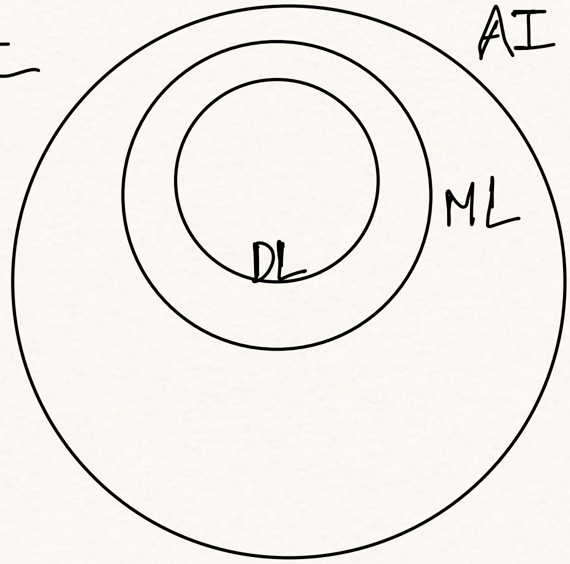
Diff. b/w AI, ML, DL



Notation

Vector : $v \in \mathbb{R}^d$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$$

$$v^T = [v_1 \ v_2 \ \cdots \ v_d]$$

Matrix : $A \in \mathbb{R}^{m \times n}$

$$A = m\left\{ \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix}}_{n} \right.$$

m row-vectors of n dimension

or

n col vectors of m dim.

Vector-vector operations

• Inner product / dot product (Diff. concepts but same for this course)

$x, y \in \mathbb{R}^d$

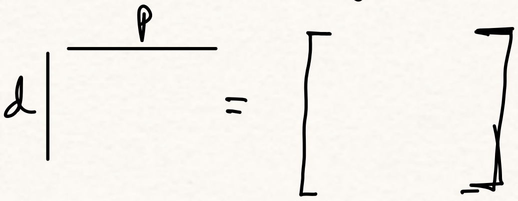
$$\sum_{i=1}^{d} x_i y_i \quad \text{or} \quad x^T y$$

(2 vectors) = • (scalar)

