LINEAR ALGEBRA

Least Squares

$$\min_{x \in \mathbb{R}^n} \|A\vec{x}-\vec{y}\|_2^2 \quad : \quad \vec{x}^* = (A^TA)^{-1}A^Ty$$

cond: A has full col rank
$y \in \mathbb{R}^m$

Norms

$f: X \to \mathbb{R}$ a norm if
1. $f(\vec{x}) \geq 0$
2. $f(\alpha\vec{x}) = |\alpha| \, f(\vec{x})$
3. $f(\vec{x}+\vec{y}) \leq f(\vec{x}) + f(\vec{y})$

l_p Norms

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

$$\|\vec{x}\|_\infty = \max |x_i|$$

$$\|\vec{x}\|_2^2 = \sum_{i=1}^{n} x_i^2$$

Cauchy - Schwarz Inequality

$|\vec{x}^T\vec{y}| \leq \|\vec{x}\|_2 \, \|\vec{y}\|_2$ (max when $\theta = 0$)

Grahm - Schmidt Algorithm

1. $\text{span}\{a_i\} = \text{span}\{q_i\}$
2. $\{q_i\}$ orthonormal set

Steps
- $q_1 = \frac{\vec{a}_1}{\|a_1\|}$
- $p_2 = (a_2^Tq_1)\,q_1$
  $s_2 = a_2 - p_2$
  $q_2 = \frac{s_2}{\|s_2\|}$

$$s_k = a_k - (a_k^Tq_1)\,q_1 - (a_k^Tq_2)\,q_2 - \dots (a_k^Tq_{k-1})\,q_{k-1}$$

QR Decomposition

$$A = \underbrace{Q}_{\text{Orthonormal}} \; \underbrace{R}_{\text{Upper Triangular}}$$

cond: A has full col rank

FTLA

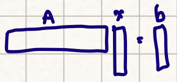
- $N(A) \oplus R(A^T) = \mathbb{R}^n$
- $N(A^T) \oplus R(A) = \mathbb{R}^m$
- $S \oplus S^\perp = \mathbb{R}^n$

If $U \oplus V = \mathbb{R}^n$, $\underset{\in \mathbb{R}^n}{\vec{x}} = \underset{\in U}{\vec{x}_1} + \underset{\in V}{\vec{x}_2}$

Min - Norm Solution

$$\min_{x \in \mathbb{R}^n} \|\vec{x}\|_2^2 \quad : \quad \vec{x}^* = A^T(AA^T)^{-1}\vec{y}$$
$$\text{s.t. } A\vec{x} = \vec{y}$$

cond: A has full ROW rank (wide)



pick 1 from ∞ soln

(A^TA) invertible when A is full c rank

Orthonormal Set $\{v_1, v_2 \cdots v_n\}$
- v_i has norm 1
- $v_i \perp$ to all v_j

$P^{-1}P = I$
$\det(I) = 1$

$$\left\|\begin{bmatrix}\vec{v}\\\vec{w}\end{bmatrix}\right\|_2^2 = \|\vec{v}\|_2^2 + \|\vec{w}\|_2^2$$

A full rank $\Rightarrow$ trivial nullspace
A not full rank $\Rightarrow$ non-trivial nullspace

$\text{proj}_x z = \langle z, x \rangle x$
$\|x\|_2 = 1$

Orthonormal Matrix Q

- $Q^TQ = I$
- $Q = Q^T$
- $Q^T = Q^{-1}$ if Q is square

Symmetric Matrices

Symmetric if

1. A is square
2. $A = A^T$

Properties

1. Real eigenvalues
2. Diagonalizable

Spectral Theorem

1. Real eigenvalues
2. Eigenspaces orthonormal
3. A orthogonally diagonalizable

$$A = U \underbrace{\Lambda}_{\text{diagonal}} \underbrace{U^T}_{\text{orthonormal}}$$

