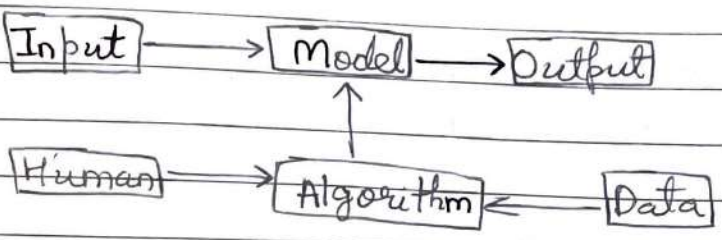
Date: / / Page no:

MLF

ML → Experience E, (Learn)
Class of Problems T, ← (with respect to)
Performance, P ← (improve)



* Train Data
* Test Data
* Validation Data.

Classification : $f = sign(w^T x + b)$ [Discrete]
$L = \frac{1}{n} \sum_i 1(f(x^i) \neq y^i)$ [Sup.]

Regression : $f = w^T x + b$ [Continuous]
$L = \frac{1}{n} \sum_i (f(x^i) - y^i)^2$ [Sup.]

Dimensionality Reduction : $f(x)$, $g(u)$ [Unsup.]
$L = \frac{1}{n} \sum_i \| g(f(x^i)) - x^i \|^2$

Density Estimation : $P(x)$ [Unsup.]
$L = \frac{1}{n} \sum_i -\log(P(x^i))$

$\boxed{Loss[f] < Loss[g]} \Rightarrow f$ is better.

Loss is aka reconstruction error.

Date: / / Page no:

Continuity : $\lim_{x \to a^-} f(x) = f(a) = \lim_{x \to a^+} f(x)$

Differentiability : $\lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} = f'(a) = \lim_{x \to a^+} \frac{f(x) - f(a)}{x - a}$

Linear approximation : $L(x) = f(a) + f'(a)(x - a)$

n^{th} order approximation :

$$L(x) = f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \cdots$$

Approximations for multivariate functions :

a) Linear : $L(x, y) = f(a, b) + \left.\frac{\partial f}{\partial x}\right|_{(a,b)} (x - a) + \left.\frac{\partial f}{\partial y}\right|_{(a,b)} (y - b)$

b) Higher : $f(x) = f(v) + \nabla f(v)^T (x - v) + \frac{1}{2}(x - v)^T \nabla^2 f(v)(x - v)$

Gradient : $\nabla f = \left[\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right]$

Directional derivative : $D_{\vec{u}} f(x, y) = \nabla f \cdot u$

$$= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y}$$

Cauchy - Schwarz Inequality :

$$-\|a\| \cdot \|b\| \leq a^T b \leq \|a\| \cdot \|b\|$$

Date: / / Page no:

Martix : $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ ← x_1, ← x_2, ← x_m (Rows)

↑ u_1 ↑ u_2 ↑ u_n (Columns)

Column Space : $C(A) = \text{span}(u_1, u_2, \cdots, u_n)$

Row Space : $C(A^T) = \text{span}(x_1, x_2, \cdots, x_n)$

Null Space : $N(A) = \{x \mid Ax = 0\}$

Left Null Space : $N(A^T) = \{y \mid A^T y = 0\}$

Length of vector : $|v| = \sqrt{v_1^2 + v_2^2 + v_3^2}$

Dot / Inner Product : $u \cdot v = u_1 v_1 + u_2 v_2 + u_3 v_3$

Rank - Nullity Theorem : Rank + Nullity = # Columns

Projections :

a) Proj. Matrix : $P = \frac{aa^T}{a^T a}$.

b) Proj of b onto a : $p = Pb = \frac{aa^T}{a^T a} b$

c) Projection of a along b : $q = \frac{a \cdot b}{|b|} \times b$

d) Error Matrix : $e = b - p$

Orthogonality : $a \cdot b = 0$ [dot product = 0]

Least Squares Solution : $Ax = b$

(i) Find A^TA and A^Tb
(ii) Find aug. matrix for $A^TAx = A^Tb$
(iii) Row Reduction
(iv) Solution $\hat{x}$ is LSS.