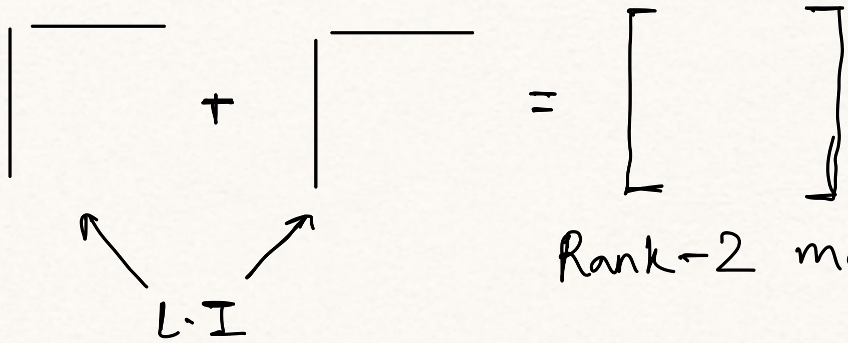
• Outer Product

$x \in \mathbb{R}^d, \; y \in \mathbb{R}^p$

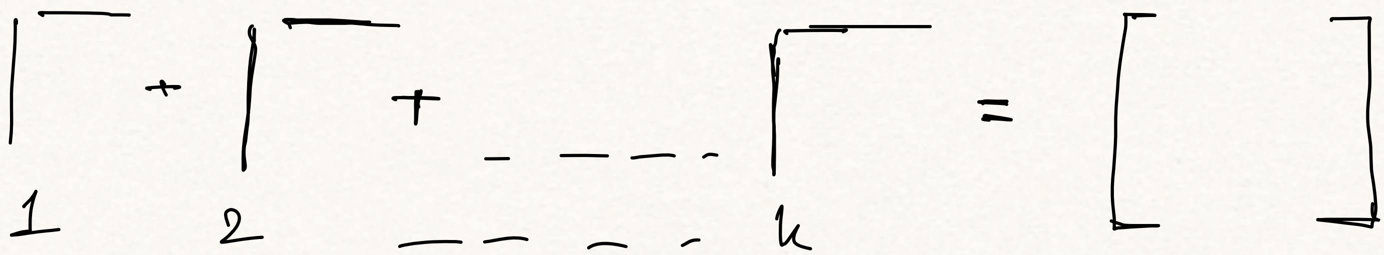
$xy^T \qquad yx^T$



Rank-1 matrix

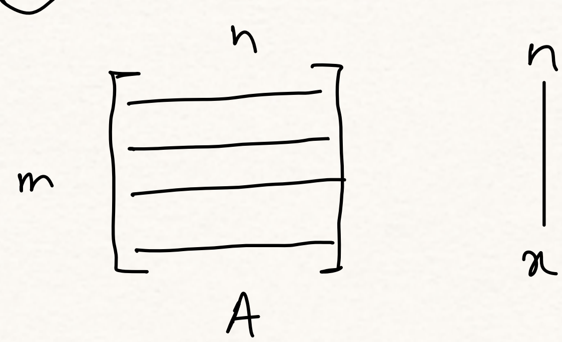


Rank-2 matrix



$\text{Rank} \leq \min(d, p, k)$

Matrix - vector operations



$A \in \mathbb{R}^{m \times n}$ $\qquad$ $x \in \mathbb{R}^n$

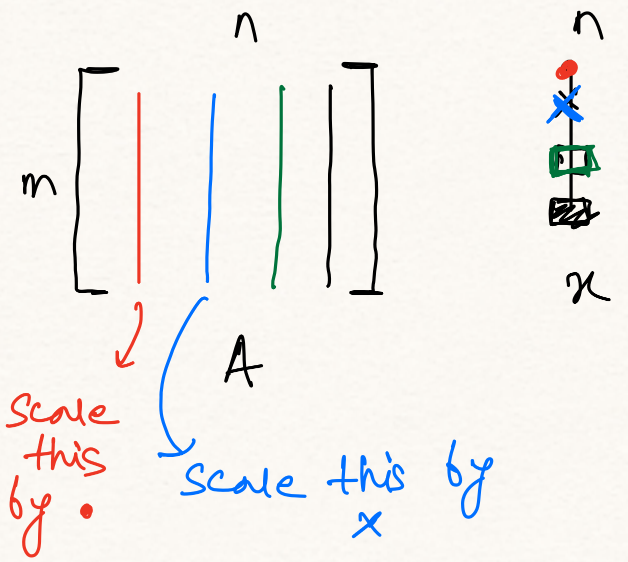
$Ax = \begin{bmatrix} \circ \\ \circ \\ \vdots \\ \circ \end{bmatrix} m$

$Ax \in \mathbb{R}^m$

Interpretation

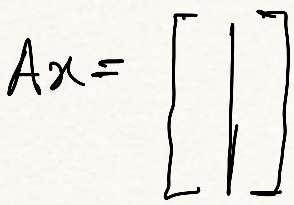
I

m. inner products of n-dim vectors.



