LA FORMULA LIST

V3 & V4

Unit-3

Norm / Length

→ $x = (x_1, x_2, x_3 \ldots x_n)$

$\|x\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$

$\|x\|^2 = x^T x$

Inner Product

→ $x = (x_1, x_2 \ldots x_n)$

$y = (y_1, y_2 \ldots y_n)$

$x^T y$ (or) $x \cdot y$ (or) $\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$

↳ $x^T y = y^T x$

Orthogonal Vectors

→ x & y are orthogonal when $\langle x, y \rangle = 0$

(zero is the only vector orthogonal to itself
zero is the only vector orthogonal to every vector)

→ $x^T y = 0$ at $90°$

→ If non-zero vectors $v_1, v_2, \ldots v_n$ are mutually orthogonal, then they are linearly independant but convex need not be true

$\Sigma c_i (v_i^T v_i) = 0 \Rightarrow$ Then $c_i = 0$

Orthonormal Vectors

→ 2 vectors are orthonormal if dot product = 0 & each vector is unit length

Orthogonal Subspaces

→ 2 subspaces V & W of same R^n are orthogonal if $v^T w = 0$ for all v & w

→ Line can be orthogonal to another line
Plane can be orthogonal to another line
but plane can't be orthogonal to another plane

Fundamental Theorem of orthogonality

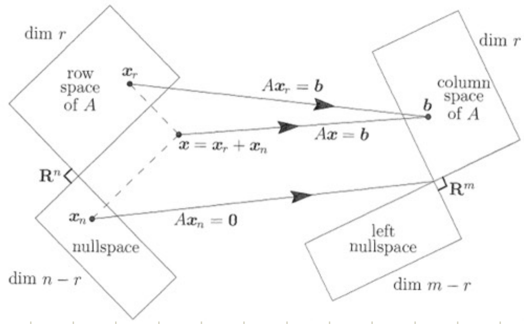
→ Row space is orthogonal to null space
Column space is orthogonal to left null space

Orthogonal Component

→ For a vector v of R^n, the space of all vectors orthogonal to v is orthogonal complement ($v^\perp$)

→ From fundamental theory of orthogonality, $N(A) = (C(A^T))^\perp$

→ $N(A) \leftrightarrow$ Orthogonal component of $C(A^T)$ (Row space)
$N(A^T) \leftrightarrow$ Orthogonal component of $C(A)$ (Column space)



Note : Let some plane be $ax + by + cz = 0$,
Then vector $\perp^{ar}$ to plane is (a, b, c)

cosine of θ

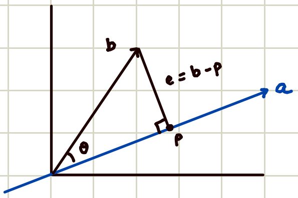
→ $\cos\theta = \dfrac{a^Tb}{||a||\,||b||}$

$\cos\theta = \cos\beta\cos\alpha + \sin\beta\sin\alpha = \dfrac{a_1b_1 + a_2b_2}{||a||\,||b||}$

→ Law of cosines $||b-a||^2 = ||b||^2 + ||a||^2 - 2||b||\,||a||\cos\theta$

If $\theta = 90°$, $||b-a||^2 = ||b||^2 + ||a||^2$

Projection onto a line



point p is the projection of b onto subspace

p must be multiple of a

i.e $p = \hat{x}a$

So, $p \perp a \Rightarrow \hat{x}a \perp a$

$e \perp a$

$b - p \perp a$

$b - \hat{x}a \perp a \Rightarrow a^T.(b - \hat{x}a) = 0$

$a^Tb - \hat{x}a^Ta = 0$

$\boxed{\hat{x} = \dfrac{a^Tb}{a^Ta}}$

& $\cos\theta = \dfrac{||p||}{||b||}$ or $\dfrac{a^Tb}{||a||\,||b||}$

Schwartz - Inequality

→ as $||e||^2 = ||b-p||^2$

$= \left|\left| b - \dfrac{a^Tb}{a^Ta}.a \right|\right|^2 = ||b||^2 + \left(\dfrac{a^Tb}{a^Ta}\right)^2 a^Ta - 2\dfrac{(a^Tb)^2}{a^Ta}$

