Машинное обучение 2. Домашнее задание 1. Васильева Наталья

Задача 2.

$$\begin{cases} x^2+1 \to \min \\ (x-2)(x-4) \le 0 \end{cases}$$

$x \in \mathbb{R}$

1) Допустимое мн-во значений:

$(x-2)(x-4)=0$

$+ \quad - \quad +$

$2 \quad 4 \quad x$

$\boxed{x \in [+2;4]}$

1) Лагранжиан: $\mathcal{L}(x,\lambda) = x^2+1+\lambda[(x-2)(x-4)] = x^2+1+\lambda(x^2-6x+8)$

Условия Каруша-Куна-Таккера:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 2x+2\lambda x-6\lambda = x+\lambda x-3\lambda=0 & (1) \\ x\in[2;4] & (2) \\ \lambda \ge 0 & (3) \\ \lambda(x-2)(x-4)=0 & (4) \end{cases}$$

(1) $x(1+\lambda)=3\lambda$

$x=\dfrac{3\lambda}{1+\lambda}$

(4) $\lambda\left(\dfrac{3\lambda}{1+\lambda}-2\right)\left(\dfrac{3\lambda}{1+\lambda}-4\right)=0$

$\lambda=0$

$\dfrac{3\lambda}{1+\lambda}=2 \Rightarrow \lambda=2$

$\dfrac{3\lambda}{1+\lambda}=4 \Rightarrow \lambda=-4$ → не удовлетвор. усл-ю $\lambda\ge 0$.

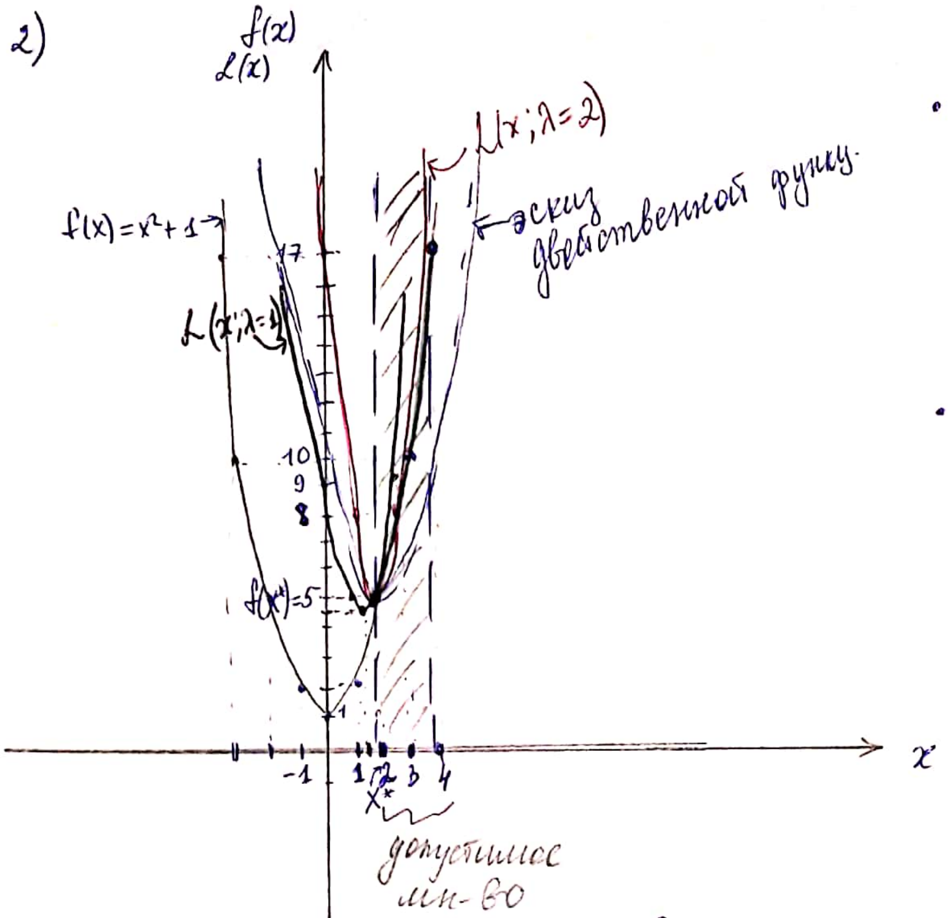
$\rightarrow \left[\begin{array}{l} \lambda=0 \Rightarrow x=0 \text{ → не удовлетвор. усл-ю } x\in[2;4] \\ \lambda=2 \Rightarrow x=\dfrac{3\cdot 2}{3}=2. \end{array}\right.$

$\lambda=2,\ x=2 \Rightarrow \lambda(x-2)(x-4)=0$ → выполн. все усл-я.