(λ_i, u_i): (eigenvalue, eigenvector)

Anything of form
A^TA , AA^T is symmetric

Diagonalizability

- ↳ When A is square
- ↳ When $\text{algebraic}_m = \text{geometric}_m$

Orthonormal

- ↳ $A^T = A^{-1}$
- ↳ $A^TA = I$
- ↳ Doesn't affect norm

Rayleigh Quotient

$$\lambda_{max}\{A\} = \max_{\|x\|=1} \vec{x}^T A \vec{x}$$

$$\lambda_{min}\{A\} = \min_{\|x\|=1} \vec{x}^T A \vec{x}$$

A is symmetric

Positive semidefinite

PSD if

1. A is symmetric
2. $\vec{x}^T A \vec{x} \geq 0$
3. Every eigenvalue is +

Properties

1. $B^{1/2} = A$, B exists

PCA, SVD

PCA

$$\min_w (err(w_1)) = \frac{1}{n}\sum_{i=1}^{n} \| x_i - w_1(w_1^T x_i)\|_2^2$$

$$= \lambda_{max}\left\{\frac{X^TX}{n}\right\}$$

SVD

$A = U\Sigma V^T$

Compact:

$$A = \underset{m \times r}{U_r} \underset{r \times r}{\Sigma_r} \underset{r \times n}{V^T}$$

Full:

$$A = \underset{m \times m}{U} \underset{m \times n}{\Sigma} \underset{n \times n}{V^T}$$

Dyadic:

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

→ U : orthonormal
left singular
rotation/ reflection

→ Σ : singular values
diagonal, all +
$\sigma_i = \sqrt{\lambda_i}$ of A^TA
scaling

→ V^T : orthonormal
right singular
v_i of A^TA
rotation/ reflection

Add basis

LOW RANK APPROX

Frobenius Norm

$$\|A\|_F = \sqrt{\sum^m \sum^n A_{ij}^2}$$

$$\|A\|_F^2 = tr(A^T A)$$

$$\|UAV\|_F = \|A_F\|$$, U,V orthonormal

$$\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2$$

$$\|A\|_F \geq \|A\|_2$$

Spectral Norm

$$\|A\|_2 = \max_{\|x\|=1} \|A\vec{x}\|_2 = \max \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}$$

$$\|A\|_2 = \sigma_{max}\{A\}$$

Eckart Young Theorem

$$\|A - A_K\|_2 \leq \|A - B\|_2 = \sigma_{K+1}$$

$$\|A - A_K\|_F^2 \leq \|A - B\|_F^2 = \sum_{i=K+1}^{r} \sigma_i^2$$

VECTOR CALC

Gradient

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}) \end{bmatrix}$$

$\nabla f(\vec{x}) \perp$ hyperplane tangent to $\vec{x}$

Jacobian

$$Df(\vec{x}) = \begin{bmatrix} \nabla f_1(\vec{x})^T \\ \nabla f_2(\vec{x})^T \\ \vdots \\ \nabla f_m(\vec{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

(transpose of the gradient)

Hessian

$$\nabla^2 f(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_n x_1}(\vec{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 x_n}(\vec{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\vec{x}) \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_i x_j}(\vec{x}) = \frac{\partial^2 f}{\partial x_j x_i}(x)$$

Taylor's Theorem

$$f(x; x_0) = f(x_0) + \frac{\partial f}{\partial x}(x_0)(x - x_0) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0)(x - x_0)^2 \cdots + \frac{1}{K!}\frac{\partial^K f}{\partial x^K}(x_0)(x - x_0)^K$$

Taylor's Approximation for Multivariate

$$f_1(x; x_0) = f(\vec{x}_0) + [\nabla f(\vec{x}_0)]^T(\vec{x} - \vec{x}_0)$$

$$f_2(x; x_0) = f(\vec{x}_0) + [\nabla f(\vec{x}_0)]^T(\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^T[\nabla^2 f(\vec{x}_0)](\vec{x} - \vec{x}_0)$$

