Crib 1

01 Background

by Alvin Wan . alvinwan.com/cs189/fa17

1 Conventions, Definitions

1. $x \in \mathbb{R}^d$ (column vector), $X \in \mathbb{R}^{n \times d}$ (n samples, d features per sample)
2. A is **positive semidefinite (PSD)** iff (all equivalent conditions):
   - unique Cholesky decomposition $A = BB^T$, where B is lower triangular
   - $\forall x^T Ax \geq 0$
   - all eigenvalues ≥ 0
3. A is **positive definite (PD)** iff $\forall x, x^T Ax > 0$ iff all eigenvalues > 0.
4. **Eigendecomposition**: $A = P\Lambda P^T = \sum_i v_i \lambda_i v_i^T$
   - D diagonal with λ_i entries, P comprised of eigenvectors.
   - a.k.a., diagonalization, spectral decomposition
   - admitted by any real, symmetric matrix
5. **Singular Value Decomposition (SVD)**: $A = U\Sigma V^T = \sum_i u_i \sigma_i v_i^T$
   - Λ contains the singular values σ_i along its diagonal, U are left singular vectors and V are right singular vectors
   - any matrix, doesn't need to be square or symmetric
6. If a matrix A is PSD, eigenvalues of A are the same as singular values of A.

2 Ordinary Least Squares (OLS)

1. $\frac{\partial x^T w}{\partial x} = \frac{\partial w^T x}{\partial x} = w$ for $w, x \in \mathbb{R}^d$ (implies $\frac{\partial Ax}{\partial x} = A^T$ for $A \in \mathbb{R}^{n \times d}, x \in \mathbb{R}^d$)
2. $\frac{\partial f(x)^T g(x)}{\partial x} = \frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x)$ for $f(x), g(x), x \in \mathbb{R}^d$ (derivation below)
3. objective: $\min_w \|Xw - y\|_2^2$, $w, x \in \mathbb{R}^d, X \in \mathbb{R}^{n \times d}, y \in \mathbb{R}^n$
4. solution: $w^* = (X^T X)^{-1} X^T y$, prediction: $\hat{y} = Xw^*$

3 Appendix: Product Rule

Let's start by deriving the product rule. First, consider $f(x) = \begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \end{bmatrix}^T$ and a similar definition for $g(x)$. Note that $f_i(x), g_i(x)$ are scalar-valued functions. e.g., $f : \mathbb{R}^d \to \mathbb{R}$.

$$\frac{\partial f(x)^T g(x)}{\partial x} = \frac{\partial \sum_i f_i(x) g_i(x)}{\partial x} = \sum_i \frac{\partial f_i(x) g_i(x)}{\partial x} \tag{1}$$

Take the partial for each i, first. Here, I will use the notation $f_{i,x_j}(x) = \frac{\partial f_i(x)}{\partial x_j}$.

$$\frac{\partial f_i(x) g_i(x)}{\partial x} = \begin{bmatrix} f_{i,x_1}(x)^T g_i(x) + f_i(x)^T g_{i,x_1}(x) \\ f_{i,x_2}(x)^T g_i(x) + f_i(x)^T g_{i,x_2}(x) \\ \vdots \\ f_{i,x_n}(x)^T g_i(x) + f_i(x)^T g_{i,x_n}(x) \end{bmatrix} = \frac{\partial f_i(x)}{\partial x} g_i(x) + f_i(x) \frac{\partial g_i(x)}{\partial x}$$

Note that $g_i(x)$ is a scalar but $\frac{g_i(x)}{\partial x}$ is a vector; the same applies to f. Continuing with (1), we plug in:

$$\begin{aligned} &= \sum_i \frac{\partial f_i(x)}{\partial x} g_i(x) + f_i(x) \frac{\partial g_i(x)}{\partial x} \\ &= \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} & \frac{\partial f_2(x)}{\partial x} & \cdots & \frac{\partial f_n(x)}{\partial x} \end{bmatrix} g(x) + \begin{bmatrix} \frac{\partial g_1(x)}{\partial x} & \frac{\partial g_2(x)}{\partial x} & \cdots & \frac{\partial g_n(x)}{\partial x} \end{bmatrix} f(x) \\ &= \frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x) \end{aligned}$$

Thanks to Jonathan Xia for helping me. :P

Example) Let us see an example of product rule.

Take $\frac{\partial x^T A x}{\partial x}$. We will take $f(x) = x$, the first x, and $g(x) = Ax$, to fit the form $\frac{\partial f(x)^T g(x)}{\partial x}$. Plug in to the formula for product rule above and solve.

$$\frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x) = \frac{\partial x}{\partial x} Ax + \frac{\partial Ax}{\partial x} x = Ax + A^T x = (A + A^T)x$$