Linear Regression : Goal ⇒ minimize L

$$L(\theta) = \frac{1}{2} \sum (x_i^T \theta - y_i)^2$$

$$A = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

feature matrix

$$L(\theta) = \frac{1}{2}(A\theta - Y)^T(A\theta - Y)$$

LSS ⇒ $(A^TA)\theta = A^TY$ or $\theta = (A^TA^{-1})A^TY$

Maximum Likelihood :

$$L(\theta) = \prod_i \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}(y_i - \theta^T x_i)^2\right)$$

Date: / / Page no:

Polynomial Regression : $\hat{y}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_m x^m$

$\hat{y}(x) = \theta^T \phi(x)$, $\phi(x) = (1, x, x^2, \cdots, x^m)$

$$A = \begin{bmatrix} \phi(x_1)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

LSS $\longrightarrow$ $\boxed{(A^TA)\theta = A^TY}$

Regularised Version (Ridge Regression) :

$$L(\theta) = \frac{1}{2} \sum_i (x_i^T \theta - y_i)^2 + \lambda \|\theta\|^2$$

$$\theta_{reg} = (A^TA + \lambda I)^{-1} A^TY$$

Eigen Values and Eigen Vectors : $Ax = \lambda x$

$|A - \lambda I| = 0$ ← Ch. eqn → $\lambda \Rightarrow$ eigen values

$(A - \lambda I)x = 0$ ← For each λ, find x ← eigen vector

Characteristic equations :

(i) 2x2 : $\lambda^2 - tr(A)\lambda + det(A) = 0$

(ii) 3x3 : $\lambda^3 - tr(A)\lambda^2 + (M_{11} + M_{22} + M_{33})\lambda - det(A) = 0$

Date: / / Page no:

Properties of Eigenvalues & Eigenvectors :

a) $\lambda_1 + \lambda_2 = tr(A)$

b) $\lambda_1 \cdot \lambda_2 = |A|$ of A

c) If $\lambda_1, \cdots, \lambda_m$ are E.V^ then $\lambda_1^k, \lambda_2^k, \cdots, \lambda_m^k$ are EV of A^k

d) If $\lambda_1, \cdots, \lambda_m$ are EV of A then $\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_m}$ are of A^{-1}

e) Symmetric matrices $\Rightarrow$ Real eigenvalues

f) Diagonal matrices $\Rightarrow$ eigenvalues are diagonal elements

g) UTM / LTM $\Rightarrow$ eigenvalues are diagonal elements

h) EV of A = EV of A^T

Diagonalisation : $\Lambda = S^{-1}AS$

a) $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$ $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues

b) $S = \begin{bmatrix} | & | & | \\ x_1 & x_2 & x_3 \\ | & | & | \end{bmatrix}$ x_1, x_2, x_3 are eigenvectors

c) $\Lambda^k = S^{-1}A^kS$

eigenvalues of similar matrices are same

Λ and A is similar matrices

Date: / / Page no:

Fibonacci Series : $0, 1, 1, 2, 3, 5, \dots \{F_K\}$

$$\boxed{F_K = F_{K-1} + F_{K-2}}$$

$$F_K = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^K - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^K$$

Orthogonally Diagonalizable Matrices :

a) Find eigenvalues & eigenvectors
b) Normalize eigenvectors
c) Rest follow as diagonalisation

$$A = Q \wedge Q^{-1} \quad \text{or} \quad A = Q \wedge Q^T$$

$$Q^T Q = I$$

Complex Matrices : $\mathbb{C}^n$: Complex numbers

$a+ib : a, b \in \mathbb{R}, \; i = \sqrt{-1} \Rightarrow i^2 = -1$

a) Conjugate : $\overline{(a+ib)} = (a-ib)$

b) Inner product : $\bar{x}^T y = x \cdot y$

$x \cdot y = \overline{y \cdot x}$
$x(cy) = c(x \cdot y)$
$(cx) \cdot y = \bar{c}(x \cdot y)$

c) Length : $\|x\|^2 = \bar{x}^T x$

d) Conjugate Transpose : $A^* = (\bar{A})^T = \overline{(A^T)}$

$(A^*)^* = A$
$(AB)^* = B^* A^*$

Complex Matrices :

a) Hermitian : $A^* = A$

b) Skew-Hermitian : $A^* = -A$

Properties : (i) All eigenvalues of HM are real

(ii) Eigenvectors corresponding to different eigenvalues are orthogonal.

(iii) All symmetric matrix are Hermitian

(iv) If EV are distinct, then matrix is diagonalizable

c) Unitary Matrices : $U^* U = U U^* = I \Rightarrow U^* = U^{-1}$

Properties : (i) $\|Ux\| = \|x\|$

(ii) eigenvalues $\rightarrow \pm 1$

(iii) eigenvectors corresponding to different eigenvalues are orthogonal

Diagonalization : $A = U \lambda U^*$

To show HM is unitarily diagonalizable we show that it is similar to an UTM.

