Семинар 1.

Вспоминаем линейную алгебру. Говорим про векторное дифференцирование.

Линейная зависимость.

Мотивация: линейная зависимость нужна для корректной минимизации функции потерь.

Задание 1.

4. Заданы векторы: $\boldsymbol{a}(1,\ 5,\ 3)$, $\boldsymbol{b}(6,\ -4, -2)$, $\boldsymbol{c}(0,\ -5,\ 7)$ и $\boldsymbol{d}(-20,\ 27,\ -35)$. Требуется подобрать числа α, β и γ так, чтобы векторы $\alpha\boldsymbol{a}$, $\beta\boldsymbol{b}$, $\gamma\boldsymbol{c}$ и $\boldsymbol{d}$ образовывали замкнутую ломаную линию, если «начало» каждого последующего вектора совместить с «концом» предыдущего.

Решение: пишем уравнение, что сумма векторов равна 0.

Проекции.

Мотивация: проекции считаются в методах снижения размерности.

Задание 2.

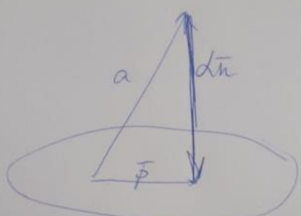
Найдите ортогональную проекцию вектора $\vec{a}\{2\,;3\,;-1\}$ на плоскость, перпендикулярную вектору $\vec{b}\{-1\,;1\,;3\}$.

Решение:

Найти проекцию вектора $\bar{a}$ на пл-ть с нормалью $\bar{n}$.

Решение:

Ищем $\bar{p}$.



$\bar{p} = \bar{a} + \alpha\bar{n}$

Домножим скалярно на $\bar{n}$:

$(\bar{p},\bar{n}) = (\bar{a},\bar{n}) + \alpha(\bar{n},\bar{n})$

0, т.к. $\perp$ $\qquad |\bar{n}|^2$

$0 = (\bar{a},\bar{n}) + \alpha|\bar{n}|^2 \Rightarrow$ находим $\alpha = \alpha_*$, подставляем:

$\boxed{\bar{p} = \bar{a} + \alpha_*\bar{n}}$

Матрицы.

Мотивация: умножение, обратные матрицы – для решения матричных уравнений (=аналитическое решение задач ML)

Задание 3.

Вычислить A*B, а также след и определитель полученной матрицы

$$A = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 5 & 6 \\ -7 & 0 & 1 \end{pmatrix}; B = \begin{pmatrix} 5 & 3 & 2 \\ 0 & 1 & -3 \\ 2 & 15 & 3 \end{pmatrix}$$

Обратные матрицы, собственные числа, собственные значения.

Мотивация: определители, собственные числа и собственные значения (для снижения размерности, а также – как часть матричного дифференцирования, например, для backprop)

Задание 4.

Решить матричное уравнение

$$\begin{pmatrix} 3 & -1 \\ 5 & -2 \end{pmatrix} \cdot X \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 14 & 16 \\ 9 & 10 \end{pmatrix}.$$

Задание 5.

Найти собственные значения и собственные векторы матрицы. В ответе указать квадрат длины собственного вектора x={x1, x2, x3}, с x2=1, соответствующего наибольшему собственному значению.

$$A = \begin{pmatrix} 4 & -3 & -3 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

Векторное дифференцирование.

Мотивация: векторное и матричное дифференцирование (решение всех задач оптимизации в ML)

Теория: что такое производная по вектору и по матрице

- При отображении вектора в число $f(x) : \mathbb{R}^n \to \mathbb{R}$

$$\nabla_x f(x) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]^T.$$

- При отображении матрицы в число $f(A) : \mathbb{R}^{n \times m} \to \mathbb{R}$

$$\nabla_A f(A) = \left(\frac{\partial f}{\partial A_{ij}}\right)_{i,j=1}^{n,m}.$$

Мы хотим оценить, как функция изменяется по каждому из аргументов по отдельности. Поэтому производной функции по вектору будет вектор, по матрице — матрица.

Полезные свойства:

1) $d(XY) = dX \cdot Y + X \cdot dY$
2) Если A - матрица константа, то $dA = 0$
3) $d(X') = dX'$
4) $d \det X = \det X tr(X^{-1}dX)'$

Задание 6.

Задача 1.2. *Пусть $a \in \mathbb{R}^n$ — вектор параметров, а $x \in \mathbb{R}^n$ — вектор переменных. Необходимо найти производную их скалярного произведения по вектору переменных $\nabla_x a^T x$.*

Решение.

$$\frac{\partial}{\partial x_i} a^T x = \frac{\partial}{\partial x_i} \sum_j a_j x_j = a_i,$$

поэтому $\nabla_x a^T x = a$.

Заметим, что $a^T x$ — это число, поэтому $a^T x = x^T a$, следовательно,

$\nabla_x x^T a = a$.

Задание 7.

Задача 1.3. *Пусть теперь $A \in \mathbb{R}^{n \times n}$. Необходимо найти $\nabla_x x^T Ax$.*

Решение.

$$\frac{\partial}{\partial x_i} x^T Ax = \frac{\partial}{\partial x_i} \sum_j x_j (Ax)_j = \frac{\partial}{\partial x_i} \sum_j x_j \left(\sum_k a_{jk} x_k \right) = \frac{\partial}{\partial x_i} \sum_{j,k} a_{jk} x_j x_k =$$

$$= \sum_{j \neq i} a_{ji} x_j + \sum_{k \neq i} a_{ik} x_k + 2a_{ii} x_i = \sum_j a_{ji} x_j + \sum_k a_{ik} x_k = \sum_j (a_{ji} + a_{ij}) x_j.$$

Поэтому $\nabla_x x^T Ax = (A + A^T)x$.