}{\lambda}\sum_{i=1}^{l+k} g_{ia}$.

Подставим полученное π_a в (12.1) :

$$1 = \sum_{a=1}^{m} \pi_a = \sum_{a=1}^{m}\left(\frac{1}{\lambda}\sum_{i=1}^{l+k} g_{ia}\right) = \frac{1}{\lambda}\sum_{i=1}^{l+k}\sum_{a=1}^{m} g_{ia} = \frac{1}{\lambda}\sum_{i=1}^{l+k} 1 = \frac{l+k}{\lambda}$$

то есть $\lambda = l + k$

В итоге получаем $\pi_a = \frac{1}{l+k}\sum_{i=1}^{l+k} g_{ia}$

<u>Рассмотрим пункт b</u>

$$\frac{dL}{d\theta_a} = \sum_{i=1}^{l} \frac{[y_i = a]\pi_{y_i}\frac{d}{d\theta_a}p(x_i|a)}{\pi_{y_i}p(x_i|y_i)} + \sum_{i=l+1}^{l+k}\frac{\pi_a\frac{d}{d\theta_a}p(x_i|a)}{p(x_i)} =$$

$$= \sum_{i=1}^{l}[y_i = a]\frac{d}{d\theta_a}\ln p(x_i|a) + \sum_{i=l+1}^{l+k}\frac{\pi_a p(x_i|a)}{p(x_i)}\frac{d}{d\theta_a}\ln p(x_i|a) =$$

$$= \sum_{i=1}^{l+k} g_{ia}\frac{d}{d\theta_a}\ln p(x_i|a) = \frac{d}{d\theta_a}\left(\sum_{i=1}^{l+k} g_{ia}\ln p(x_i|a)\right) = 0$$

Значит $\theta_a = \arg\max_{\theta}\sum_{i=1}^{l+k} g_{ia}\ln p(x_i|a)$, где $a \in \{1,\ldots,m\}$.

В итоге получаем EM-алгоритм в общем виде:

Е-шаг:

$$g_{ij} = \begin{cases} [y_i = j] & \textbf{при } 1 \le i \le l \\ \frac{\pi_j p(x_i|j)}{p(x_i)} & \textbf{при } l+1 \le i \le l+k \end{cases} \tag{12.2}$$

М-шаг:

$$\pi_a = \frac{1}{l+k}\sum_{i=1}^{l+k} g_{ia}, \text{ где } a \in \{1,\ldots,m\}. \tag{12.3}$$

$$\theta_a = \arg\max_{\theta}\sum_{i=1}^{l+k} g_{ia}\ln p(x_i|a), \text{ где } a \in \{1,\ldots,m\} \tag{12.4}$$

Теперь запишем для смеси гауссиан

Плотность гауссовского распределение при параметрах (μ_y, Σ_y).

$$p(x_i\,|\,j) = \frac{\exp(-\frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j))}{\sqrt{(2\pi)^n \det \Sigma_j}}$$

Подставим эту плотность в (12.2) и получим Е-шаг:

$$g_{ij} = \begin{cases} [y_i = j] & \textbf{при } 1 \le i \le l \\ \dfrac{\frac{\pi_j \exp(-\frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j))}{\sqrt{(2\pi)^n \det \Sigma_j}}}{\sum_{s=1}^{m} \frac{\pi_s \exp(-\frac{1}{2}(x_i-\mu_s)^T\Sigma_s^{-1}(x_i-\mu_s))}{\sqrt{(2\pi)^n \det \Sigma_s}}} & \textbf{при } l+1 \le i \le l+k \end{cases}$$

Формула (12.3) остается без изменении, поэтому будем просчитывать

$$\text{a)} \quad \frac{dL}{d\mu_j} = 0 \quad \text{и} \quad \text{b)}\ \frac{dL}{d\Sigma_j} = 0.$$