⇒ Реш-е: $\boxed{\lambda=2;\ x^*=2}$

$\boxed{f(x^*)=2^2+1=5}$

2)



Из графика видно, что минимум $f(x)$ на допустимом мн-ве достигается в его левой границе $x^*=2$.

- $f(x)=x^2+1$

верш. параб. $x_B=-\frac{b}{2a}=0$
$y_B=1$

Точки:

x	1	2	3	4
$f(x)$	2	5	10	17

- график лагранжиана.

$\lambda=1 \Rightarrow$

$\Rightarrow \mathcal{L}(x,\lambda=1)=2x^2-6x+9$

$\left.\begin{array}{l} x_B=1.5 \\ y_B=4.5 \end{array}\right\}$ вершина

точки:

x	2	3
$\mathcal{L}(x)$	5	9

$\lambda=2 \Rightarrow$

$\Rightarrow \mathcal{L}(x,\lambda=2)=3x^2-12x+17$

вершина: $x_B=\frac{12}{6}=2$

$y_B=5$

Точки:

x	3	4
$\mathcal{L}(x)$	8	17

Для $\lambda=3$ верш.: $x_B=2\frac{1}{4}$
$y_B=4.75$

$f(x_*)=5 \geq \inf\limits_x \mathcal{L}(x,\lambda)$

$\downarrow$

Для перебранных λ (1, 2, 3)

$\inf \mathcal{L}(x,\lambda) \leq 5$ $\quad$ ($\lambda=1 \Rightarrow \inf\mathcal{L}(\cdot)=4.5$;
$\lambda=2 \Rightarrow \inf\mathcal{L}(\cdot)=5$;
$\lambda=3 \Rightarrow \inf\mathcal{L}(\cdot)=4.75$)

$\Rightarrow$ Нер-во выполнено

$f(x_*)=5 \geq \inf\limits_x \mathcal{L}(x,\lambda)$.

3) <u>Двойственная задача.</u>

Двойственная ф-я: $g(\lambda)=\inf\limits_x \mathcal{L}(x,\lambda) \leq f_0(x_*)$

дает нижнюю оценку на min в исходн. задаче.
(наиб. из всех нижних граней)

Двойств. задача к исх. задаче $\begin{cases} g(\lambda) \to \max\limits_\lambda \\ \lambda \geq 0 \end{cases}$

Минимум лагранжиана достигается в т. $x = \frac{3\lambda}{1+\lambda}$, подставим его в $L(x,\lambda)$ и получим двойств. ф-ю:

$$g(\lambda) = L\left(\frac{3\lambda}{1+\lambda}; \lambda\right) = \left(\frac{3\lambda}{1+\lambda}\right)^2 + 1 + \lambda\left[\left(\frac{3\lambda}{1+\lambda}\right)^2 - 6\left(\frac{3\lambda}{1+\lambda}\right) + 8\right]$$

$$\begin{cases} g(\lambda) = \left(\frac{3\lambda}{1+\lambda}\right)^2 + 1 + \lambda\left[\left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right)\right] \to \max\limits_{\lambda} \\ \lambda \geq 0 \\ \left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right) \leq 0 \end{cases}$$

$$\left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right) = 0$$

$\Rightarrow \lambda = 2$ $\quad \lambda = -4$

не удовл. усл-ю $\lambda \geq 0$

$\left(\frac{3\lambda}{1+\lambda}\right)$ + − 2 λ

$\Rightarrow \boxed{\lambda \geq 2}$

$$\begin{cases} g(\lambda) = \left(\frac{3\lambda}{1+\lambda}\right)^2 + 1 + \lambda\left[\left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right)\right] \to \max\limits_{\lambda} \\ \lambda \geq 2 \end{cases}$$