Schur's Theorem : Any matrix $A_{n \times n}$ is similar to UTM

i.e. $A = UTU^*$

Spectral Theorem : A HM 'A' is unitarily diagonalizable

$$\boxed{U^* A U = D}$$

For symmetric matrix,

$Q^TAQ = D$, $Q^TQ = I$ ⟶ Orthogonally diagonalizable

- HM ⟶ ⟵ Unitarily diagonalizable

Quadratic Form : $f(x) = ax^2 + 2bxy + cy^2$

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

OR,

$$f(x,y) = v^TAv, \quad v = \begin{bmatrix} x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

Partial Derivatives :

a) 1st order : $f_x = \frac{\partial f}{\partial x}$, $f_y = \frac{\partial f}{\partial y}$

b) 2nd order : $f_{xx} = \frac{\partial^2 f}{\partial x^2}$, $f_{xy} = \frac{\partial^2 f}{\partial x \partial y}$

$$f_{yy} = \frac{\partial^2 f}{\partial y^2}, \quad f_{yx} = \frac{\partial^2 f}{\partial y \partial x}$$

Date: / / Page no:

Maxima & Minima : At (p,q) if 1st order derivative = 0, then its stationary pt.

$$D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix} = f_{xx} f_{yy} - f_{xy}^2$$

a) minima $\Rightarrow f_{xx} > 0$, $D > 0$

b) maxima $\Rightarrow f_{xx} < 0$, $D > 0$

c) saddle $\Rightarrow D < 0$

d) inconclusive $\Rightarrow D = 0$

Definiteness : $f = v^T A v$, $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$, $v = \begin{bmatrix} x \\ y \end{bmatrix}$

+ve	$f > 0$	$a > 0$, $ac - b^2 > 0$	EV > 0
-ve	$f < 0$	$a < 0$, $ac - b^2 < 0$	EV < 0
+ve semi	$f \geq 0$	$a > 0$, $ac - b^2 = 0$	EV ≥ 0
-ve semi	$f \leq 0$	$a < 0$, $ac - b^2 = 0$	EV ≤ 0
indefinite	$f > 0$, $f < 0$	$ac - b^2 < 0$	EV > 0, EV < 0

Date: / / Page no:

Singular Value Decomposition (SVD):

- Given A,

a) Find A^TA

b) Find eigenvalues of A^TA & sort them in descending order of their absolute values. Find singular values by taking square roots of eigenvalues.

$$\boxed{\sigma_i = \sqrt{\lambda_i}}$$

c) Construct S by placing σ_i's in descending order along its diagonal. Calculate S^{-1}

$$S = \begin{bmatrix} \sigma_i & & 0 \\ & \searrow & \\ 0 & & \end{bmatrix}$$

d) Use eigenvalues & compute eigenvectors of A^TA. Place these along columns of V and find V^T.

$$V = \begin{bmatrix} | & | \\ x_1 & x_2 \\ | & | \end{bmatrix}$$

x_1 : eigenvector of λ_1
x_2 : eigenvector of λ_2

e) Find $U = AVS^{-1}$

Result : $\boxed{A = USV^T}$

Date: / / Page no:

Principal Component Analysis (PCA):

a) $X = [x_1 \cdots x_n] \qquad x_i = \begin{bmatrix} fe_i \\ \vdots \\ fe_m \end{bmatrix}$

b) $\bar{X} = \sum \frac{1}{n} x_i$ (Mean vector)

c) $X - \bar{X} = [x_1 - \bar{x}, \cdots, x_n - \bar{x}]$

d) $C = \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})^T$ (cov. matrix) → symm(m×m)

e) Find eigenvalues & eigenvectors of C

$$\Lambda = [\lambda_1 \ \lambda_2 \cdots \lambda_n], \quad U = [u_1, \cdots, u_n]$$

f) Find transformed data points

$$\tilde{x}_i = \sum \alpha_j u_j, \quad \alpha_j = x_i^T u_j$$

g) Reconstruction Error: $J = \frac{1}{n} \sum \| x_i - \tilde{x}_i \|^2$

h) Projected Variance: $\lambda_1 + \lambda_2 + \cdots + \lambda_k$

Gradient Descent : Given $f(x)$, x_0, n, ϵ

$$x_n = x_{n-1} - \epsilon f'(x_{n-1})$$

x_0 = initial guess
n = # iterations
ϵ = step.