<u>Рассмотрим пункт a</u>

$$\ln p(x_i\,|\,j) = -\frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j) - \frac{1}{2}\ln(2\pi)^n - \frac{1}{2}\ln\det\Sigma_j$$

$$\frac{d}{d\mu_j}\ln p(x_i\,|\,j) = \Sigma_j^{-1}(x_i-\mu_j)$$

$$\frac{dL}{d\mu_j} = \sum_{i=1}^{l+k} g_{ij}\frac{d}{d\mu_j}\ln p(x_i\,|\,j) = \sum_{i=1}^{l+k} g_{ij}\Sigma_j^{-1}(x_i-\mu_j) = \Sigma_j^{-1}\sum_{i=1}^{l+k} g_{ij}(x_i-\mu_j) = 0$$

В итоге находим параметр $\mu_j = \dfrac{\sum_{i=1}^{l+k} g_{ij}x_i}{\sum_{i=1}^{l+k} g_{ij}}$.

<u>Рассмотрим пункт b</u>

Введем обозначения $\Lambda = \Sigma^{-1}$ и вместо $\dfrac{dL}{d\Sigma}$ будем считать $\dfrac{dL}{d\Lambda}$.

Тогда перепишем прологарифмированную плотность

$$\ln p(x_i\,|\,j) = -\frac{1}{2}(x_i-\mu_j)^T\Lambda_j(x_i-\mu_j) + \frac{1}{2}\ln det\Lambda_j - \frac{n}{2}\ln 2\pi$$

Причем сделаем следующее преобразования

$$(x_i-\mu_j)^T\Lambda_j(x_i-\mu_j)=trace(\Lambda_j(x_i-\mu_j)(x_i-\mu_j)^T)$$

Тогда

$$\frac{d}{d\Lambda_j}\ln p(x_i|j)=-\frac{1}{2}\frac{d}{d\Lambda_j}trace(\Lambda_j(x_i-\mu_j)(x_i-\mu_j)^T)+\frac{1}{2\det\Lambda_j}\frac{d}{d\Lambda_j}\det\Lambda_j$$

Учитывая, что $\frac{d}{dA}trace(AB)=B^T$ и $\frac{d}{dA}\det A=A^{-1}\det A$, получаем

$$\frac{d}{d\Lambda_j}\ln p(x_i|j)=-\frac{1}{2}(x_i-\mu_j)(x_i-\mu_j)^T+\frac{1}{2}\Lambda^{-1}$$

$$\frac{dL}{d\Lambda_j}=\sum_{i=1}^{l+k}g_{ij}\frac{d}{d\Lambda_j}\ln p(x_i|j)=-\frac{1}{2}\sum_{i=1}^{l+k}g_{ij}\Big((x_i-\mu_j)(x_i-\mu_j)^T-\Lambda_j^{-1}\Big)=0$$

$$\Lambda_j^{-1}=\frac{\sum_{i=1}^{l+k}g_{ij}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{l+k}g_{ij}}$$

Таким образом нашли $\Sigma_j=\dfrac{\sum_{i=1}^{l+k}g_{ij}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{l+k}g_{ij}}$

В итоге получаем EM-алгоритм для смеси гауссиан :

E-шаг:

$$g_{ij}=\begin{cases}[y_i=j] & \textbf{при } 1\le i\le l\\ \dfrac{\frac{\pi_j\exp(-\frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j))}{\sqrt{(2\pi)^n\det\Sigma_j}}}{\sum_{s=1}^{m}\frac{\pi_s\exp(-\frac{1}{2}(x_i-\mu_s)^T\Sigma_s^{-1}(x_i-\mu_s))}{\sqrt{(2\pi)^n\det\Sigma_s}}} & \textbf{при } l+1\le i\le l+k\end{cases}$$

M- шаг:

$$\pi_j=\frac{1}{l+k}\sum_{i=1}^{l+k}g_{ij}\qquad\qquad \mu_j=\frac{\sum_{i=1}^{l+k}g_{ij}x_i}{\sum_{i=1}^{l+k}g_{ij}}$$

$$\Sigma_j=\frac{\sum_{i=1}^{l+k}g_{ij}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{l+k}g_{ij}}$$