$= b^Tb + \dfrac{(a^Tb)^2}{a^Ta} - \dfrac{2(a^Tb)^2}{a^Ta}$

$= \dfrac{(b^Tb)(a^Ta) - (a^Tb)^2}{a^Ta} \geq 0$

$\Rightarrow (b^Tb)(a^Ta) \geq (a^Tb)^2 \Rightarrow |a^Tb| \leq ||a||\,||b||$

→ The equality holds true iff $\vec{b}$ is a multiple of $\vec{a}$

Projection Matrix

→ $P = \dfrac{aa^T}{a^Ta}$

$Pb = \dfrac{aa^T}{a^Ta}b = a\dfrac{a^Tb}{a^Ta} = a\hat{x} = p$

To project b onto a, multiply by projection matrix $P : p = Pb$

→ P is symmetric matrix

$P^n = P$ for $n = 1, 2, 3 \ldots$

$\text{rank}(P) = 1$

$\text{tr}(P) = 1$

Transpose from inner products

→ $(Ax)^Ty = x^TA^Ty = x^T(A^Ty)$

$(ABx)^Ty = (Bx)^T(A^Ty) = x^T(B^TA^Ty)$

Projections & least squares

→ Failure of Gaussian elimination with multiple eq^n's & one variable

$a_1 x = b_1$

$a_2 x = b_2$ ⇒ system solvable if b is multiple of a

$a_3 x = b_3$

If system is inconsistent, then we choose a value of a which minimizes an avg. error E in m equations

sum of squares ⇒ $E^2 = \sum_{i=1}^{m} (b_i - a_i x_i)^2$

If there is exact sol^n, E = 0

so min error at $\frac{dE^2}{dx} = 0$ ⇒ $\sum_{i=1}^{m} 2(b_i - a_i x_i)\, a_i = 2\sum_{i=1}^{m} a_i^2 x - 2\sum_{i=1}^{m} a_i b_i = 0$ ⇒ $\sum_{i=1}^{m} a_i^2 x = \sum_{i=1}^{m} a_i b_i$

$a^T a (\hat{x}) = a^T b$ ⇒ $\hat{x} = \frac{a^T b}{a^T a}$

Least Squares - Multiple Variables

→ Consider an inconsistent system of linear equations

$A_{m \times n}\, x_{n \times 1} = b_{m \times 1}$

Since b lies outside C(A) & we project it onto C(A) to get p in C(A) closest to b

System reduced to $A\hat{x} = p$

$\boxed{A^T A \hat{x} = A^T b}$ ⇒ Normal Equation

→ When $Ax = b$ is inconsistent, least squares solution minimizes to $A^T A \hat{x} = A^T b$

→ When $A^T A$ is invertible exactly when columns of A are linearly independant, $\hat{x} = (A^T A)^{-1} A^T b$

Projection of b onto column space $p = A\hat{x} = A(A^T A)^{-1} A^T b$

Orthogonal bases

→ Basis consisting of mutually orthogonal vectors

Orthonormal basis

→ Basis consisting of unit length, mutually orthogonal vectors

Orthogonal matrix

→ Matrix with orthonormal columns is called Q $(m \geq n)$

→ If m = n, matrix is orthogonal

Properties of Q

1) If Q has orthonormal columns, then $Q^T Q = I$

$$\begin{bmatrix} - q_1^T - \\ - q_2^T - \\ \vdots \\ - q_n^T - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1 \end{bmatrix} = I$$

2) If an orthogonal matrix is square matrix with orthonormal columns

$Q^T = Q^{-1}$

3) If Q is a tall matrix Q^T = left inverse of Q ⇒ $Q^T Q = I$

4) Multiplication by any Q preserves length

$\|x\| = \|Qx\|$

5) Q preserves inner products and angles

$(Qx)^T (Qy) = x^T Q^T Q y = x^T y$

6) If $q_1, q_2, \dots q_n$ are orthonormal bases of $\mathbb{R}^n$, then any vector b in $\mathbb{R}^n$ can be expressed as
$b = x_1 q_1 + x_2 q_2 + \dots + x_n q_n$