$$\frac{\partial g(\lambda)}{\partial \lambda} = 2\left(\frac{3\lambda}{1+\lambda}\right)\left(\frac{3(1+\lambda) - 3\lambda \cdot 1}{(1+\lambda)^2}\right) + \lambda \cdot \left[\left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right)\right]' + \left[\left(\frac{3\lambda}{1+\lambda} - 2\right)\left(\frac{3\lambda}{1+\lambda} - 4\right)\right] = \ldots =$$

$$= -\frac{(\lambda-2)(\lambda+4)}{(1+\lambda)^2} = 0$$

$(1+\lambda)^2 \neq 0$: $\lambda = -1$, $\lambda \geq 2$

$$\left[\begin{array}{ll} \lambda - 2 = 0 & \Rightarrow \boxed{\lambda^* = 2} \\ \lambda + 4 = 0 & \Rightarrow \lambda = -4 \to \text{не удовл. } \lambda \geq 0 \end{array}\right.$$

$\boxed{\lambda^* = 2}$

$\left.\frac{3\lambda}{1+\lambda}\right|_{\lambda=2} = \frac{6}{3} = 2. \Rightarrow g(\lambda^*) = 4 + 1 + 0 = 5.$

$\boxed{g(\lambda^*) = 5}$

$\boxed{f(x^*) = 5}$

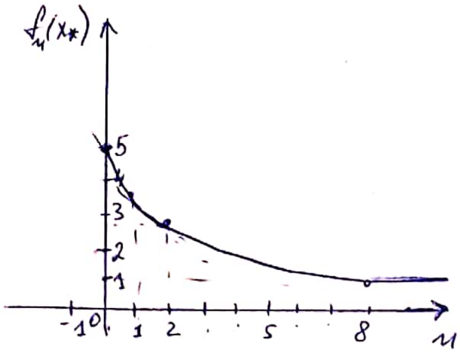
$\Downarrow$

$\boxed{g(\lambda^*) = f(x^*)}$ – имеем сильную двойственность.

(строгая двойств. выполнена)

1) $x^2+1 \to \min$

$(x-2)(x-4) \le u$ $\to$ $\underline{x^2-6x+8-u=0}$ $\to$ с $\uparrow u$, границы допустим. мн-ва двигаются в левую и правую сторону $\Rightarrow f(x^*)\downarrow$ (видно из графика)



$u \ge 0$

$u<0 \Rightarrow$ реш-й у $x^2-6x+8-u=0$ нет или единств.

$u=-1 \Rightarrow x^*=3 \Rightarrow$

$u=1 \Rightarrow x\in[1.58;4.41] \to \min$ в т. $x^*=1.58$ $\Rightarrow f(x^*)=3.5$

$u=2 \Rightarrow x\in[1.27;4.73] \to x^*=1.27 \Rightarrow f(x^*)=2.6$

$u=10 \Rightarrow x\in[-0.32;6.32] \Rightarrow$ ~~min~~

$x^*=0 \Rightarrow f(x^*)=1$

Когда левая граница допустимого мн-ва x достигает 0, (и далее уменьшается), то минимум достигается в вершине параболы: $\boxed{x^*=0,\ f(x^*)=1}$

(правая граница допустимого мн-ва не уменьшается при этом)

При $u=8$:

$\Rightarrow x^2-6x=0$

$x(x-6)=0$ $\quad$ + 0 − 6 + x

$\boxed{x\in[0;6]}$

$\hookrightarrow$ в $x=0 \to \min$

При $u=8$ достигается глобальн. мин. x^2+1 и далее график $f_u(x_*)$ выходит в насыщение (везде $=1$)

Задача 3.

$K(x,z) = \cos(x-z) = \cos x\cdot\cos z + \sin x\cdot\sin z$. (Воспользуемся св-вом косинуса разности)

1) $K(x,z) = x = f(x)\cdot \underbrace{f(z)}_{1}$, где $f(x)=x$ $\to$ по 4п. Теоремы 1 явл. ядром.
ядро $\nearrow$ $\quad$ $\hookrightarrow$ вещ. ф-я от x

или по опред. ядра $K(x,z)=\langle\varphi(x),\varphi(z)\rangle$, где $\varphi(x)=x$, $\varphi(z)=1$

$\varphi(z)=1$ – конст. явл. ядром по св-ву 4 теор. $(\sqrt{1}\cdot\sqrt{1})$

Аналогично $K(z,x)=z$ - ядро.

