Автор работы	1	2	3	4	5	6	7	8

1. [2 балла] Докажите, что функция
$$f(x) = \log\left(\sum_{i=1}^{n} e^{x_i}\right)$$
выпуклая, используя любой дифференциальный критерий выпуклости.

2. [3 балла] Пусть $X \in \mathbb{R}^{m\times n}$, где $\text{rk}X = n$, $\Omega \in \mathbb{S}^m_{++}$, и $W \in \mathbb{R}^{k\times n}$. Найдите матрицу $G \in \mathbb{R}^{k\times m}$, являющуюся решением следующей задачи оптимизации:
$$f(G) = \text{tr}\left(G\Omega G^\top\right) \to \min_{GX=W}$$

3. [1 балл] Приведите пример μ-сильно выпуклой L-гладкой функции, для которой градиентный спуск сойдётся из любой стартовой точки ровно за одну итерацию. Укажите необходимый для этого шаг метода (он не должен зависеть от точки старта). Ответ обоснуйте.

4. [2 балла] Упростите выражение:
$$\sum_{i=1}^{n} \langle X^{-1} w_i, w_i \rangle, \quad \text{где} \quad X = \sum_{i=1}^{n} w_i w_i^T, w_i \in \mathbb{R}^n, \det X \neq 0.$$

5. [2 балла] Предложите метод решения линейной системы уравнений большой размерности:
$$Ax = b,$$
где матрица A симметрична и положительно определена, с помощью одного из методов оптимизации. Пусть известны собственные значения матрицы $\lambda_{\text{min}}(A), \lambda_{\text{max}}(A)$. Приведите оценку скорости сходимости метода (здесь предполагается, что вы предложите метод из курса, тогда доказывать скорость сходимости не нужно).

6. [3 балла] Рассмотрите задачу оптимизации:
$$\begin{aligned} f(x) &\to \min_{x\in\mathbb{R}^n} \\ \text{s.t. } a &\preceq x \preceq b \end{aligned}, \qquad a \in \mathbb{R}^n, b \in \mathbb{R}^n, \; a \prec b.$$
Выпишите явно нетривиальную итерацию проксимального градиентного метода для неё.

7. [2 балла] Сходимость алгоритма Франк-Вульфа в курсе показана только для гладких функций. В этой задаче мы рассмотрим, является ли гладкость необходимой. Рассмотрим следующую негладкую задачу с $f : \mathbb{R}^2 \to \mathbb{R}$, $f = \max\{x_1, x_2\}$:
$$\min_{x_1^2 + x_2^2 \leq 2} f(x_1, x_2).$$
Предположим, что мы стартуем из точки $(0, 0)$ и запускаем алгоритм Франк-Вульфа (с любым правилом шага). Поскольку функция не является гладкой, мы будем использовать произвольный субградиент вместо градиента. Сходится ли этот алгоритм к оптимуму? Ответ обоснуйте.

8. [5 баллов] Рассмотрим выпуклую гладкую задачу минимизации конечной суммы:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \to \min_{x \in \mathbb{R}^p}$$

Алгоритм SGD выбирает $i \in [n]$ равномерно и устанавливает $\nabla f_i(x_k)$ как стохастический градиент. Иногда можно ускорить SGD, выполняя сэмплирование не равномерно, а по значимости.

а. [1 балл] Рассмотрим произвольное распределение вероятностей $p = (p_1, \ldots, p_n)$ с $p_i > 0$ и $\sum_{i=1}^{n} p_i = 1$. Мы выбираем i согласно распределению p и определяем g_k как:

$$g_k := \frac{1}{p_i n}\nabla f_i(x_k) \tag{IS}$$

Тогда покажите, что g_k является несмещенной оценкой градиента, то есть $\mathbb{E}[g_k | x_k] = \nabla f(x_k)$.

б. [2 балла] Напомним, что стандартный единичный симплекс определяется как

$$\Delta_n := \left\{ y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i = 1, y_i \geq 0 \forall i \right\}.$$

Для некоторых фиксированных констант $c_i \in \mathbb{R}$ для $i \in [n]$. Пусть $y^\star$ будет оптимумом следующей задачи оптимизации:

$$\min_{y \in \Delta_n} \sum_{i=1}^{n} \frac{c_i^2}{y_i} \tag{P}$$

Используя общие условия локального экстремума первого порядка, докажите, что

$$y_i^\star = \frac{|c_i|}{\sum_{j=1}^{n} |c_j|}, \forall i \in [n]$$

является решением поставленной задачи (P).

в. [2 балла] Используя результат из предыдущего пункта, вычислите оптимальную вероятность сэмплирования $p^\star$ для того, чтобы минимизировать дисперсию $\mathbb{E}[\|g_k - \nabla f(x_k)\|^2]$ стохастического градиента g_k, определенного в (IS).