Тестовые вопросы к лекции 1

Дедлайн (жёсткий): 8 февраля 2023 года, 23:59 по Москве

Ответьте на следующие 5 вопросов (пояснения приветствуются, но не обязательны). Каждый вопрос стоит 0.4 балла. Просьба присылать ответы (желательно в формате PDF) на почту **Горбунова Эдуарда**: ed-gorbunov@yandex.ru. Кроме того, просьба **указывать следующую тему письма:**

«Методы оптимизации в ML, весна 2023. Тест 1»

1. Предположим, что в задаче со слайдов 2-9 из первой презентации, на складах хранится 2 типа груза. А именно, пусть на складе i имеется $a_{i,1}$ первого типа груза и $a_{i,2}$ второго типа груза. Кроме того, пусть заводу j требуется теперь $b_{i,1}$ первого типа груза и $b_{i,2}$ второго типа груза. Стоимость перевозки единицы первого типа груза между складом i и заводом j пропорциональна расстоянию между ними c_{ij}, а стоимость перевозки второго типа груза – пропорциональна квадрату расстояния между ними c_{ij}^2. Пусть x_{ij} – это количество первого типа груза, перевезённого со склада i на завод j, а y_{ij} – количество второго типа груза, перевезённого со склада i на завод j.

   Какая из следующих задач является задачей минимизации суммарной перевозки грузов со складов на заводы (чтобы обеспечить заводы необходимым количеством обоих грузов).

   (a)
$$\min_{x,y} \left\{ f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}(x_{ij}+y_{ij}) \right\}$$
$$\sum_{j=1}^{m} x_{ij} \le a_{i,1}, \quad \sum_{j=1}^{m} y_{ij} \le a_{i,2}, \quad \forall i = 1,\ldots,n$$
$$\sum_{i=1}^{n} x_{ij} \ge b_{j,1}, \quad \sum_{i=1}^{n} y_{ij} \ge b_{j,2}, \quad \forall j = 1,\ldots,m$$

   (b)
$$\min_{x,y} \left\{ f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}^2 y_{ij} \right\}$$
$$\sum_{j=1}^{m} x_{ij} \le a_{i,1}, \quad \sum_{j=1}^{m} y_{ij} \le a_{i,2}, \quad \forall i = 1,\ldots,n$$
$$\sum_{i=1}^{n} x_{ij} \ge b_{j,1}, \quad \sum_{i=1}^{n} y_{ij} \ge b_{j,2}, \quad \forall j = 1,\ldots,m$$

(c)

$$\min_{x,y} \quad \left\{ f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}^2 y_{ij}^2 \right\}$$
$$\sum_{j=1}^{m} x_{ij} \le a_{i,1}, \quad \sum_{j=1}^{m} y_{ij} \le a_{i,2}, \quad \forall i = 1,\dots,n$$
$$\sum_{i=1}^{n} x_{ij} \ge b_{j,1}, \quad \sum_{i=1}^{n} y_{ij} \ge b_{j,2}, \quad \forall j = 1,\dots,m$$

(d)

$$\min_{x}\max_{y} \quad \left\{ f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}^2 y_{ij} \right\}$$
$$\sum_{j=1}^{m} x_{ij} \le a_{i,1}, \quad \sum_{j=1}^{m} y_{ij} \le a_{i,2}, \quad \forall i = 1,\dots,n$$
$$\sum_{i=1}^{n} x_{ij} \ge b_{j,1}, \quad \sum_{i=1}^{n} y_{ij} \ge b_{j,2}, \quad \forall j = 1,\dots,m$$

2. Рассмотрим задачу со слайдов 18-29 (классификация рукописных цифр) из первой презентации. В этом примере мы рассмотрели лосс-функции

$$f_i(w) = (z_{3,0} - 0)^2 + \ldots + (z_{3,j-1} - 0)^2 + (z_{3,j} - 1)^2 + (z_{3,j+1} - 0)^2 + \ldots + (z_{3,9} - 0)^2,$$

где j – это цифра, написанная на i-й картинке. Ключевое свойство этой функции в том, что она тем больше, чем меньше ответ нейросети соответствует правильному ответу. Какая из функций ниже **не соответствует** этому свойству?