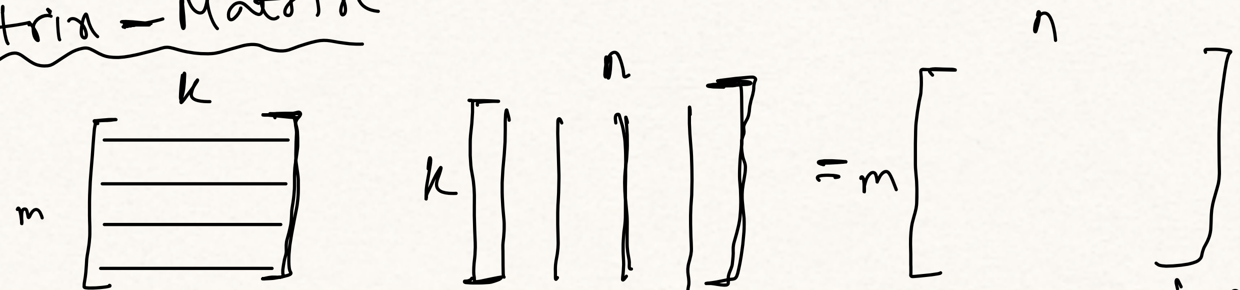
Scaling means multiplying each element of that vector with another element.

Interpretation II

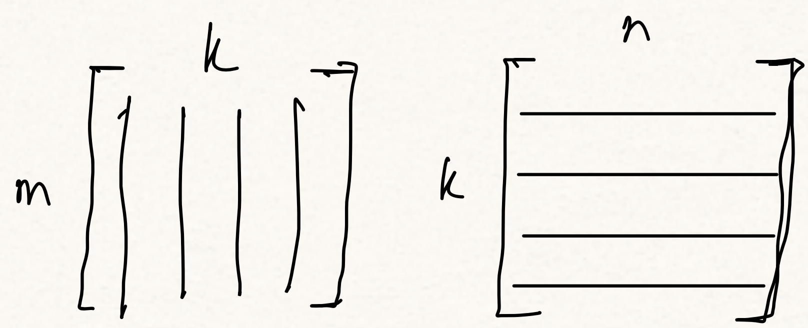


↓ sum of all the 'scaled' m-dim. vectors

Matrix – Matrix



inner-product



$$c_1 \Big| \overline{\;r_1\;} + c_2 \Big| \overline{\;r_2\;} + \ldots\; c_k \Big| \overline{\;r_k\;}$$

outer-product

* Both results will be same.

Application in ML

① Represent Data

$X = $ design matrix, rows $H_1, H_2, \ldots, H_n$; columns → features

$y = $ value vector: $\begin{bmatrix} \text{price of } H_1 \\ \vdots \\ \text{price of } H_n \end{bmatrix}$

② Covariance Matrices

$S \in \mathbb{R}^{d \times d}$

③ Calculus :

Gradients – Vector

Hessians – Matrix (Symmetric)

