😳 Линейная алгебра и матрично-векторное дифференцирование

Векторы и матрицы

- Вектором далее называем массив (ie набор) чисел:

$$x^T = [x_1, x_2, ...x_n], x \in \mathbb{R}^n$$

- Матрица - множество векторов(как столбцов и строк) в таблице, далее отдельные виды матриц обозначаются как:
  - Симметрическая (если $A^T = A$) $A \in \mathbb{S}^n$
  - Положительно определенная(полуопределенная): $\forall x \neq 0 \Rightarrow x^T A x > 0 (\geq 0)$. Тогда $A \in \mathbb{S}^n_{++} (\mathbb{S}^n_{+})$. Аналогично определяется и отрицательная определенность и множества $\mathbb{S}^n_{--}, \mathbb{S}^n_{-}$.

Матричное умножение (matmul)

В стандартном виде умножение матриц имеет сложность $O(n^3)$, однако существует алгоритм Штрассена, позволяющий выполнять умножение с $O(n^{\log_2 7})$.

Матричная экспонента $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k, e^{A+B} \neq e^A e^B$ (следует из некоммутативности умножения матриц)

Скалярное произведение (dot product)

Обозначение: $\langle x, y \rangle = x^T y = \sum x_i y_i = y^T x = \langle y, x \rangle$

Свойства:

- $\langle x, y \rangle = \langle y, x \rangle$
- $\langle x, Ay \rangle = \langle Ax, y \rangle$

Вычислительно скалярное произведение считать проще, чем матричное, а значит для матриц A_1, A_2, A_3 и вектора x $\Rightarrow$ выражение $A_1 A_2 A_3 x$ счиатать лучше справа налево, а не наоборот.

Норма

Вектора

Количественная норма вектора, обозначается как $\|x\|$:

- $\|\alpha x\| = |\alpha| \ \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|x\| = 0 \Rightarrow x = 0$

Расстояние между векторами: $d(x, y) = \|x - y\|$

Зачастую используется **p-норма** (по-умолчанию p= 2): $\|x\|_p = \sqrt[p]{\sum_{i=0}^{n} |x_i|^p}$

Матричная норма

Аналогичная норма в пространстве матриц данного размера, самые часто используемые нормы:

- Норма Фробениуса: $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$
- Спектральная норма $\|A\|_2 = \sup_{\{x \neq 0\}} \frac{\|Ax\|_2}{\|x\|_2}$. Также $\|A\|_2 = \sigma_1(A) = \sqrt{\lambda_{\max(A^T A)}}$ ($\sigma_1(A)$ - максимальное сингулярное значение, может быть получено из синглуярного SVD разложения, например)

При этом для вектора $x \Rightarrow x^T x \neq x x^T$, поэтому будем использовать обозначение **внутреннего** (ie скалярного, inner/dot product) произведения, если результат - действительное число, и **внешнего** (ie тензорного, outer product), если результат будет матрицей/тензором.

Стандартное (фробениусово) **скалярное произведение матриц** $X, Y \in \mathbb{R}^{m \times n}$: $\langle X, Y \rangle = \operatorname{tr}(X^T Y) = \sum_i \sum_j X_{ij} Y_{ij} = \operatorname{tr}(Y^T X) = \langle Y, X \rangle$. Из него следует $\langle X X \rangle = \|X\|_F^2$

Собственные векторы и собственные значения

λ - Собственное значение $A \in \mathbb{R}^{n \times n}$, если $\exists q \neq 0 : Aq = \lambda q$

Теорема: если A $\succeq$ $(\succ)$ (ie Положительно полуопределенная или положительно определенная) $\Leftrightarrow$ все собственные значения неотрицательные (положительные)

Матричные разложения

Спектральное разложение $A \in \mathbb{S}_n$
Такая матрица A м/б представлена в виде:

$$A = Q \Lambda Q^T$$
$$\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$$
$$Q = (q_1, q_2, ..., q_n)$$

где λ_i - собственное значение связанное с собственным веткором q_i. При этом Q - ортогональная, т е $Q^T Q = I$

Синглярное (SVD) разложение $A \in \mathbb{R}^{m \times n}$: $\operatorname{rk} A = r$
Представление любой матрицы в виде произведения трех:

$$A = U \Sigma V^T$$

где U, V - ортогональные, а $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, ... \sigma_r) : \sigma_1 \geq \sigma_2 \geq ... \sigma_r$. Столбцы U называют левыми сингулярными векторами, столбцы V - правыми. Сингулярные значения в Σ : $\sigma(A) = \sqrt{\lambda(A^T A)}$. Исходная матрица A преставима в виде суммы матриц ранга 1:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Здесь используется именно внешнее произведение векторов u, v.

