1. Дана функция $f : \mathbb{R}^3 \to \mathbb{R}$ с минимальным значением в $x_* = (2, 0, 1)^\top$. На ней запускается метод оптимизации из точки x_0, про который известно, что

$$\|x_{k+1} - x_*\|_2 \leq \frac{1}{2}\|x_k - x_*\|_2^2$$

где $\{x_k\}_{k=0}^{\infty}$ - последовательность итераций метода. При каком значении x_0 можно гарантировать сходимость итераций метода к x_*?

(a) $x_0 = (-1, 1, 2)^\top$
(b) $x_0 = (4, 0, 0)^\top$
(c) $x_0 = (1, 2, 2)^\top$
(d) $x_0 = (3, 1, 1)^\top$

$$\|x_k - x_*\|_2 \leq \frac{1}{2}\|x_{k-1} - x_*\|_2^2 \leq \frac{1}{2 \cdot 2^2}\|x_{k-2} - x_*\|_2^4 \leq$$

$$\leq \frac{1}{2 \cdot 2^2 \cdot 2^4}\|x_{k-3} - x_*\|_2^8 \leq ... \leq \frac{1}{2^{2^k - 1}}\|x_0 - x_*\|_2^{2^k}$$

Для сходимости необходимо $\lim\limits_{k\to\infty} \frac{1}{2^{2^k-1}}\|x_0 - x_*\|_2^{2^k} = 0$

$$\lim_{k\to\infty} \frac{1}{2^{2^k-1}}\|x_0 - x_*\|_2^{2^k} = 0 \Rightarrow \lim_{k\to\infty} \left(\frac{\|x_0 - x_*\|_2}{2}\right)^{2^k} = 0 \Rightarrow \frac{\|x_0 - x_*\|_2}{2} < 1$$

Последнему условию соотвествует только вектор $(3, 1, 1)^\top$

Ответ: d

2. Рассмотрим квадратичную задачу оптимизации

$$f(x) = \frac{1}{2}x^\top A x$$

где матрица A является симметричной положительно определенной. Для решения этой задачи применим квазиньютоновский метод Бройдена (SR1). Обозначим $s_k = x_{k+1} - x_k,\ y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. Схема обновления матрицы H_k, которая аппроксимирует гессиан, имеет вид

$$H_{k+1} = H_k + \Delta H_k,\ \Delta H_k = \begin{cases} \frac{(s_k - H_k y_k)(s_k - H_k y_k)^\top}{\langle s_k - H_k y_k, y_k\rangle}, & s_k - H_k y_k \neq 0 \\ 0, & s_k - H_k y_k = 0 \end{cases}$$

При каком начальном приближении H_0 метод Бройдена совпадает с методом Ньютона?

(a) $H_0 = I$
(b) $H_0 = A$

(c) $H_0 = A^{-1}$
(d) $H_0 = (A^\top A)^{-1}$

$\nabla^2 f(x) = A, \quad [\nabla^2 f(x)]^{-1} = A^{-1}$. Т.к. метод Ньютона на квадратичной функции сходится за одну итерацию, то в качестве H_0 можно взять A^{-1}, в таком случае и метод Бройдена сойдется за одну итерацию

Ответ: c

3. Рассмотрим задачу оптимизации с ограничениями

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$s.t. \; g(x) \leq 0, \; h(x) = 0$$

где $g(x) : \mathbb{R}^n \to \mathbb{R}, \; h(x) : \mathbb{R}^n \to \mathbb{R}$. Какой вид имеет двойственная задача?

(a) $\max\limits_{\lambda, \mu \geq 0} \varphi(\lambda, \mu), \; \varphi(\lambda, \mu) = \min\limits_x [f(x) + \lambda h(x) + \mu g(x)]$.
(b) $\max\limits_{\lambda, \mu \geq 0} \varphi(\lambda, \mu), \; \varphi(\lambda, \mu) = \max\limits_x [f(x) + \lambda h(x) + \mu g(x)]$.
(c) $\min\limits_{\lambda, \mu \geq 0} \varphi(\lambda, \mu), \; \varphi(\lambda, \mu) = \min\limits_x [f(x) + \lambda h(x) + \mu g(x)]$.
(d) $\min\limits_{\lambda, \mu \geq 0} \varphi(\lambda, \mu), \; \varphi(\lambda, \mu) = \max\limits_x [f(x) + \lambda h(x) + \mu g(x)]$.

Ответ: a (из определения двойственной задачи и функции Лагранжа)

4. При построении модели для прогнозирования оттока сотрудников используются следующие 3 признака: возраст, время работы в компании и изменение заработной платы в течении последнего года. Требуется предсказать, уйдет ли сотрудник из компании в течение следующих трех месяцев. Был обучен линейный классификатор SVM (для простоты будем считать, что до точности $\varepsilon = 0$) и получен вектор параметров x^*. Оказалось, что у всех сотрудников, соотвествующих опорным объектам, заработная плата в течение последнего года не менялась. Какое значение вектора весов x^* в SVM заведомо невозможно?

(a) $x^* = (0.5, 3.3, 0)^\top$.
(b) $x^* = (-1.7, 5.2, 0)^\top$.
(c) $x^* = (2.8, -0.5, 0.4)^\top$.
(d) $x^* = (150.5, 103.0, 0)^\top$.

$x^* = \sum\limits_{i=1}^{m} \lambda y_i a_i$, где y_i - опорные вектора. Т.к. у опорных векторов из условия третья компонента равна нулю, то и у вектора весов третья компонента должна быть равна нулю. Этому условию не соотвествует вектор $(2.8, -0.5, 0.4)^\top$

Ответ: c

5. Обучение SVM может быть записано в виде минимизации Hinge Loss с l2-регуляризацией.

$$\min_x \sum_{i=1}^{m} \max\{1 - M_i(x), 0\} + \frac{\|x\|_2^2}{2C}$$

Выберите верное утверждение.

(a) Чем больше C, тем сильнее регуляризация и больше ширина разделяющей полосы.

(b) Чем больше C, тем сильнее регуляризация и меньше ширина разделяющей полосы.

(c) Чем больше C, тем слабее регуляризация и больше ширина разделяющей полосы.

(d) Чем больше C, тем слабее регуляризация и меньше ширина разделяющей полосы.

Ответ: d (обсуждалось на лекции)

Ответы:

1. d
2. c
3. a
4. c
5. d