2) $\cos(x)\cdot\cos(z)$; $\overset{\cos(x)}{\text{вещественн.}}$ ф-я от $x \Rightarrow$ $\overset{\cos(x)\cdot\cos(z)}{\text{по 4 п. теоремы 1}}$ явл. ядром.
где $\boxed{f(x)=\cos(x),\ f(z)=\cos(z)}$ $\quad$ $\cos(z)$-аналогично, вещ. ф-я от z

Аналогично $\sin x\cdot\sin z$ явл. ядром, как произвед. 2 вещест. ф-й (п.4)

3) $\underbrace{\underbrace{\cos x \cdot \cos z}_{\text{п. 4 (ядро)}} + \underbrace{\sin x \cdot \sin z}_{\text{п. 4 (ядро)}}}_{\text{ядро по п. 1}}$ искомая ф-я $\downarrow$ $\to$ ядро как сумма ядер по п. 1 теоремы 1.

Задача 4.

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

Проверим условия теор. Мерсера:

1) $K(x,z) = K(z,x)$ – симметрична $\left(\frac{1}{1+e^{-z \cdot x}} = \frac{1}{1+e^{-xz}}\right)$

2) Неотриц. определенность означает $\forall \ell$ $(x_1, \dots, x_\ell) \in \mathbb{R}^d$ м-ца $(K(x_i, x_j))_{i,j=1}^{\ell}$ неотриц. определена (т.е. все её главные миноры неотрицательны (≥ 0)).

Попробуем подобрать такую выборку, чтобы не выполнялось 2 усл-е.

$\Delta_1 = \frac{1}{1+e^{-xz}} > 0$ (1-й минор всегда >0) $\Rightarrow$ поэтому как минимум следует рассмотреть $\ell = 2$

$\Delta_2 = K_{11} \cdot K_{22} - K_{12} \cdot K_{21}$

Пусть $\ell = 2$ (длина выборки), $d = 1$ (кол-во признаков)

$$X = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow K = (K(x_i, x_j))_{i,j=1}^{2} = \begin{pmatrix} \frac{1}{1+1} & \frac{1}{2} \\ \frac{1}{2} & \underbrace{\frac{1}{1+e^{-2}}}_{0.88} \end{pmatrix} \Rightarrow \det K \approx 0.19 > 0$$

Пусть

$$X = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow K = \begin{pmatrix} \frac{1}{1+e^{-1}} & \frac{1}{1+e^{-2}} \\ \frac{1}{1+e^{-2}} & \frac{1}{1+e^{-4}} \end{pmatrix} \Rightarrow \det K = \frac{e}{1+e} \cdot \frac{e^4}{(1+e^4)} - \frac{e^2}{(1+e^2)^2} \underline{< 0}$$

$\Rightarrow$ Удалось подобрать такую выборку, что матрица $(K(x_i, x_j))_{i,j=1}^{\ell}$ не является неотрицательно определенной. (нарушено 2 усл-е теоремы Мерсера) $\Rightarrow$ $\frac{1}{1+e^{-xz}}$ не явл. ядром.

(*) 1) $K(z,x) = z \cdot x$ – ядро как произвед. веществен. ф-й по п. 4 теоремы 1.

2) $K(x,z) = -x \cdot z$ – домнож. на отриц. конст. $\to$ нельзя построить ядро с помощью домнож. на отриц. конст. $\Rightarrow$ не явл. ядром.

Задача 5.

полиномиал. ядро.

• $K_1(x,z) = (1+xz)^2 = 1 + 2xz + x^2z^2 = \langle\varphi(x), \varphi(z)\rangle$

Объект x представим в спрямляющем пр-ве:

$\varphi(x) = (1, \underset{\sim}{2x}, x^2)$. Все мономы до степени 2. (при этом также при xz есть коэффициент)

набор различных признаков в исх. пр-ве

• $K_2(x,z) = (1 + xz + x^2z^2)$

спрямляющее пр-во: $\varphi(x) = (1, \underset{\sim}{x}, x^2)$

Отличается от 1-го спрямл. пр-ва тем, что нет коэф-та перед xz.

• $k_1 + k_2 = 1 + 2xz + z^2x^2 + 1 + xz + x^2z^2 = 2 + 3xz + 2x^2z^2$.