For multivariate, $f(x,y)$, (x_0, y_0), n, ϵ

a) $\nabla f = \begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \end{bmatrix}$

b) $x_n = x_{n-1} - \epsilon \nabla f$

Newton's Method : Given $f(x)$, n

n = # iterations

$$x_n = x_{n-1} - \frac{f'(x_{-1})}{f''(x_{-1})}$$

Date: / / Page no:

Taylor's Series : $f(x+\eta d) = f(x) + \eta d f'(x) + \frac{1}{2}\eta^2 d^2 f''(x) + \cdots$

→ G.D : $x_{t+1} = x_t + \eta d_t$, $d = -f'(x_t)$

$$f(x+\eta d) \sim f(x) + \eta d f'(x) \quad \text{(for small } \eta\text{)}$$

(d: descent direction)

$\Rightarrow d f'(x) < 0$ [$\because d = -f'(x)$]

For higher dimensions :

Taylor series : $f(x+\eta d) \sim f(x) + \eta d^T \nabla f(x)$, $d = -\nabla f(x)$

($d^T \nabla f(x)$: small; $\nabla f(x)$: steepest ascent; $d^T \nabla f(x) < 0$)

Location vector : $(t+1)$ step along dir^n, d,

$$x_{t+1} = x_t + \eta d^T \nabla f(x_t), \quad d = -\nabla f(x_t)$$

Constrained Optimization :

a) Minimize $f(x)$ s.t. $g(x) \leq 0$

→ (i) Descent direction : $d^T \nabla f(x) < 0$
(ii) Feasible direction : $g(x) \leq 0$

b) If x^* is optimal then,

→ (i) $g(x^*) \leq 0$
(ii) No descent direction is feasible.

Date: / / Page no:

Optimality : Minimize $f(x)$ s.t. $g(x) \leq 0$

a) **necessary** : at x^*, $d^T \nabla g(x^*) = 0$
b) **optimal config.** : $\nabla f(x^*) = -\lambda \nabla g(x^*)$

Method of Lagrange Multipliers :

a) **Minimize** : $f(x)$ under $g(x) = 0$
b) **necessary** : at x^*, $d^T \nabla g(x^*) = 0$
c) **optimal config.** : $\nabla f(x^*) = \lambda \nabla g(x^*)$
(L.M)

Convexity : $S \subseteq \mathbb{R}^d$ is convex if
$\forall x_1, x_2 \in S, \ \lambda x_1 + (1-\lambda) x_2 \in S, \ \forall \lambda \in [0,1]$

Property : If $S_1, S_2 \subseteq \mathbb{R}^d$, then $S_{12} = S_1 \cap S_2 = \{x : x \in S_1, x \in S_2\}$
Intersection of convex sets is convex.

Convex Hull : $S = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}^d$
$z = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n$

z is a convex combination of S if $\exists \lambda_1, \lambda_2 \dots, \lambda_n$
s.t. $\lambda_i \geq 0, \ \sum_i \lambda = 1$

$$CH(S) = \{z : z = \sum_i \lambda_i x_i, \ \exists \ x_i \in S, \ \lambda_i \geqslant 0, \ \sum_i \lambda_i = 1\}$$

- Convex Hull is a convex set.

Date: / / Page no:

Convex Functions : a) A fxn f is convex if epi(f) is convex set.

$$epi(f) = \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^d : z \geq f(x) \right\}$$

b) f is convex if $f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$ holds $\forall\ \lambda \in [0,1]$

c) $f : \mathbb{R}^d \to \mathbb{R}$ is convex iff for a point y :

differentiable

$$f(y) \geq f(x) + (y-x)^T \nabla f(x)$$

d) $f : \mathbb{R}^d \to \mathbb{R}$ → $\det(H) \geq 0$, $H \Rightarrow$ Hessian matrix

$EV(H) \geq 0$ → [Positive / Positive semi-definite]

e) For convex fxn, if x^* is local minima and z is global minima then,

$$f(z) = f(x^*) \leq f(x),\ \forall\ x \in [x^* - \delta, x^* + \delta],\ \delta > 0.$$

Linear Programming (LPP) : Constraints → Linear
Objective → Linear.

