1. В каком случае субградиент функции $f(x)$ в точке x является непустым множеством?

(a) $f(x) = -\sqrt{1-x^2},\ x = 0$
(b) $f(x) = -\sqrt{1-x^2},\ x = 1$
(c) $f(x) = \max(x, 2-x),\ x = 1$
(d) $f(x) = -\sqrt{1-x^2},\ x = \frac{1}{4}$

Скорее всего в вопросе имелся ввиду субдифференциал, ведь субградиент - это не множество, а число, а субдифференциал - это множество всех субградиентов.

Будем использовать определение субградиента a: $f(x) - f(x_0) \geq \langle a, x - x_0 \rangle \quad \forall x \in X$
(a) $-\sqrt{1-x^2} + 1 \geq ax, \quad a = 0$
(b) $-\sqrt{1-x^2} \geq a(x-1)$

$$\sqrt{(1-x)(1+x)} \leq a(1-x)$$

$$\sqrt{1+x} \leq a\sqrt{1-x}$$

$$a \geq \sqrt{\frac{1+x}{1-x}}, \quad x \to 1 \Rightarrow a \geq \infty, \quad \partial f(1) = \varnothing$$

(c) $\partial f(1) = [-1, 1]$
(d) $\partial f(1/4) = \nabla f(1/4) = \frac{\sqrt{15}}{15}$
Комментарий: видимо также в вопросе имелось ввиду "в каком случае субградиент является **пустым** множеством"
Ответ: b

2. Выберите неверное утверждение
(a) $f(x) = |x-2|,\ \partial f(2) = [-1, 1]$
(b) $f(x) = \max(4x-3, 5-x),\ \partial f(2) = [-1, 4]$
(c) $f(x) = \sqrt{x},\ \partial f(3) = \left\{\frac{1}{2\sqrt{3}}\right\}$
(d) $f(x) = \begin{cases} \frac{3-x}{8},\ x < 3 \\ (x-3)^2,\ x \geq 3 \end{cases},\ \partial f(3) = [-1/8, 0]$

$f(x) = |x-2| = \max(-x+2, x-2),\ \partial f(x) = \begin{cases} -1, & x < 2 \\ [-1, 1] & x = 2 \\ 1, & x > 2 \end{cases},\ a$ - верное утверждение

$f(x) = \max(4x-3, 5-x),\ \partial f(x) = \begin{cases} -1, & x < 1.6 \\ [-1, 4] & x = 1.6 \\ 4, & x > 1.6 \end{cases},\ b$ - неверное утверждение

$f(x) = \sqrt{x},\ \partial f(3) = \{\nabla f(3)\} = \left\{\frac{1}{2\sqrt{3}}\right\},\ c$ - верное утверждение

$f(x) = \begin{cases} \frac{3-x}{8},\ x < 3 \\ (x-3)^2,\ x \geq 3 \end{cases},\ \partial f(x) = \begin{cases} -1/8, & x < 3 \\ [-1/8, 0], & x = 3 \\ 2(x-3), & x > 3 \end{cases},\ d$ - верное утверждение

Ответ: b

3. Как обсуждалось на лекции, слишком маленький размер шага градиентного спуска может приводить к медленной сходимости, а слишком большой - к расходимости. Рассмотрим функцию $f(x) = 4x^2 - 20x + 7$ и запустим на ней градиентный спуск из точки $x^0 = 4$ с размером шага α. Существует некоторый "критический"размер шага γ, такой что при $\alpha > \gamma$ градиентный метод расходится, т.е. с каждой итерацией удаляется от решения. Какого значение γ?

(a) $\gamma = 0.5$
(b) $\gamma = 0.25$
(c) $\gamma = 2$
(d) $\gamma = \frac{5}{6}$

$$\nabla f(x) = 8x - 20$$

$$x_{k+1} = x_k - \alpha(8x_k - 20) = (1 - 8\alpha)x_k + 20\alpha$$

Для сходимости необходимо, чтобы отображение было сжимающим, то есть $|1 - 8\alpha| < 1$, откуда $\alpha \in (0, 0.25)$. При $\alpha = 0.25$ будут прыжки между точками 1 и 4. При $\alpha > 0.25$ будет происходить расхождение.

