1 Матрично-векторное дифференцирование

Определение 1. Говорят, что функция $f\colon \mathbb{R}^m \to \mathbb{R}^n$ дифференцируема в точке x, если существует такой линейный оператор $L\colon \mathbb{R}^m \to \mathbb{R}^n$, что для любых достаточно малых по норме $dx \in \mathbb{R}^n$ выполнено:

$$f(x + dx) = f(x) + L(dx) + o(\|dx\|)$$

Задача 1.1. Пусть $x \in \mathbb{R}^n$. Найдите дифференциалы следующих функций:

a) $a^T x$, $(a \in \mathbb{R}^n)$

b) $x^T Ax$, $(A \in \mathbb{R}^{n\times n})$

c) $\log \|Ax - b\|^2$, $(A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m)$

d) $a \times x$ $(a, x \in \mathbb{R}^3)$

Задача 1.2. Пусть $X \in \mathbb{R}^{m\times n}$. Найдите дифференциалы следующих функций:

a) $a^T Xb$, $(a \in \mathbb{R}^m$, $b \in \mathbb{R}^n)$

b) $a^T X^T BXa$, $(a \in \mathbb{R}^m$, $B \in \mathbb{R}^{n\times n})$

c) $\|Xa - b\|^2$, $(a \in \mathbb{R}^n$, $b \in \mathbb{R}^m)$

d) $\mathrm{tr}(AX)$

e)* X^{-1}

f)* $\log\det(X)$

2 Многомерная линейная регрессия

Задача 2.1. Пусть $X \in \mathbb{R}^{n\times d}$ и $y \in \mathbb{R}^n$. Рассмотрим стандартную задачу линейной регрессии $f_w(x) = w^T x$:

a) Вспомните, как выглядит функция потерь $MSE(w)$ для линейной регрессии и найдите её градиент.

b) Запишите условия для минимума и найдите оптимальный вектор весов w^*.

c) Какие требования на X должны выполняться, чтобы w^* существовал?

d) Рассмотрим линейную регрессию и модифицированную функцию потерь $MSE_\lambda(w) = MSE(w) + \lambda\|w\|^2$.Найдите оптимальный вектор весов w^*.

3 Градиентный спуск

Определение 2. Функция f называется выпуклой, если для любых двух точек $x, y \in \mathbb{R}^n$ выполняется неравенство $f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle$.

Задача 3.1. Докажите, что градиент – это направление наискорейшего роста.

Определение 3. Функция f называется Липшицевой с константой L, если для любых двух точек $x, y \in \mathbb{R}^n$ выполняется неравенство $\|f(x) - f(y)\| \leq L\|x - y\|$.

Задача 3.2. Пусть f выпукла в области $D \subset \mathbb{R}^n$ и имеет Липшицев градиент. Рассмотрим градиентный спуск с шагом λ. Докажите, что если $\lambda < \frac{1}{L}$, то $f(x_k)$ убывает на каждой итерации.

Подсказка. Воспользуйтесь неравенством формулой Тейлора.