Karush-Kuhn-Tucker Conditions (KKT) :

a) Stationarity : $\nabla f(x) + \sum_i^n u_i \nabla g(x) + \sum_j^m v_j \nabla h(x) = 0$

b) Complementary slackness : $u_i g_i = 0\ \forall i$

c) Primal feasibility : $g_i(x) \leq 0\ \forall i$

d) Dual feasibility : $u_i \geq 0\ \forall i$

Date: / / Page no:

Types of possible solutions for LPP :

a) Infeasible
b) Unbounded
c) Feasible

max $\xrightarrow{\text{dual}}$ min
$\xleftarrow{\text{dual}}$

Weak duality : $g(\lambda^*) \leq f(x^*)$
Strong duality : $g(\lambda^*) = f(x^*)$

Dual \ Primal	Finite optimal	Unbounded	Feasible
Finite optimal	✓	X	X
Unbounded	X	X	✓
Feasible	X	✓	✓

Sample Space : Ω = set of outcomes

Events : $E \subseteq \Omega$, $F = P(\Omega)$

Probability : (Ω, F, P) , $P : F \rightarrow [0,1]$

Axioms : a) $P(\Omega) = 1$
b) $P(A) \geq 0$
c) $A_1, A_2, \dots, A_n$ are disjoint
$P(A_1 \cup A_2 \dots, A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$

Properties : a) $P(A^c) = 1 - P(A)$
b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
c) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B)$
$- P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

Date: / / Page no:

Conditional Probability : $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Total Probability : $B_1, B_2, \dots, B_n$ are mut. exc & exh.

$$\forall_{\substack{i,j \\ i \neq j}} \; B_i \cap B_j = \phi$$

$$B_1 \cup B_2 \cup \dots \cup B_n = \Omega$$

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$
$$= P(A|B_1)\,P(B_1) + \dots + P(A|B_n)\,P(B_n)$$

Independent events : $P(A \cap B) = P(A) \cdot P(B)$

Bayes's Theorem : $P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\bar{A})\,P(\bar{A})}$

Probability Mass fxn : $X : \Omega \to \mathbb{R}$ (Random variable)

$f_X : \mathbb{R} \to [0,1]$

$$f_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

a) $f_X(x) \in [0,1]$

b) $\sum_x f_X(x) = 1$

c) $\sum_x f_X(x) = \sum_x P(\{\omega \in \Omega : X(\omega) = x\}) = P(\Omega) = 1$

Cumulative Distributive fxn : $F_X(x) = P(X \leq x)$

a) $F_X(x) \in [0,1]$, b) $F_X(-\infty) = 0$, c) $F_X(\infty) = 1$

Date: / / Page no:

Expectation : $E[X] = \sum_x x f_X(x)$

Conditional Expectation : $E[X|A] = \sum_x x f_{X|A}(x)$

Linearity : a) $E[X+Y] = E[X] + E[Y]$

b) $E[aX] = aE[X]$

Variance : $Var[X] = E[(X-EX)^2] = E[X^2] - (EX)^2$

Joint Distributions : $f_{XY}(x,y) = P(X=x, Y=y)$

$$f_{X,Y}(x,y) = P(\{\omega \in \Omega : X(\omega)=x\} \cap \{\omega \in \Omega : Y(\omega)=y\})$$

$$F_{XY}(x,y) = P(X \leq x, Y \leq y)$$

Marginal & Conditional Distributions :

$$f_X(x) = P(X=x) = \sum_y f_{XY}(x,y)$$

$$f_Y(y) = P(Y=y) = \sum_x f_{XY}(x,y)$$

$$f_{X|Y}(x|y) = P(X=x \mid Y=y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Independent Random Variables :

$$f_{XY}(x,y) = f_X(x)\, f_Y(y)$$

$$\Updownarrow$$

$\forall a,b$ $\{X=a\}$ & $\{Y=b\}$ are independent

$$\Updownarrow$$

$$\forall g,h \quad E[g(x) \cdot h(y)] = E[g(x)] \cdot E[h(y)]$$

Sum of Random Variables : $Z = X+Y$

$$f_Z(z) = P(Z=z) = \sum_x f_X(x)\, f_Y(z-x)$$

Conditional Expectation : $X : \Omega \to \mathbb{R}$

$A \subseteq \Omega$

$E[X|A]$ defined

$Y : \Omega \to \mathbb{R}$

$E[X|Y]$ can be viewed as fxn of Y.

