Let $A \in \mathbb{R}^{n \times n}$ be a matrix.

Def: If $Ax = \lambda x$ for some scalar $\lambda \in \mathbb{C}$ and some nonzero vector $x \in \mathbb{R}^n$, then λ is called an <u>eigenvalue</u> of A, and x is called an <u>eigenvector</u> of A associated to λ.

Let $u, v \in \mathbb{R}^n$ be two vectors

Def: The <u>dot product / inner product</u> between u and v is

$$u \cdot v = u^T v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Def: The <u>ℓ^p-norm</u> of u is

$$|u|_{\ell^p}^p = |u_1|^p + \cdots + |u_n|^p$$

$p=1$: $\quad |u|_1 = |u|_{\ell^1} = |u_1| + \cdots + |u_n|$

$p=2$: $\quad |u|_2^2 = |u|_{\ell^2}^2 = |u_1|^2 + \cdots + |u_n|^2$

Def: The <u>transpose</u> of $A = (A_{ij})_{m\times n}$ is $A^T = (A_{ji})_{n\times m}$

A matrix A is called <u>symmetric</u> if $A = A^T$.

A matrix U is called <u>orthogonal</u> if $U^T U = U U^T = I$

Remark: U is orthogonal if and only if the columns of U form an orthonormal basis.

Theorem: (Eigenvalue Decomposition for Symmetric Matrices)

If $A \in \mathbb{R}^{n\times n}$ is sym, then there exists an orthogonal U such that

$$UAU^T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

where $\lambda_1, \cdots \lambda_n$ are eigenvalues of A.

Def: Let $A \in \mathbb{R}^{m\times n}$. $\sigma \geq 0$ is called a <u>singular value</u> of A if σ^2 is an eigenvalue of A^TA or AA^T

Remark (1) The non-zero eigenvalues of A^TA and AA^T agree.

Theorem (Singular Value Decomposition) Let $A \in \mathbb{R}^{m\times n}$. There exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ s.t.

$$UAV^T = \begin{cases} \left(\begin{array}{ccccc|c} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & 0 \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{array}\right)_{m \times n} & n \geq m \\ \left(\begin{array}{cccccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ \hline & & & 0 & & \end{array}\right)_{m \times n} & m \geq n \end{cases}$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ and r is the rank of A.

Proposition: (1) $\operatorname{Ker}(A) = \operatorname{Ran}(A^T)^{\perp}$

$\operatorname{Ker}(A^T) = \operatorname{Ran}(A)^{\perp}$

(2) $\operatorname{Ker}(A)^{\perp} = \operatorname{Ran}(A^T)$

$\operatorname{Ker}(A^T)^{\perp} = \operatorname{Ran}(A)$

2 Linear Methods in Regression.

2.1 Linear Regression

We denote a sample point by
$x = (x_1, \cdots, x_p)^T \in \mathbb{R}^p$
where p is the # of features.

- Noise-free model: linear regression assumes $y = f(x, \beta^*)$ is a linear function with parameter β^*:

$$y = f(x; \beta^*) = x^T \beta^* \left(= (\beta^*)^T x \right) = \beta_1^* x_1 + \cdots + \beta_p^* x_p$$

Remark: A general linear function of x is
$$y = \underbrace{\beta_0^*}_{\text{const.}} + \beta_1^* x_1 + \cdots + \beta_p^* x_p$$

In this case, define $\tilde{x} \stackrel{\text{def}}{=} (1, x_1, \cdots, x_p)$
$\tilde{\beta}^* = (\beta_0^*, \beta_1^*, \cdots, \beta_p^*)$, then

$y = \tilde{x}^T \tilde{\beta}^*$

- Noisy model: in the presence of noise ϵ (random) is

$$y = x^T \beta^* + \epsilon.$$

Goal of linear regression: estimate β^* from the noisy measurement.

Given the sample points:

$$\{(x^{(1)}, y_1), (x^{(2)}, y_2), \dots, (x^{(n)}, y_n)\}$$

Need to choose β s.t.

$$\begin{cases} x^{(1)T} \beta = y_1 \\ \vdots \\ x^{(n)T} \beta = y_n \end{cases} \quad ①$$

Set $X \stackrel{\text{def}}{=} \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \\ | & | & & | \end{bmatrix}_{p \times n}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Then ① $\Leftrightarrow$ $$X^T \beta = y$$

Need to solve for β.