Задача 13

Рассмотрим выборку, имеющее следующее распределение

$$P(y_i) = P(i < z)P(y_i \,|\, i < z) + P(i \geq z)P(y_i \,|\, i \geq z)$$

Причем известно, что для всех $i < z$

$$P(y_i \,|\, i < z) = [y_i = 1]\theta_1 + [y_i = 0](1 - \theta_1),$$

а для всех $i \geq z$

$$P(y_i \,|\, i \geq z) = [y_i = 1]\theta_2 + [y_i = 0](1 - \theta_2)$$

Обозначим

$$w_i := P(i < z),$$
$$P_1(y_i) := P(y_i \,|\, i < z),$$
$$P_2(y_i) := P(y_i \,|\, i \geq z)$$

E-шаг:

$$g_{i1} = P(i < z \,|\, y_i) = \frac{P(i < z)P(y_i \,|\, i < z)}{P(y_i)} = \frac{w_i P_1(y_i)}{P(y_i)}$$

$$g_{i2} = P(i \geq z \,|\, y_i) = \frac{P(i \geq z)P(y_i \,|\, i \geq z)}{P(y_i)} = \frac{(1 - w_i)P_2(y_i)}{P(y_i)}$$

M-шаг:

Запишем задачу выпуклого программирования

$$\begin{cases} -\sum_{i=1}^{n} \ln p(y_i) \to \min_{w,\theta} \\ 1 = w_1 > w_2 > \ldots > w_n = 0 \end{cases}$$

Запишем Лагранжиан

$$L(w, \theta) = -\sum_{i=1}^{n} \ln p(y_i) + \sum_{i=1}^{n-1} \lambda_i (w_{i+1} - w_i);$$

Необходимые условия седовой точки функции Лагранжа:

$$1) \frac{dL}{dw_i} = -\frac{d}{dw_i} \ln p(y_i) - \lambda_i + \lambda_{i-1} = 0 \quad \text{при } i = 2, \ldots, n-1$$

$$2)\frac{dL}{d\theta_1} = -\sum_{i=1}^{n}\frac{d}{d\theta_1}\ln p(y_i) \quad \text{и} \quad 3)\frac{dL}{d\theta_2} = -\sum_{i=1}^{n}\frac{d}{d\theta_2}\ln p(y_i)$$

<u>1)Ищем вектор весов w</u>

$$\frac{d}{dw_i}\ln p(y_i) = \frac{p_1(y_i) - p_2(y_i)}{p(y_i)} = \frac{g_{i1}}{w_i} - \frac{g_{i2}}{1 - w_i}$$

Из-за условий дополняющей нежесткости, следует, что все $\lambda_i = 0$, так как $w_i \neq w_{i+1}$ при $i = 1,\dots,n-1$.

Значит

$$0 = \frac{dL}{dw_i} = -\frac{d}{dw_i}\ln p(y_i) = -\frac{g_{i1}}{w_i} + \frac{g_{i2}}{1 - w_i} \implies w_i = g_{i1} \text{ при } 1 < i < n$$

<u>2)Ищем параметр θ_1</u>

$$\frac{d}{d\theta_1}\ln p(y_i) = \frac{w_i\frac{d}{d\theta_1}p_1(y_i)}{p(y_i)} = \frac{w_i([y_i = 1] - [y_i = 0])}{p(y_i)} =$$

$$= \frac{w_i([y_i = 1] - [y_i = 0])}{p_1(y_i)} \cdot \frac{p_1(y_i)}{p(y_i)} = \frac{[y_i = 1] - [y_i = 0]}{p_1(y_i)} \cdot g_{i1} =$$

$$= g_{i1} \cdot \frac{[y_i = 1] - [y_i = 0]}{([y_i = 1] - [y_i = 0])\theta_1 + [y_i = 0]} =$$

$$= g_{i1} \cdot \frac{1}{\theta_1 + \frac{[y_i = 0]}{[y_i = 1] - [y_i = 0]}} = g_{i1} \cdot \frac{1}{\theta_1 - [y_i = 0]} =$$

$$= g_{i1} \cdot \left(\frac{1}{\theta_1} \cdot [y_i = 1] - \frac{1}{1 - \theta_1} \cdot [y_i = 0]\right)$$