Covariance :

a) $Cov[X,Y] = E[(X-EX)(Y-EY)]$

b) $cov[X,X] = var[X]$

c) $cov[X,Y] = 0 \iff X, Y$ uncorrelated.

d) $cov[X,Y] = E[XY] - EX \cdot EY$

If X, Y independent, $cov[X,Y] = 0$

Bernoulli Distribution : $f(1) = p$

$f(0) = 1-p$

$X \sim Bernoulli(p)$

$E[X] = p$

$Var[X] = p(1-p)$

Binomial Distribution : $Bin(n,P)$

$X_1, X_2, \dots, X_n$ are independent $Bern(p)$ RVs

$X \sim Bin(n,P)$

$$f_X(R) = P(X=R) = {}^nC_R\, p^R (1-p)^{n-R}$$

$$\sum_{R=0}^{n} f(R) = 1 \quad , \quad E[X] = np$$

$Var[X] = npq$, $q = 1-p$

Poisson Distribution : $X \sim Poisson(\lambda)$

$$f_X(R) = P(X=R) = \frac{e^{-\lambda} \lambda^R}{R!}$$

$$\sum_{R=0}^{\infty} f_X(R) = 1 \quad , \quad E[X] = \lambda$$

Geometric Distribution : $X \sim Geom(P)$

$$f_X(R) = P(X=R) = (1-P)^{R-1}P$$

$$\sum_{R=1}^{\infty} (1-P)^{R-1}P = 1 \quad , \quad E[X] = \frac{1}{P}$$

Variance Properties : $Var[X] = EX^2 - (EX)^2$

a) $Var[aX] = a^2 Var[X]$

b) $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$

Continuous Random Variables :

a) PDF : $f_X(x) = \frac{P(X \in [x, x+dx])}{dx}$

b) CDF : $F_X(x) = P(X \leq x)$

Properties :

(i) $f_X(x) \geq 0$

(ii) $F_X(-\infty) = 0$

(iii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

(iv) $F_X(\infty) = 1$

(v) F_X is increasing.

c) Conditional PDF : $f_{X|A}(x) = \frac{P(X \in [x, x+dx] \mid A)}{dx}$

Expectation & Variance :

a) $E[X] = \int_{-\infty}^{\infty} x \cdot f_x(x)\, dx$

(i) $E[X+Y] = EX + EY$

(ii) $Y = g(X)$, $E[Y] = \int_{-\infty}^{\infty} g(x) \cdot f_x(x)\, dx$

b) $Var[X] = E[(X-EX)^2] = EX^2 - (EX)^2$

c) $SD[X] = \sqrt{Var[X]}$

(i) $Var[X+Y] \neq Var[X] + Var[Y]$

(ii) $Var[aX] = a^2\, Var[X]$

(iii) $Var[X] \geq 0$

Conditional & Total Expectation :

a) $E[X|A] = \int_{-\infty}^{\infty} x \cdot f_{X|A}(x)\, dx$

b) $E[X] = E[X|A]\, P(A) + E[X|A^c]\, P(A^c)$

Density Functions : $f_{X,Y} = \frac{P(X \in [x, x+dx],\ Y \in [y, y+dy])}{dx\, dy}$

$F_{X,Y}(x,y) = P(X \leq x,\ Y \leq y)$

(i) $f_{X,Y}(x,y) \geq 0$	(i) $F_{XY}(-\infty, -\infty) = 0$
(ii) $\int_x \int_y f_{XY}(x,y)\, dx\, dy = 1$	(ii) $F_{XY}(\infty, \infty) = 1$

Marginals & Conditionals :

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Sum of Random Variables : $Z = X + Y$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\,dx$$

(convolution operation)

Maximum of Independent R.Vs :

$$Z = \max(X, Y)$$

$$F_Z(z) = P(Z \le z) = F_X(z) \cdot F_Y(z)$$

Minimum of Independent RVs :

$$Z = \min(X, Y)$$

$$F_Z(z) = P(Z \le z) = 1 - (1 - F_X(z)) \cdot (1 - F_Y(z))$$

Covariance & Correlation :

a) $cov[X, Y] = E[XY] - (EX)(EY)$

b) $\rho[X, Y] = \dfrac{cov[XY]}{\sqrt{var[X] \cdot var[Y]}}$

* Independent $\Rightarrow$ uncorrelated

c) $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ ~~$cov[X] =$~~ $\begin{bmatrix} var[x_1] & cov[x_1, x_2] & \cdots \\ \vdots & & \\ & & \end{bmatrix}$

$$cov[X] = \begin{bmatrix} var[x_1] & cov[x_1, x_2] & \cdots & cov[x_1, x_n] \\ \vdots & & & \\ cov[x_n, x_1] & \cdots & & var[x_n] \end{bmatrix}$$

Uniform Distribution : $X \sim Unif(a, b)$

$$f_x(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \frac{a+b}{2}, \quad Var[X] = \frac{(b-a)^2}{2}$$

