$x, y \quad x, y \leftarrow \mathbb{R}^N$ jgscott.github.com / SSC 383D /

$\langle x, y \rangle = x^T y = \sum_{i=1}^{N} x_i y_i$

$\|x\|_2 = (x^T x)^{1/2}$ Euclidean norm = length of a hypoteneus

$f: \mathbb{R}^N \to \mathbb{R}$

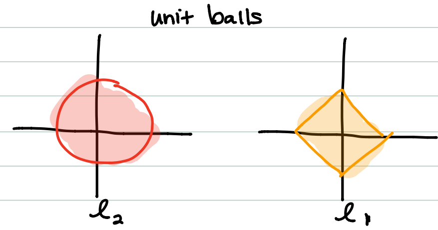
1) $f(x) \geq 0$
2) $f(x) = 0 \leftrightarrow x = 0$
3) $f(tx) = |t| f(x)$
4) $f(x+y) \leq f(x) + f(y)$

$f(x) = \|x\|$

Norms are useful for calculating distances

$\text{dist}(x,y) = \|x-y\|$

$B \subset \mathbb{R}^N = \{x \in \mathbb{R}^N : \|x\| \leq 1\}$



$\|x\|_p = \{|x_1|^p + \ldots + |x_N|^p\}^{1/p}$

$\|x\| = (x^T Q x)^{1/2}$

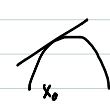
$Q \in S_+^N$

Gradients/Hessions

$f: \mathbb{R}^N \to \mathbb{R}$

$\nabla f(x)$ gradient

$\nabla f(x) = \left\{ \frac{\partial f(x)}{\partial x_i} \right\}_{i=1}^{N}$ $\quad f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0)$



$f(x) = \frac{1}{2} x^T P x + x^T q + r$

$\nabla f(x) = Px + q$ gradient of above quadratic function

$H f(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{\substack{i = 1 \ldots N \\ j = 1 \ldots N}}$

Linear Algebra

Let $A \in \mathbb{R}^{n \times p}$ "set of matrices of real #s w/ n rows + p columns)

$R(A) = \text{range} = \{ Ax \cdot x \leftarrow \mathbb{R}^p \}$ subspace of $\mathbb{R}^n$

$\text{rank}(A) = \text{dimension}(R(A)) \leq \min(n, p)$

$N(A) = \text{null space} = \{ x : Ax = 0 \}$ subspace of $\mathbb{R}^p$

V is a subspace of $\mathbb{R}^n$

$V^{\perp} = \{ x : z^T x = 0 \ \forall \ z \in V \}$ orthoganol complement is useful for understanding linear regression

Basic result

$N(A) = R(A^T)^{\perp}$

$R(A) = N(A^T)^{\perp}$

take any $X \in \mathbb{R}^N$

V subs of $\mathbb{R}^N$

$X = V_1 + V_2$

$V_0 = V$

$V_1 = V^{\perp}$

$y = X\beta + \varepsilon$ fitted values + residuals form orthoganol pair

$y = \hat{y} + \varepsilon$

Decompositions from linear algebra

$A \in S^N$ S = "symmetric" N = # dimensitions

Any such A

$A = Q \Lambda Q^T$ spectral decomposition (useful in PCA)

Q = orthoganol matrix (columns are orthoganol)

$Q(q \dots q_N)$

$q_i^T q_j = \emptyset$ if $i \neq j$

$q_i^T q_j = 1$ if $i = j$

$\Lambda = \text{diagonal}(\lambda_1 \dots \lambda_N)$ λ_i = eigenvalues

$A q_i = \lambda_i q_i$

A symmetric positive definite: $A \in S^N_{++}$ covariance matrices

$x^T A x > 0$ for all $x \neq 0$

positive semi-definite $A \in S^N_{+}$ $(x^T A x \geq 0)$

$\Sigma \in S^N_{++}$

$\Sigma = LL^T$

L lower triangular $\cdot$ $\begin{pmatrix} \ell_{11} & & & 0 \\ & \ell_{22} & & \\ \text{stuff} & & \ell_{33} & \\ & & & \ell_{nn} \end{pmatrix}$

Cholesty decomposition

$X \in \mathbb{R}^{n \times p}$

$X = UDW^T$ $\qquad r = \text{rank}(x)$

$U^T U = I \;, W^T W = I, \; D = \text{diag}(d_1 \ldots d_r)$

Stats / Probability

Sampling distributions

Confidence intervals

linear regression

analysis of variance ANOVA, R^2

residuals

$\mathcal{Y}$ = sample space

Θ = parameter space

$y \in \mathcal{Y}$

$\theta \in \Theta$

$p(\theta)$ prior distribution

$p(y|\theta)$ sampling model / likelihood

$p(y, x)$ joint prob of x & y

$= p(y)\, p(x|y)$

$p(x)\, p(y|x)$

$$\underset{\text{posterior}}{p(y|x)} = \frac{\overset{\text{prior}}{p(y)}\, p(x|y)}{p(x)}$$ Bayes Rule

θ = unknown rate (Poisson)

$\Theta = [0, \infty) = \mathbb{R}^+$

$$p(Y = k | \theta) = \frac{\theta^k}{k!} e^{-\theta}$$

$p(\theta) = \text{Gamma}(\theta | a, b)$

$$= \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$

$$p(y_1 \ldots y_N | \theta) = \prod_{i=1}^{N} p(y_i | \theta)$$

$$= \prod_{i=1}^{N} \left[\frac{\theta^{y_i}}{y_i!} e^{-\theta} \right] = \frac{\theta^{\Sigma y_i} e^{-n\theta}}{K}$$

$$p(\theta|y_1 \dots y_N) = \frac{1}{p(x)} \cdot p(\theta) \cdot p(y|\theta)$$

$$= \frac{1}{p(x)} \frac{b^a}{\Gamma(a)} \frac{\theta^{a-1} e^{-b\theta} \theta^{\Sigma y_i} e^{-n\theta}}{k}$$

$$= \frac{1}{K_1} \theta^{\Sigma y_i + a - 1} e^{-(b+n)\theta}$$

$$= \text{Gamma}(a + \Sigma y_i, b+n)$$