Далее важную роль играет **число обусловленности** матрицы $k = \|A\| \cdot \|A^{-1}\|$ также $k(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, если же $A \in \mathbb{S}_{++}^n \Rightarrow k(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$

Определитель и след

- $\det A = \prod_i^n \lambda_i$
- $\det(AB) = \det A \det B$
- $\det A^{-1} = \frac{1}{\det(A)}$
- $\operatorname{tr} A = \sum_i^n \lambda_i$
- $\operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA)$

Аппроксимация Тейлора 1-го порядка

Пусть:
- f - дифференцируемая функция $\mathbb{R}_n \to \mathbb{R}$

- $\nabla f(x_0)$ — градиент в точке x_0

Тогда можно линейно приблизить функцию f вблизи x_0:

$$f^{I}_{x_0}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0)$$

Квардратичная аппроксимация (2-го порядка):

$$f^{II}_{x_0}(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

где $\nabla^2 f(x_0)$ - матрица Гессе в точке x_0.

Матричное дифференцирование

Пусть $f : \mathbb{R}^n \to R$, определим:
- **Градиент** $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}\right)$
- **Матрица Гессе/гессиан**:

Матрица Гессе функции $f(x, y)$:

$$H(f) = \nabla^2 f = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Если $H \succ 0 \Rightarrow f$ - выпуклая, если $H \prec 0 \Rightarrow f$ - вогнутая.
- **Якобиан**: если $f : \mathbb{R}^n \to \mathbb{R}^m$ матрица Якоби имеет вид:

$$J(f, x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \vdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \vdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \vdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Обозначения

Если $f(x) : X \to Y, \frac{\partial f}{\partial x} \in G$
- $X = \mathbb{R}, Y = \mathbb{R} \Rightarrow G = \mathbb{R}$ - производная $f'(x)$
- $X = \mathbb{R}^n, Y = \mathbb{R} \Rightarrow G = \mathbb{R}^n$ - градиент $\frac{\partial f}{\partial x}$
- $X = \mathbb{R}^n, Y = \mathbb{R}^m \Rightarrow G = \mathbb{R}^{m \times n}$ - якобиан $\frac{\partial f_i}{\partial x_j}$
- $X = \mathbb{R}^{m \times n}, Y = \mathbb{R} \Rightarrow G = \mathbb{R}^{m \times n}$ - матричная производная $\frac{\partial f}{(\partial x_{ij}}$

Дифференциалы

Линейная часть приращения функции:

$$df(x) = \lim_{\|dx\| \to 0} f(x + dx) - f(x) = \langle \nabla f(x), dx \rangle$$

Свойства дифференциалов:
- $d(\alpha X) = \alpha dX$
- $d(AXB) = A(dX)B$
- $d\langle X, Y \rangle = \langle dX, Y \rangle + \langle X, dY \rangle$

- $d\left(\frac{X}{\varphi}\right) = \frac{\varphi(dX) - (d\varphi)X}{\varphi^2}$
- $d(\det X) = \det X \langle X^{-T}, dX \rangle$
- $d(\operatorname{tr} X) = \langle I, dX \rangle$
- $df(g(X)) = \frac{df}{dg} dg(x)$

Пример 1: $f(x) = \langle x, Ax \rangle - b^T x + c$

$df = d(\langle x, Ax \rangle - b^T x + c) = d(\langle Ax, x \rangle) - d(\langle b, x \rangle) + dc =$

$= \langle Adx, x \rangle + \langle Ax, dx \rangle - \langle b, dx \rangle = \langle A^T x, dx \rangle + \langle Ax, dx \rangle -$

$-\langle b, dx \rangle = \langle (A^T + A)x - b, dx \rangle \Rightarrow \nabla f = (A^T + A)x - b$

Пример 2: $f(x) = \ln(\langle x, Ax \rangle)$

$df = \frac{d\langle x, Ax \rangle}{\langle x, Ax \rangle} = \frac{\langle (A + A^T)x, dx \rangle}{\langle Ax, x \rangle}$

В силу неотрицательности аргумента логарифма $\Rightarrow A \in \mathbb{S}^n_{++}$(положительно определенная матрица)

$df = \langle \frac{2Ax}{\langle Ax, x \rangle}, dx \rangle) \Rightarrow \nabla f = \frac{2Ax}{\langle Ax, x \rangle}$

Пример 3 (матричная производная): $f(x) = \langle S, X \rangle - \ln \det X$

$df = d(\langle S, X \rangle - \ln \det X) = d(\langle S, X \rangle) - d(\ln \det X) = \langle S, dX \rangle -$

$-\frac{d \det X}{\det X} = \langle S, dX \rangle - \frac{\det X \langle X^{-T}, dX \rangle}{\det X} = \langle S - X^{-T}, dX \rangle$

$\Rightarrow \nabla f = S - X^{-T}$