Exponential Distribution : $X \sim exp(\lambda)$

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_x(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

$$E[X] = \frac{1}{\lambda} \quad , \quad Var[X] = \frac{1}{\lambda^2}$$

Normal Distribution : $Z \sim N(0,1)$

$$f_z(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right)$$

$$A = \int_{-\infty}^{\infty} e^{-x^2/2} dx \Rightarrow A^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} \, r \, dr \, d\theta$$

$$A^2 = 2\pi \Rightarrow A = \sqrt{2\pi}$$

Gaussian Distribution : $Z \sim N(0,1)$

$$X = \sigma Z + \mu \quad , \quad Z = \frac{X - \mu}{\sigma}$$

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$E[X] = \mu \quad , \quad Var[X] = \sigma^2$$

Standard Normal Vector :

$Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, ---, $Z_d \sim N(0,1)$

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_d \end{bmatrix}$$

$$f_Z(z) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_i^2\right) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|z\|^2\right)$$

$$X = \underbrace{\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}}_{A} Z \qquad X_1 = Z_1$$

$$X_2 = \rho Z_1 + \sqrt{1-\rho^2}\, Z_2$$

$$Z = \underbrace{\begin{bmatrix} 1 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}}_{A^{-1}} X$$

$$\det(A) = \sqrt{1-\rho^2}, \quad \det(A^{-1}) = \frac{1}{\sqrt{1-\rho^2}}$$

$E[X_1] = E[X_2] = 0$

$cov[X_1, X_2] = \rho$

$var[X_1] = 1$, $var[X_2] = 1$, $E[X] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$cov[X] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = AA^T, \quad \det(\Sigma) = 1-\rho^2$$

$$[cov(X)]^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}, \quad \det(\Sigma^{-1}) = \frac{1}{1-\rho^2}$$

$$f_x(x) = f_z(A^{-1}x)\,|Det(A^{-1})|$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x_2-\rho x_1)^2\right)$$

$$= f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1)$$

$X_1 \sim N(0,1)$

$X_2|(X_1=x_1) \sim N(\rho x_1, 1-\rho^2)$

1) Bivariate Normal : $X = AZ$

$$f_x(x) = f_z(A^{-1}x)\,|Det(A^{-1})|$$

$$\Sigma = AA^T = E[XX^T] = cov[X]$$

$$Det(A^{-1}) = \frac{1}{\sqrt{Def(\Sigma)}}, \quad f_x(x) = \frac{1}{2\pi\sqrt{det(\Sigma)}} \exp\left(\frac{-1}{2}x^T\Sigma^{-1}x\right)$$

$$\Sigma = \begin{bmatrix} a^2 & \rho ab \\ \rho ab & b^2 \end{bmatrix}$$

$$X = AZ + \mu$$

$$Z = A^{-1}(X-\mu)$$

$$f_x(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Date: / / Page no:

2 Multivariate Normal :

$$X = AZ + \mu$$

$$Z = A^{-1}(X - \mu)$$

$$\Sigma = AA^T$$

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$X \sim N(\mu, \Sigma)$$

Properties : a) $Y = a^T X \sim N(a^T\mu, a^T\Sigma a)$

b) $Y = AX \sim N(A\mu, A\Sigma A^T)$

c) X_i, X_j are independent $\Leftrightarrow \Sigma_{i,j} = 0$

Parameter Estimation & Max. Likelihood :

$$P = \{P_\theta : \theta \in \Theta\}$$

$X_1, X_2, \cdots X_n$ are iid from P_θ for $\theta \in \Theta$

$$L(\theta) = P(X_1 = x_1, \cdots X_n = x_n | \theta)$$

$$= \prod_{i=1}^{n} f_{X_i}(x_i|\theta) = \prod_{i=1}^{n} P_\theta(x_i)$$

$$\log(L(\theta)) = \sum_{i=1}^{n} \log(P_\theta(x_i))$$

$$R(\theta) = -\log(L(\theta))$$

(04) Bern

P_θ

$R($

$\hat{\theta}$

(05) Unif

Bernoulli Bias : $P = \{ Bern(\theta) : \theta \in [0,1] \}$

$$x_1, x_2, \dots, x_n \in \{0,1\}$$

$$P_\theta(x) = \begin{cases} \theta & x=1 \\ 1-\theta & x=0 \end{cases} = \theta^x (1-\theta)^{1-x}$$

$$R(\theta) = -\sum_{i=1}^{n} \log(P_\theta(x_i)) = a \log \frac{1}{\theta} + (n-a) \log \frac{1}{1-\theta}$$

