Машинное обучение, ФКН ВШЭ
Матрично-вектороное дифференцирование и градиентный спуск

Задача 1. Пусть $f(X) = \ln\det X$, где $X \in \mathbb{R}^{n\times n}$. Найдите производную $\nabla_X f(X)$.

Задача 2. Пусть $f(x) = x^T \exp(xx^T)x$, где $x \in \mathbb{R}^n$, а $\exp(B)$ — матричная экспонента, $B \in \mathbb{R}^{n\times n}$. Матричной экспонентой обозначают ряд

$$I_n + \frac{B}{1!} + \frac{B^2}{2!} + \frac{B^3}{3!} + \frac{B^4}{4!} + \ldots = \sum_{k=0}^{\infty} \frac{B^k}{k!}.$$

Найдите производную $\nabla_x f(x)$.

Задача 3. Пусть $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, x \in \mathbb{R}^n$. Найдите производную $\nabla_x f(x)$ функции

$$f(x) = \sin\|Ax + b\|_2$$

Задача 4. Рассмотрим симметричную матрицу $A \in \mathbb{R}^{n\times n}$ и ее спектральное разложение $A = Q\Lambda Q^T$. Пусть $\lambda \in \mathbb{R}^n$ - это диагональ матрицы Λ (то есть вектор, составленный из собственных значений A). Найдите производные:

1. $\nabla_A \operatorname{tr}(A)$
2. $\nabla_Q \operatorname{tr}(A)$

Задача 5. Рассмотрим задачу обучения линейной регрессии с функцией ошибки Log-Cosh:

$$Q(w) = \frac{1}{\ell}\sum_{i=1}^{\ell} \log(\cosh(w^T x_i - y_i))$$

Выпишите формулу для градиента $\nabla_w Q(w)$. Запишите ее в матричном виде, используя матрицу объекты-признаки X и вектор целевых переменных y.

Задача 6. В случае Ridge-регрессии минимизируется функция со штрафом:

$$Q(w) = (y - xw)^T(y - xw) + \lambda w^2,$$

где λ — положительный параметр, штрафующий функцию за слишком большие значения w.

1. Найдите дифференциал $\mathrm{d}Q(w)$, выведите формулу для оптимального w.
2. Найдите $\mathrm{d}^2Q(w)$. Убедитесь, что мы оказались в точке минимума.
3. Выпишите шаг градиентного спуска в матричном виде.

Задача 7. Найдите симметричную матрицу X наиболее близкую к матрице A по норме Фробениуса, $\sum_{i,j}(x_{ij} - a_{ij})^2$. Тут мы просто из каждого элемента вычитаем каждый и смотрим на сумму квадратов таких разностей. То есть решите задачку условной матричной минимизации

$$\begin{cases} ||X - A||_2^2 \to \min_A \\ X^T = X \end{cases}$$

Hint: Надо будет выписать Лагранджиан. А ещё пригодится тот факт, что $\sum_{i,j}(x_{ij} - a_{ij})^2 = ||X - A||_2^2 = \mathrm{tr}((X - A)^T(X - A))$.