$\Rightarrow \varphi(x) = (2, 3x, 2x^2)$

Задача 1

Двойств. ф-н для гребневой регрессии:

$$Q(a) = \frac{1}{2}\|Ka - y\|^2 + \frac{\lambda}{2}a^TKa \to \min_a$$

$K = \Phi\Phi^T$. – все попарные скалярн. произв. объектов. $\boxed{K = K^T}$ (симметричн.)

Выведем некотор. св-ва матричн. дифференцирования:

$K \in \mathbb{R}^{n\times n}$, $a \in \mathbb{R}^n$

• а) $\nabla_a a^TKa = ?$

$$\frac{\partial}{\partial a_i} a^T\underbrace{Ka}_{\substack{n\times 1 \\ (\text{вектор})}} = \frac{\partial}{\partial a_i}\sum_j a_j (Ka)_j = \frac{\partial}{\partial a_i}\sum_j a_j\Big(\sum_k K_{jk}a_k\Big) = \frac{\partial}{\partial a_i}\sum_{jk} a_j K_{jk} a_k = \frac{\partial}{\partial a_i}\Big[\sum_{j\neq i} a_j k_{ji} a_i +$$

$$+ \sum_{\substack{k\neq j \\ (j=i)}} a_i k_{ik} a_k + \underset{(i=j=k)}{k_{ii}a_i^2}\Big] = \sum_{j\neq i} a_j k_{ji} + \sum_{k\neq i} a_k \cdot k_{ik} + 2k_{ii}a_i = \sum_j a_j k_{ji} + \sum_k k_{ik}a_k =$$

$(K + K^T)\cdot a$

$$= \underbrace{\sum_j (k_{ji} + k_{ij})a_j}_{\text{произв. по 1 элементу}} \Rightarrow \boxed{\nabla_a a^TKa = (K + K^T)a}$$

в матричном виде.

, т.к. $K = K^T$ симметр. $\Rightarrow$ св-во а) $\boxed{\nabla_a a^TKa = 2Ka}$

• б) $\nabla_x a^Tx$, где $a \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ – векторы

свойство б) $\boxed{\nabla_x a^Tx = a}$

$a^Tx = \langle a, x\rangle = x^Ta = a_1x_1 + a_2x_2 + \ldots + a_nx_n$

$\frac{\partial a^Tx}{\partial x_1} = a_1 \ \ldots \quad \nabla_x a^Tx = (a_1\ a_2\ \ldots\ a_n)^T = a \quad \Leftrightarrow \nabla_x\langle a, x\rangle = \nabla_x\langle x, a\rangle = a$

FOC: $\nabla_a Q(a) = \nabla_a\left[\frac{1}{2}\|Ka-y\|^2 + \frac{\lambda}{2}a^TKa\right] = \nabla_a\left[\frac{1}{2}(Ka-y)^T(Ka-y) + \frac{\lambda}{2}a^TKa\right]$

- $\nabla_a\left((Ka-y)^T(Ka-y)\right) = \nabla_a\left(\underbrace{(Ka)^T}_{= a^TK^T}Ka - (Ka)^Ty - y^TKa + \underbrace{y^Ty}_{}\right) =$

не зависит от a
$\nabla_a y^Ty = 0$

$$= \nabla_a\left(\overset{(1\times n)\,(n\times n)\,(n\times 1)}{\underbrace{a^TK^TKa}} - \underbrace{a^TK^Ty}_{(y^TKa)^T} - \underbrace{y^TKa}_{(K^Ty)^Ta}\right) = \underbrace{(K^TK + K^TK)}_{2K^TK}a - \underbrace{K^Ty - K^Ty}_{-2K^Ty} = 0$$

по св-ву а) $(K^TK + K^TK)a$; по св-ву б) K^Ty ; по св-ву б) $\nabla_a \Rightarrow K^Ty$

- $\nabla_a \frac{\lambda}{2}a^TKa = \lambda\cdot Ka$
по св-ву а)

$\Rightarrow \nabla_a Q(a) = K^TKa - K^Ty + \lambda Ka = 0$

$$a(K^TK + \lambda K) = K^Ty$$

$$a\cancel{K}(\underset{= K}{K^T} + \lambda I) = \cancel{K}^Ty$$

$$a(K + \lambda I) = y$$

$$\boxed{a^* = (K+\lambda I)^{-1}y.}$$