$q_1^T b = x_1 q_1^T q_1 \Rightarrow x_1 = \frac{q_1^T b}{q_1^T q_1} = q_1^T b$

$x_2 = q_2^T b$ and so on

$b = (q_1^T b) q_1 + q_2^T b q_2 + \dots$
$= x_1 q_1 + x_2 q_2 + \dots$
$b = Q x \Rightarrow x = Q^{-1} b$
$x = Q^T b \quad (Q^{-1} = Q^T)$

Gram - Schmidt Orthogonalization

→ Gram - Schmidt process orthonormalizes a set of vectors in an inner product space (Euclidian space)
→ Procedure :
i) Given 3 linearly independant vectors a, b, c in some vector space
ii) Normalize the first vector
$q_1 = \frac{a}{\|a\|}$ (q_1 has unit length)

iii) Remove component of b in direction of q_1 and normalize
$B = b - (q_1^T b) q_1$
$q_2 = \frac{B}{\|B\|}$

iii) Remove components of c in direction of q_1 & q_2 and normalize
$C = c - (q_1^T c) q_1 - (q_2^T c) q_2$
$q_3 = \frac{C}{\|C\|}$

QR Factorization

→ If $A_{m \times n}$ is a matrix with linearly independant columns, then $A_{m\times n} = Q_{m\times n} R_{n\times n}$

$Q_{m\times n}$ ↓ Matrix w orthonormal vectors

$R_{n\times n}$ ↓ upper Triangular & Invertible matrix

If $A = \begin{bmatrix} | & | & | \\ a & b & c \\ | & | & | \end{bmatrix}$

Then, $a = (q_1^T a) q_1$
$b = (q_1^T b) q_1 + (q_2^T b) q_2$
$c = (q_1^T c) q_1 + (q_2^T c) q_2 + (q_3^T c) q_3$

and, $A_{m\times n} = \underbrace{\begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix}}_{Q_{m\times n}} \underbrace{\begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ 0 & q_2^T b & q_2^T c \\ 0 & 0 & q_3^T c \end{bmatrix}}_{R_{n\times n}}$

Eigen Values and Eigen Vectors

→ Till now we used some matrix A & multiply to matrix x to get matrix b $(Ax = b)$

→ Now we replace A with some scalar number λ $(\underbrace{\lambda}_{\text{eigen value}} \underbrace{x}_{\text{eigen vector of } A} = b)$

→ Solution for $Ax = \lambda x$

Then $(A - \lambda I)x = 0$

x : nullspace of $A - \lambda I$

$A - \lambda I$ must be singular

Properties

i) Eigen values of A are roots of characteristic eq^n → $(\det(A - \lambda I) = 0)$

ii) Determinant of matrix is product of eigen values

iii) Trace of matrix is sum of eigen values

iv) A^k : λ_i^k

v) A^{-1} : $\dfrac{1}{\lambda_i}$

vi) $A^T = A$

vii) If any eigen value is zero, then A is singular

viii) If $Av = \lambda v$

Then $A(cv) = c(Av) = c\lambda v = \lambda(cv)$

ix) The diagnonal elements are eigen values for upper triangular, lower triangular, diagonal matrix

$\lambda_i = a_{ii}$

→ Direct formula to find eigen values,

for 2×2 matrix, $\lambda^2 - tr(A)\lambda + \det(A) = 0$

for 3×3 matrix, $\lambda^3 - tr(A)\lambda^2 + \min\det(A)\lambda - \det(A) = 0$

$$\min\det(A) \Rightarrow \text{ex:} \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix} \Rightarrow (7\times3 - (-4\times-4)) + ((8\times3) - (2\times2)) + ((8\times7) - (-6\times-6)) = 45$$