Jacobians – Matrix

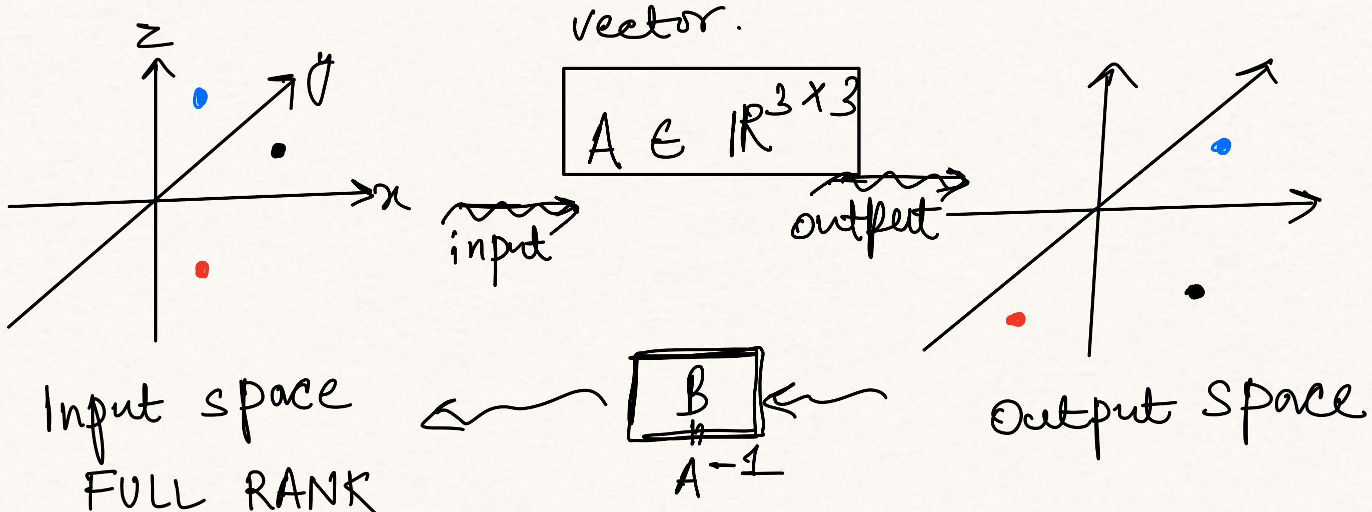
④ Kernel Method

Geometrical Interpretation

$A \in \mathbb{R}^{m \times n}$ $\quad$ $A(x) \in \mathbb{R}^m$

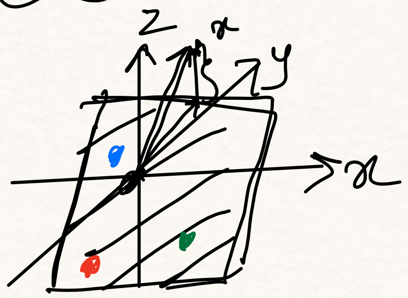
$x \in \mathbb{R}^n$

↳ Interpret this as func^n which takes vector as input and outputs a vector.

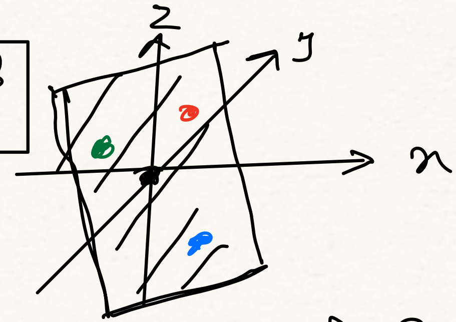


- If A is full-rank then B exists.
- A is full-rank $\iff$ one to one mapping b/w I/O.

RANK DEFICIENT (Rank-2)



$\boxed{A \in \mathbb{R}^{3\times 3}}$



Rank-2 interpretation $\Rightarrow \exists$ a 2-D subspace in input & output space which have one to one I/O mapping.

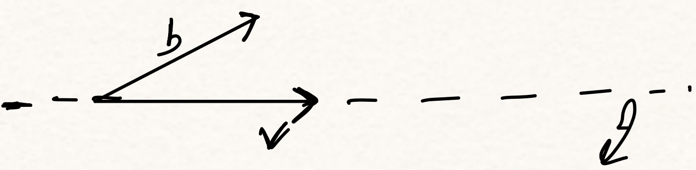
What about points outside this subspace?

Decomposition of this vector.

$$x = \text{Proj}(x; \text{Row space}) + \text{Proj}(x; \text{null space})$$

$$\begin{aligned} A(x) &= A(x_R + x_N) \\ &= A(x_R) + \underbrace{A(x_N)}_{=0} \\ &= A(x_R) \end{aligned}$$

Projection



subspace spanned by v.