($a = \sum x_i$)

$$\hat{\theta}_{ML} = \frac{a}{n} = \frac{\sum x_i}{n}$$

Uniform Limits : $P = \{ Unif(a,b) : a, b \in \mathbb{R} \}$

$x_1, x_2, \dots, x_n$

$$P_\theta(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases} = \frac{1}{b-a} 1(x \in [a,b])$$

$$R(\theta) = -\sum_{i=1}^{n} \log(P_\theta(x_i))$$

$$= \sum_{i=1}^{n} -\log\left(\frac{1}{b-a}\right) - \log(1(x_i \in [a,b]))$$

If $a < \min(x_i)$ & $b > \max(x_i)$

$$R(\theta) = n \log(b-a)$$

($\max x_i$, $\min x_i$)

Date: / / Page no:

Normal Mean & Variance : a) $P = \{ N(\mu, I) : \mu \in \mathbb{R} \}$

$$P_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right)$$

$$R(\theta) = \sum_{i=1}^{n} \frac{1}{2}(x_i - \mu)^2 + C$$

$$\mu = \frac{1}{n} \sum x_i$$

b) $P = \{ N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+ \}$

$$P_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \frac{(x_i - \mu)^2}{\sigma^2}\right)$$

$$R(\theta) = \frac{n}{2} \log(\sigma^2) + \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}$$

$$\mu = \frac{1}{n} \sum x_i$$

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{n}$$

Multivariate Normal : $P = \{ N(\mu, \Sigma) : \mu \in \mathbb{R}^d, \Sigma \in S_d^+ \}$

$$\mu = \frac{1}{N} \sum x_i$$

$$\Sigma = \frac{1}{N} \sum (x_i - \mu)(x_i - \mu)^T$$

Linear Regression with Gaussian Noise :

$X \in \mathbb{R}^d, \; Y \in \mathbb{R}$

$$Y = w^T X + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$P_{Y|X} = \{ N(w^T X, \sigma^2) : w \in \mathbb{R}^d \}$$

$$P(Y_1 = y_1, \dots, y_n \mid X_1 = x_1, \dots, x_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(y_i - w^T x_i)^2\right)$$

$$R(w) = \sum_{i=1}^{n} \frac{1}{2\sigma^2}(y_i - w^T x_i)^2 + c$$

Gaussian Mixture Model :

$$f_X(x) = \sum_{R=1}^{K} \pi_R \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu_R)^T \Sigma_R^{-1} (x-\mu_R)\right)$$

$$P(X) = \sum_{R=1}^{k} \pi_R N(x \mid \mu_R, \Sigma_R)$$

Max. Likelihood estimation :

$$P(Data) = \prod_{i=1}^{N} \left(\sum_{R=1}^{K} \pi_R N(x_i \mid \mu_R, \Sigma_R) \right)$$

Inequalities and Laws :

a) Markov Inequality : $P(X \geq t) \leq \mu / t$

$$EX := \int_0^\infty x f_X(x)\, dx \geq t \int_t^\infty f_X(x)\, dx$$

b) Chebyshev Inequality : $P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}$

$$\leq \frac{E[(X-\mu)^2]}{t^2}$$

$$= \frac{var[X]}{t^2}$$

c) Hoeffding Inequality : $a \leq X_i \leq b$

$$EX_i = \mu$$

$$\bar{X}_n = \frac{1}{n} \sum X_i \quad , \quad var[\bar{X}_n] = \frac{1}{n^2} n\, var[X_i] = \frac{\sigma^2}{n}$$

$$P(|\bar{X}_n - \mu| \geq t) \leq \frac{var[\bar{X}_n]}{t^2} = \frac{\sigma^2}{nt^2}$$

$$P(|\bar{X}_n - \mu| \geq t) \leq 2 \exp\left(\frac{-2nt^2}{(b-a)^2}\right)$$

d) Convergence in Probability :

$$\lim_{n \to \infty} P(|X_n - X| \geq t) = 0 \quad \forall \, t > 0$$

e) Convergence in Distribution :

$$\lim_{n \to \infty} |F_{X_n}(x) - F_X(x)| = 0 \quad \forall \, x$$

f) Law of Large Numbers :

$X_1, X_2, \dots, X_n$ iid from D

$$EX_i = \mu \,, \quad \bar{X}_n = \frac{1}{n} \sum X_i$$

$$\bar{X}_n \xrightarrow{P} \mu$$

$$P(|\bar{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$$

g) Central Limit Theorem :

$X_1, X_2, \dots, X_n$ iid from D

$$EX_i = \mu, \quad var[X_i] = \sigma^2$$

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \,, \quad Y_n \xrightarrow{D} N(0, \sigma^2)$$