Main Theorem

$\min_{\vec{x} \in \Omega} f(\vec{x})$: $\nabla f(\vec{x}^*) = 0$, $f: \mathbb{R}^n \to \mathbb{R}$ is differentiable

Directional Derivative

$f: R^n \to R \ , \ \|\vec{u}\|_2 = 1$

$D_{\vec{u}} f(\vec{x}) = u^T [\nabla f(\vec{x})]$

REGRESSION

$A(\vec{x} + \delta\vec{x}) = (\vec{y} + \delta\vec{y})$

$\frac{\|\delta\vec{x}\|_2}{\|\vec{x}\|_2} = \|A\|_2 \, \|A^{-1}\|_2 \cdot \frac{\|\delta\vec{y}\|^2}{\|\vec{y}\|^2}$

A is invertible

Condition Number

$\kappa(A) = \frac{\sigma_1\{A\}}{\sigma_n\{A\}}$

$K(A^TA) = \frac{\lambda_{max}\{A^TA\}}{\lambda_{min}\{A^TA\}}$

Ridge Regression

$\min_{\vec{x} \in \mathbb{R}^n} \{ \|A\vec{x} - \vec{y}\|_2^2 + \lambda \|\vec{x}\|_2^2 \}$

$\vec{x}^* = (A^TA + \lambda I)^{-1} A^T\vec{y}$

A^TA is PSD, $\lambda > 0$, $(A^TA + \lambda I)$ PSD so invertible

SVD: $\vec{x}^* = \sum_{i=1}^{r} \frac{\sigma_i\{A\}}{\sigma_i\{A^2\} + \lambda} (\vec{u}_i^T \vec{y}) \cdot \vec{v}_i$

Tikhanov Regression

$\min_{\vec{x} \in \mathbb{R}^n} \{ \|W_1(A\vec{x} - \vec{y})\|_2^2 + \|W_2(\vec{x} - \vec{x}_0)\|_2^2 \}$

$\vec{x}^* = (A^TW_1^2A + W_2^2)^{-1}(A^TW_1^2\vec{y} + W_2^2\vec{x}_0)$

MLE as Tikhanov

$\operatorname{argmax} p_{\vec{x}}(\vec{y}) = \operatorname{argmin}_{\vec{x} \in \mathbb{R}^n} \|\Sigma_{\vec{w}}^{-1/2}(A\vec{x} - \vec{y})\|_2^2$ ⟶ what $\vec{x}$ makes data most likely

MAP as Tikhanov

$\operatorname{argmax} p(\vec{x}|\vec{y}) = \operatorname{argmin}_{\vec{x} \in \mathbb{R}^n} \{ \|\Sigma_{\vec{w}}^{-1/2}(A\vec{x} - \vec{y})\|_2^2 + \|\Sigma_{v}^{-1/2}(\vec{x} - x_0)\|_2^2$ ⟶ what $\vec{x}$ is most likely

CONVEXITY

Convex Set

$\vec{C}$ is convex if $\vec{x}_1, \vec{x}_2 \in C \ , \ \theta \in [0,1]$

$\underbrace{\theta\vec{x}_1 + (1-\theta)\vec{x}_2}_{\text{point on line segment b/w } x_1, x_2} \in C$

↳ $C_1 \cap C_2$ is convex if C_1, C_2 convex

Hyperplane

$\vec{a}^T\vec{x} = b$, all convex

Convex Function

$f: \mathbb{R}^n \to \mathbb{R}$ convex if

1. $\operatorname{dom}(f)$ convex
2. $f(\theta x_1 + (1-\theta)x_2) \leq \theta f(\vec{x}_1) + (1-\theta) f(x_2)$

SOCP
constraints $\| \text{vecter/scalar affine} \|_2^2 \leq$ Scalar affine