Projection matrix(v)

$$= \left[\frac{vv^T}{v^Tv}\right] b$$

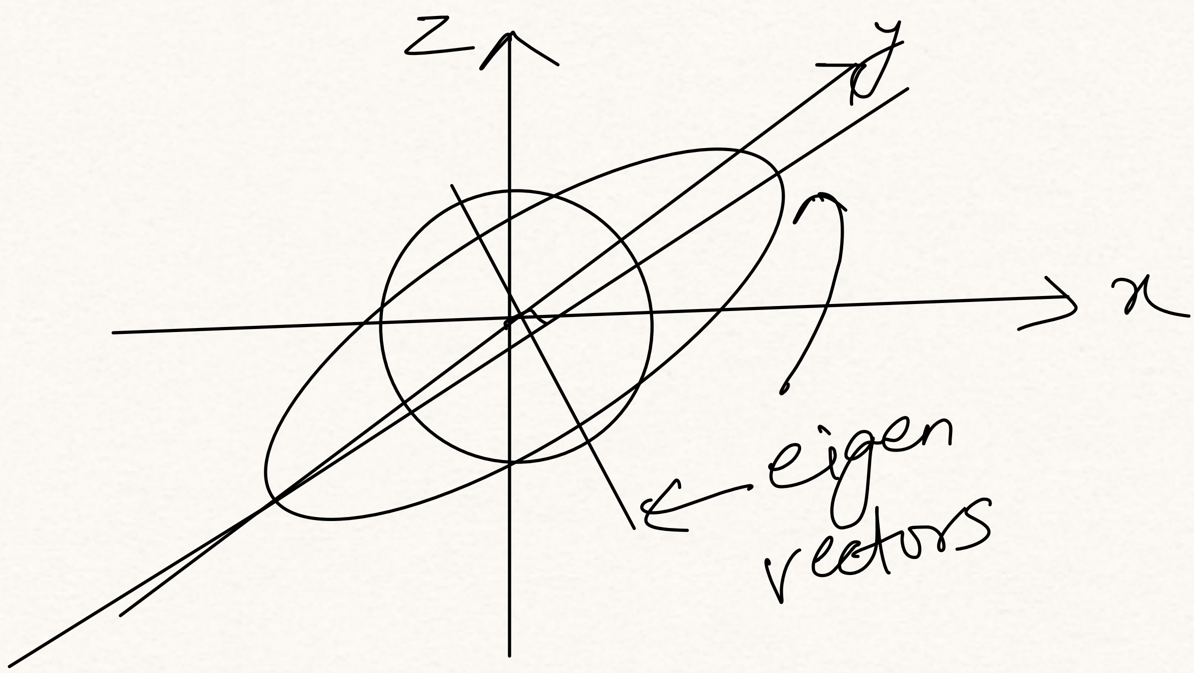
$$\frac{vv^T}{v^Tv} b = \left(\frac{v}{\|v\|}\right)\left(\frac{v}{\|v\|}\right)^T b$$

$$= [\tilde{V}\,\tilde{V}^T]\,b$$

$$= \tilde{V}\,\underbrace{(\tilde{V}^T b)}_{\text{length}}$$

$$X = \underbrace{\begin{bmatrix} | & | & | & | \\ | & | & | & | \end{bmatrix}}_{\text{set of L.I vectors}} \qquad X(X^TX)^{-1}X^T \rightarrow \text{projection matrix.}$$

$A \in \mathbb{R}^{3\times 3} \rightarrow$ symmetric



Determinant = Product of all eigen values

$$= \frac{\text{Vol. of output shape}}{\text{Vol. of input shape}}$$

For Non-full rank, atleast one dim. will be 0.

Thus, det. $= \frac{0}{x} = 0$.

Spectrum: collection of Eigen values

Spectral Theorem

$A \in \mathbb{R}^{d \times d}, \ A = A^T$

→ Real valued eigen values
→ orthonormal eigen vectors.

Hessians, Covariance Matrix, Kernel

square & symmetric.