Cayley - Hamilton Theorem

→ Every square matrix A satisfies the characteristic equation $|A - \lambda I| = 0$

Replace λ with A in polynomial, solve for A^{-1}

Diagonalization of a matrix

→ Suppose $n\times n$ matrix A has n linearly independant eigen vectors

If these eigen vectors are columns of a matrix S, then $S^{-1}AS$ is a diagonal matrix Λ

S : Eigenvector matrix

Λ : Eigenvalue matrix

$$AS = A\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

Note: If $A = S\Lambda S^{-1}$

Then $A^k = S\Lambda^k S^{-1}$

→ $AS = S\Lambda$

$S^{-1}AS = \Lambda$ (or) $A = S\Lambda S^{-1}$

→ Any matrix with distinct eigen values can be diagonalized

diagonaling matrix S isn't unique

Order of eigenvectors in S & eigenvalues is Λ is automatically same

Quadratic Forms

→ Let q be a real polynomial in variables $x_1, x_2, \dots x_n$ such that every term in q has degree 2

$q(x_1, x_2, \dots x_n) = \sum_{i=1}^{n} c_i x_i^2 + \sum_{i<j} d_{ij} x_i x_j$ where $c_i, d_{ij} \in \mathbb{R}$

↳ Quadratic form

→ If there aren't any cross-product terms $(x_i x_j)$ (or all $d_{ij} = 0$) then q is called diagonal

→ $2 \times 2 \Rightarrow a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2$

$3 \times 3 \Rightarrow a_{11} x_1^2 + a_{22} x_2^2 + a_{33} x_3^2 + 2a_{12} x_1 x_2 + 2a_{23} x_2 x_3 + 2a_{31} x_3 x_1$

Positive Definiteness

→ If a quadratic form of a matrix, the function f is allowed to vanish at $(0,0)$ when $f(x,y)$ is strictly +ve at all other points it is called positive definite

→ Matrix is strictly positive definite if

a) All eigen values > 0

b) All subdeterminants > 0

c) All pivots > 0

d) $x^T A x > 0$

→ For quadratic form,

$ax^2 + 2bxy + cy^2$ is positive definite iff $a > 0$ and $ac > b^2$

$ax^2 + 2bxy + cy^2$ is negative definite iff $a < 0$ and $ac < b^2$

$ax^2 + 2bxy + cy^2$ is positive semi-definite iff $a > 0$ and $ac = b^2$

$ax^2 + 2bxy + cy^2$ is negative semi-definite iff $a < 0$ and $ac = b^2$

Singular Value Decomposition

→ Any non-matrix A can be factored into

$A = U \Sigma V^T$

orthogonal ↙ ↓ ↳ Orthogonal

Eigen Values (Diagonal)

$U \rightarrow m \times m$ orthogonal matrix with eigen vectors of AA^T as columns

$\Sigma \rightarrow m \times n$ diagonal matrix with singular values of A as diagonal elements

$V \rightarrow n \times n$ orthogonal matrix with eigen vectors of $A^T A$ as columns

Steps

a) Short Matrix $(m < n)$

i) Find AA^T

ii) Find eigen values of AA^T

iii) Find eigen vectors x_1 & x_2

iv) Normalise $\hat{x}$

v) Find singular values $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2} \dots$ where $\lambda_1 > \lambda_2 > \dots > \lambda_m$ and form Σ with same order as A

vi) $v_i = \frac{A^T u_i}{\sigma_i}$ or $v_i^T = \frac{u_i^T A}{\sigma_i}$. Find v_1 & v_2

vii) Find v_3 by orthogonality $v_1 \perp v_3$ & $v_2 \perp v_3$

viii) Write $A = U \Sigma V^T$

a) Long Matrix $(m > n)$

i) Find $A^T A$

ii) Find eigen values of $A^T A$

iii) Find eigen vectors x_1 & x_2

iv) Normalise $\hat{x}$

v) Find singular values $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2} \dots$ where $\lambda_1 > \lambda_2 > \dots > \lambda_m$ and form Σ with same order as A

vi) $u_i = \frac{A v_i}{\sigma_i}$. Find u_1 & u_2

vii) Find u_3 by orthogonality $u_1 \perp u_3$ & $u_2 \perp u_3$

viii) Write $A = U \Sigma V^T$

Pseudo - Inverse

→ If A is a matrix, then pseudoinverse A^+ is a generalization of the inverse (Follows SVD)