Ответ: b

4. Пусть имеются два датасета для линейной регрессии: (A_1, b_1) и (A_2, b_2). В первом $m_1 = 10^7$ объектов и $n = 1000$ признаков, т.е. $A_1 \in \mathbb{R}^{m_1 \times n},\ b_1 \in \mathbb{R}^{m_1}$. Во втором датасете $m_2 = 10^5$ объектов и тоже $n = 1000$ признаков, т.е. $A_2 \in \mathbb{R}^{m_2 \times n},\ b_2 \in \mathbb{R}^{m_2}$. При этом известно, что

$$\lambda_{max}(A_1^\top A_1) = 50,\ \lambda_{min}(A_1^\top A_1) = 0.1$$

$$\lambda_{max}(A_2^\top A_2) = 300,\ \lambda_{min}(A_2^\top A_2) = 0.001$$

Будем считать, что время вычисления градиента пропорционально количеству объектов в обучающей выборке. Мы запускаем градиентный спуск на каждой из задач линейной регрессии:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2m_1}\|A_1x - b_1\|_2^2$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2m_2}\|A_2x - b_2\|_2^2$$

Оцените (с точки зрения теории), на какой из двух задач (1), (2) градиентному спуску понадобится меньше времени для достижения целевой точности $\varepsilon = 10^{-6}$

(a) На задаче (1)
(b) На задаче (2)

$$\kappa_1 = \frac{\lambda_{max}(A_1^\top A_1)}{\lambda_{min}(A_1^\top A_1)} = \frac{50}{0.1} = 500, \quad \kappa_2 = \frac{\lambda_{max}(A_2^\top A_2)}{\lambda_{min}(A_2^\top A_2)} = \frac{300}{0.001} = 300000$$

В сильно выпуклом случае количество итераций градиентного спуска есть $O(\kappa \log \frac{1}{\varepsilon})$
По условию задачи подсчет градиента (а следовательно и ассимптотическое время работы

одной итерации) есть $O(m)$
Тогда быстрее спуск сойдется на той задаче, у которой значение κm будет меньше

$$\kappa_1 m_1 = 500 \cdot 10^7 = 5 \cdot 10^9, \quad \kappa_2 m_2 = 300000 \cdot 10^5 = 3 \cdot 10^{10}$$

Ответ: a

5. На функции $f(x) : \mathbb{R}^n \to \mathbb{R}$ запускается градиентный спуск с шагом $\alpha = 1$ из точки $x_0 \in \mathbb{R}^n$. Обозначим $x_* = \arg\min_x f(x)$. В каких из примеров ниже выполняется сходимость по аргументу, т.е. $\|x_k - x_*\| \xrightarrow{k\to\infty} 0$?
(a) $x \in \mathbb{R}^1,\ f(x) = \frac{x^4}{10},\ x_0 = 1$
(b) $x = (x[1], x[2])^\top \in \mathbb{R}^2,\ f(x) = (x[1])^2 + (x[2])^2,\ x_0 = (2,2)^\top$
(c) $x \in \mathbb{R}^1,\ f(x) = \frac{x^2}{4},\ x_0 = 1$

(a) $\nabla f(x) = \frac{2x^3}{5}, \quad x_{k+1} = x_k - \frac{2x_k^3}{5}$
Аналитически доказать, что последовательность сходится не вышло, но численно кажется, что сходимость есть
(b) $f(x) = \|x\|_2^2, \quad \nabla f(x) = 2x, \quad x_{k+1} = x_k - 2x_k = -x_k, \quad x_k = (-1)^k \cdot (2,2)^\top$
Очевидно, что эта последовательность не сходится, следовательно на этой задаче сходимости по аргументу нет
(c) $\nabla f(x) = \frac{x}{2}, \quad x_{k+1} = x_k - \frac{x_k}{2} = \frac{x_k}{2}, \quad x_k = \frac{x_0}{2^k} = \frac{1}{2^k}$
Эта последовательность сходится к нулю, а ноль и является точкой минимума, следовательно сходимость по аргументу есть
Ответ: a, c

Ответы:
1. b
2. b
3. b
4. a
5. a, c