Quadratic Forms

$A \in \mathbb{R}^{d \times d}$, $x \in \mathbb{R}^d$

$x^T A x$ — Quadratic Form

$x^T B x = x^T A x \mid B = B^T$

$B = \frac{1}{2} A + \frac{1}{2} A^T$

Definitiveness

$x^T A x > 0 \quad \forall x \neq 0 \rightarrow$ A is positive definite

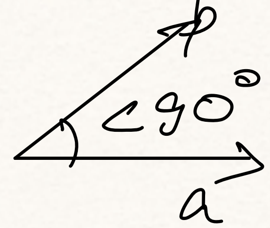
$\geq 0 \quad \forall x \neq 0 \rightarrow$ A is positive semi definite

$< 0 \quad$ " $\quad$ " negative definite

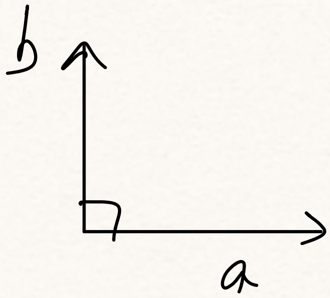
$\leq 0 \quad$ " $\quad$ " " semi-def.

$>, < \quad$ " $\quad$ " indefinite

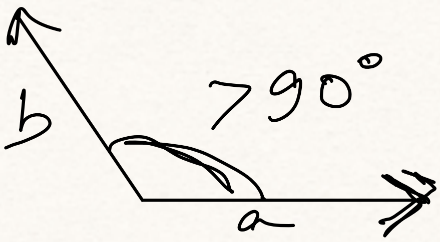
$a^T b$



> 0



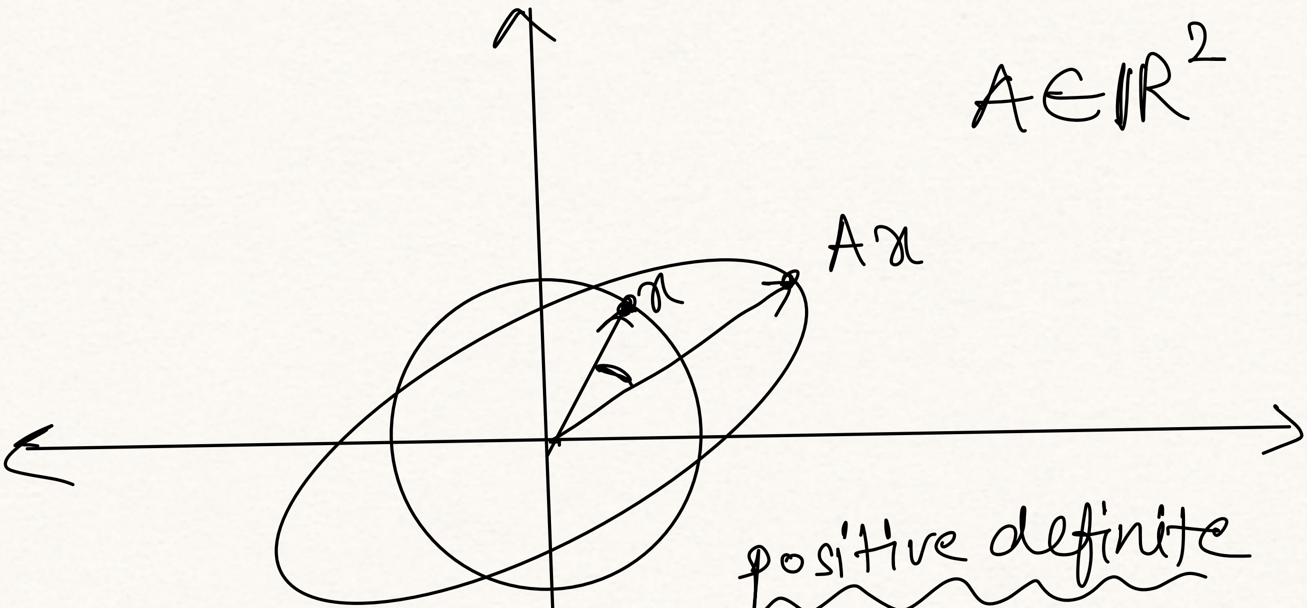
$= 0$



< 0

$$\underbrace{(x)}_{\text{input}}{}^T \underbrace{Ax}_{\text{output}} \rightarrow \text{dot product of input \& output}$$

$A \in \mathbb{R}^2$



positive definite

⇒ For any input x the angle b/w x & output Ax will be $< 90°$.