→ It satisfies $AA^+A = A$

$A^+AA^+ = A^+$

$(AA^+)^T = AA^+$

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

Steps

i) Compute $A = U \Sigma V^T$

ii) Take reciprocals of non-zero values in Σ ex: $(3, 2, 0) \to (1/3, 1/2, 0)$

iii) Compute $A^+ = V \Sigma^+ U^T$

Principal Component Analysis (PCA)

→ Statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components

Steps

i) Start with data matrix X

ii) Subtract mean from each row $\bar{X} = X - \text{mean}(X)$

iii) $\bar{X} = U \Sigma V^T$

iv) Columns of U are principal directions (principal components)

v) Squares of singular values σ_i^2 give variance for each component

vi) $Z = U^T \bar{X}$ ⇒ Coordinates in new principal component space

Image Processing

→ Grayscale Images - 2x2 Matrix

Color Images - 3x3 Matrix

→ 3x3 Averaging (blur) filter:

$K = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ ⇒ Filter computes average of a pixel & its 8 neighbours (only for interior pixels)

ex:

$$\begin{bmatrix} 52 & 55 & 61 & 59 & 79 \\ 62 & 59 & 55 & 104 & 94 \\ 63 & 65 & 66 & 113 & 144 \\ 64 & 70 & 70 & 126 & 154 \\ 69 & 73 & 74 & 139 & 161 \end{bmatrix} \longrightarrow \begin{bmatrix} 52 & 55 & 61 & 59 & 79 \\ 62 & 63 & 75 & 94 & 94 \\ 63 & 68 & 81 & 110 & 144 \\ 64 & 74 & 91 & 121 & 154 \\ 69 & 73 & 74 & 139 & 161 \end{bmatrix}$$

$$\frac{(59 + 55 + 104 + 65 + 66 + 113 + 70 + 70 + 126)}{9} = 81$$

→ Sharping Kernel (Laplace variant)

$K = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$

Minimising Multivariable function

→ Energy is a positive definite quadratic function & derivative of a quadratic is linear & we set these derivatives to 0 which makes us reach point of minima

Steps

i) Convert A & B into P (quadratic form) with $\frac{1}{2}x^T A x - x^T b$

ii) Now perform partial differentiation wrt all the variables in x

iii) Using these equations find solution / critical points
(You can only get solution when A is positive definite)

Backpropagation

→ Algorithm to train neural networks by adjusting weights based on error in predictions

Stochiatic Gradient Descent (SGD)

→ Optimizing algorithm that updates the parameters by following negative gradient direction to minimize the loss function

Steps

i) Given x, w, b

ii) Calculate $y = w^T x + b$

iii) Compute $\hat{y} = \sigma(y)$ (Sigmoid function) $= \frac{1}{1+e^{-y}}$ by substituting y

iv) Compute Loss $L(\hat{y}, t) = \frac{1}{2}(\hat{y} - t)^2$ (t will be given)

v) Calculate $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial y} \cdot \frac{\partial y}{\partial w} = (\hat{y} - t) \cdot (\sigma(y)(1-\sigma(y))) \cdot x = (\hat{y}-t) \cdot \hat{y}(1-\hat{y}) \cdot x$

$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial b} = (\hat{y} - t) \cdot \sigma(y)(1-\sigma(y)) = (\hat{y} - t)\hat{y}(1-\hat{y})$

vi) Finally update weights (η will be given)

$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w_{old}}$

x : Input vector

w : Initial weights

b : bias

t : true label

η : learning rate

~~PI - 370~~

~~PCA - 396~~

~~Image Process - 433~~

~~Minimizing multiv - 444~~

~~Stochastic Gradient Descent - 467~~