(a) $f_i(w) = |z_{3,0} - 0|^3 + \ldots + |z_{3,j-1} - 0|^3 + |z_{3,j} - 1|^3 + |z_{3,j+1} - 0|^3 + \ldots + |z_{3,9} - 0|^3$

(b) $f_i(w) = (z_{3,0} - 0) + \ldots + (z_{3,j-1} - 0) + (z_{3,j} - 1) + (z_{3,j+1} - 0) + \ldots + (z_{3,9} - 0)$

(c) $f_i(w) = \max\left\{|z_{3,0} - 0|, \ldots, |z_{3,j-1} - 0|, |z_{3,j} - 1|, |z_{3,j+1} - 0|, \ldots, |z_{3,9} - 0|\right\}$

(d) $f_i(w) = \sqrt{(z_{3,0} - 0)^2 + \ldots + (z_{3,j-1} - 0)^2 + (z_{3,j} - 1)^2 + (z_{3,j+1} - 0)^2 + \ldots + (z_{3,9} - 0)^2}$

3. На слайде 4 из второй презентации приведена оценка: если функции $f(x,\xi)$ выпуклы и M-Липшицевы, Q имеет диаметр D и $\hat{x}^* = \operatorname{argmin}_{x\in Q} \hat{f}(x)$, тогда с вероятностью хотя бы $1-\delta$

$$f(\hat{x}^*) - \min_{x\in Q} f(x) = O\left(\sqrt{\frac{M^2D^2n\ln(m)\ln(n/\delta)}{m}}\right). \tag{1}$$

Пусть $M = 1$, $D = 10$, $n = 1000$. Каким нужно выбрать m, чтобы выражение, стоящее внутри $O(\cdot)$ было меньше $\varepsilon = 10^{-3}$, а равенство равенство (1) выполнялось с вероятностью хотя бы 0.999?

(a) Достаточно взять $m = 10^8$

(b) Достаточно взять $m = 10^{10}$
(c) Достаточно взять $m = 10^{12}$
(d) Достаточно взять $m = 10^{14}$

4. Рассмотрим задачу со слайдов 6-8 из второй презентации. Какое предположение нужно сделать о шуме ξ_i, чтобы следуяя аналогичному подходу (оценка максимального правдоподобия), получить задачу

$$\hat{x} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{m} |y_i - \langle a^i, x \rangle|^{3/2}?$$

(a) Плотность распределения ξ_i должна иметь вид: $\rho(x) \sim \exp\left(-\alpha|x|^3\right)$
(b) Плотность распределения ξ_i должна иметь вид: $\rho(x) \sim \exp\left(-\alpha|x|^{2/3}\right)$
(c) Плотность распределения ξ_i должна иметь вид: $\rho(x) \sim \exp\left(-\alpha|x|^{3/2}\right)$
(d) Плотность распределения ξ_i должна иметь вид: $\rho(x) \sim \exp\left(-\alpha|x|^{-2/3}\right)$

5. Функция $f(x)$ заданная на множестве $\mathbb{R}^n$ и принимающая значения в $\mathbb{R}$ является M-Липшицевой относительно ℓ_2-нормы, если

(a) для всех $x, y \in \mathbb{R}^n$ выполняется $|f(x) - f(y)| \leq M\|x - y\|_2^2$
(b) для всех $x, y \in \mathbb{R}^n$ выполняется $|f(x) - f(y)| \leq M\|x - y\|_\infty^2$
(c) для всех $x, y \in \mathbb{R}^n$ выполняется $|f(x) - f(y)| \leq M\|x - y\|_2$
(d) для всех $x, y \in \mathbb{R}^n$ выполняется $|f(x) - f(y)| \leq M\|x - y\|_1^2$