$$\sum_{i=1}^{n}\frac{d}{d\theta_1}\ln p(y_i) = \sum_{i=1}^{n} g_{i1} \cdot \frac{1}{\theta_1 - [y_i = 0]} =$$

$$= \frac{1}{\theta_1}\sum_{i=1}^{n} g_{i1}[y_i = 1] - \frac{1}{1 - \theta_1}\sum_{i=1}^{n} g_{i1}[y_i = 0] =$$

$$= \frac{1}{\theta_1}\sum_{i=1}^{n} g_{i1}[y_i = 1] - \frac{1}{1 - \theta_1}\sum_{i=1}^{n} g_{i1}(1 - [y_i = 1]) =$$

$$= \frac{1}{\theta_1(1-\theta_1)} \sum_{i=1}^{n} g_{i1}[y_i = 1] - \frac{1}{1-\theta_1} \sum_{i=1}^{n} g_{i1}$$

$$0 = \frac{dL}{d\theta_1} = -\sum_{i=1}^{n} \frac{d}{d\theta_1} \ln p(y_i) =$$

$$= -\frac{1}{\theta_1(1-\theta_1)} \sum_{i=1}^{n} g_{i1}[y_i = 1] + \frac{1}{1-\theta_1} \sum_{i=1}^{n} g_{i1}$$

$$\theta_1 = \frac{\sum_{i=1}^{n} g_{i1}[y_i = 1]}{\sum_{i=1}^{n} g_{i1}}$$

3) Аналогично ищем параметр θ_2

$$\theta_2 = \frac{\sum_{i=1}^{n} g_{i2}[y_i = 1]}{\sum_{i=1}^{n} g_{i2}}$$

Ответ:

Е-шаг :

$$g_{i1} = P(i < z \,|\, y_i) = \frac{P(i < z)P(y_i \,|\, i < z)}{P(y_i)} = \frac{w_i P_1(y_i)}{P(y_i)}$$

$$g_{i2} = P(i \geq z \,|\, y_i) = \frac{P(i \geq z)P(y_i \,|\, i \geq z)}{P(y_i)} = \frac{(1-w_i)P_2(y_i)}{P(y_i)}$$

М-шаг:

$w_i = g_{i1}$ при $1 < i < n$; $w_1 = 1$; $w_n = 0$;

$$\theta_1 = \frac{\sum_{i=1}^{n} g_{i1}[y_i = 1]}{\sum_{i=1}^{n} g_{i1}};$$

$$\theta_2 = \frac{\sum_{i=1}^{n} g_{i2}[y_i = 1]}{\sum_{i=1}^{n} g_{i2}}$$

Задача 14

$$\sum_{k=1}^{K} \sum_{i=1}^{l} [a(x_i) = k] \|x_i - c_k\|^2 \to \min_{c_k, w}$$

1) зафиксируем c_k, нужно минимизировать по w.

$$\forall i \ \sum_{k=1}^{K} [a(x_i) = k] \|x_i - c_k\|^2 \to \min_{w} \implies a(x_i) = \arg \min_{k=1,\ldots,K} \|x_i - c_k\|^2$$

2) зафиксируем w, нужно минимизировать по c_k.

$$\forall k \ f(c_k) = \sum_{i=1}^{l} [a(x_i) = k] \|x_i - c_k\|^2 \to \min_{c_k}$$

$$\frac{df(c_k)}{dc_k} = \frac{d}{dc_k} \sum_{i=1}^{l} [a(x_i) = k] \|x_i - c_k\|^2 = 0$$

$$\sum_{i=1}^{l} [a(x_i) = k](x_i - c_k) = 0 \implies c_k = \frac{\sum_{i=1}^{l} [a(x_i) = k] x_i}{\sum_{i=1}^{l} [a(x_i) = k]}$$ - точка экстремума

$$\frac{d^2 f(c_k)}{dc_k^2} = \sum_{i=1}^{l} [a(x_i) = k] > 0 \implies c_k$$ - точка минимума