For a P.D matrix

$\Rightarrow$ All eigenvalues > 0

Link b/w definitiveness & spectrum.

P.S.D

$\rightarrow$ " " ≥ 0

N.D

$\rightarrow$ " " < 0

N.S.D

$\rightarrow$ " " ≤ 0

I.D

$\rightarrow$ " " $>, < 0$

Decomposition of Matrices

- Singular value Decomposition
- Eigen value Decomposition.
- Cholesky Decomposition.

	A	Decomposition
SVD	Any	$A = USV^T$
EVD	square	$A = UDU^{-1}$

$$A(x) = U(S(V^T(x)))$$

Decomposed func^{n}s/ matrices

$U, V \Rightarrow$ orthonormal matrices

$S \rightarrow$ Diagonal

$D \rightarrow$ Diagonal

For EVD

$U \rightarrow$ set of eigen vectors

$D \rightarrow$ set of eigen values

U D

A	Step 1	Step 2	Step 3
SVD	V^T (Rotation 1)	Scaling along axis (Real valued)	U (Rotation 2)
EVD	U^{-1} (Rotation)	Scaling along axes (Rotation if e.v is complex)	U (inverse of step 1)

A arbitrary → S.V.D

square → S.V.D & E.V.D may not be identical

square & symmetric → S.V.D & E.V.D same

Matrix Calculus

func^ns	Eg.	value	first der.	Sec. der.
$f: \mathbb{R} \to \mathbb{R}$	x^2	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$
$f: \mathbb{R}^d \to \mathbb{R}$	loss func^n	$\mathbb{R}$	$\mathbb{R}^d$ [Gradient]	$\mathbb{R}^{d \times d}$ [Hessian] S^d (symmetric with dim. d)
$f: \mathbb{R}^d \to \mathbb{R}^p$	N.N layer	$\mathbb{R}^p$	$\mathbb{R}^{d \times p}$ [Jacobian]	$\mathbb{R}^{d \times p \times p}$ [Higher order Tensor]

Gradient → direc^n of steepest ascent

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_d} f(x) \end{bmatrix}$$

$$\parallel$$

$$\nabla_x f(x_1, x_2, \ldots, x_d)$$

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

A vector

For matrices,

$$\nabla_A f(A) = \begin{bmatrix} \frac{\partial}{\partial a_{11}} f(A) & - & - & - \\ - & & & - \\ - & - & - & \frac{\partial}{\partial a_{mn}} f(A) \end{bmatrix}$$

$f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$

$f: \mathbb{R}^d \rightarrow R$

$\nabla_x^2 f(x)$

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & - - - & \frac{\partial^2 f(x)}{\partial x_1 \partial x_d} \\ \vdots & & \\ \frac{\partial^2 f(x)}{\partial x_d \partial x_1} & - - - & \frac{\partial^2 f(x)}{\partial x_d \partial x_d} \end{bmatrix}$$

<u>Eg</u>: b is some constant.

$\nabla_x (b^T x)$

$$= \begin{bmatrix} \vdots \\ \frac{\partial}{\partial x_i}(b^T x) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{\partial}{\partial x_i}(b_1 x_1 + \dots b_i x_i + \dots b_d x_d) \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_d \end{bmatrix} = b$$

Product Rule

$$\nabla_x x^T A x = \nabla_x x^T A x + \nabla_x x^T A x$$

$$= Ax + A^T x$$

$$= x(A + A^T)$$

$$= 2Ax \rightarrow \text{if } A \text{ is symm.}$$

$$\nabla_A \log |A| = A^{-1}$$

$$\frac{d}{dx} \log(x) = x^{-1}$$