

**Intro: Random Experiment:** An outcome for each run.

**Sample Space  $\Omega$ :** Set of all possible outcomes.

**Event:** Subsets of  $\Omega$ .

**Prob. of Event  $A$ :**  $P(A) = \frac{\text{Number of outcomes in } A}{\text{Number of outcomes in } \Omega}$

**Axioms:**  $P(A) \geq 0 \forall A \in \Omega$ ,  $P(\Omega) = 1$ ,

If  $A \cap B = \emptyset$ , then  $P(A \cup B) = P(A) + P(B) \forall A, B \in \Omega$

**Cond. Prob.**  $P(A|B) = \frac{P(A \cap B)}{P(B)}$

\*  $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

**Independence:**  $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

**Total Prob. Thm:** If  $H_1, H_2, \dots, H_n$  form a partition of  $\Omega$ , then  $P(A) = \sum_{i=1}^n P(A|H_i)P(H_i)$ .

**Bayes' Rule:**  $P(H_k|A) = \frac{P(H_k \cap A)}{P(A)} = \frac{P(A|H_k)P(H_k)}{\sum_{i=1}^n P(A|H_i)P(H_i)}$

\*Posteriori:  $P(H_k|A)$ , Likelihood:  $P(A|H_k)$ , Prior:  $P(H_k)$

**1 RV: CDF:**  $F_X(x) = P[X \leq x]$

**PMF:**  $P_X(x_j) = P[X = x_j] \ j = 1, 2, \dots$

**PDF:**  $f_X(x) = \frac{d}{dx} F_X(x)$

\*  $P[a \leq X \leq b] = \int_a^b f_X(x) dx$  IS THIS CORRECT?

**Cond. PMF:**  $P_X(x|A) = P[X = x|A] = \frac{P[X=x, A]}{P[A]}$  IS THIS CORRECT?

**Cond. PDF:**  $f_X(x|A) = \frac{f_{X,A}(x,a)}{f_A(a)}$  IS THIS CORRECT?

**Exp.:**  $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx \mid \sum_{k=-\infty}^{\infty} h(k) P_X(x_i=k)$

**Variance:**  $\sigma_X^2 = \text{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$

**Cond. Exp.:**  $E[X|A] = \int_{-\infty}^{\infty} x f_X(x|A) dx$

**2 RVs: Joint PMF:**  $P_{X,Y}(x, y) = P[X = x, Y = y]$

**Joint PDF:**  $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

\*  $P[(X, Y) \in A] = \int \int_{(x,y) \in A} f_{X,Y}(x, y) dx dy$

**Exp.:**  $E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$

**Correlation (Corr.):**  $E[XY]$

**Covar.:**  $\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$

**Corr. Coeff.:**  $\rho_{X,Y} = E \left[ \left( \frac{X - \mu_X}{\sigma_X} \right) \left( \frac{Y - \mu_Y}{\sigma_Y} \right) \right] = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}$

**Marginal PMF:**  $P_X(x) = \sum_{j=1}^{\infty} P_{X,Y}(x, y_j)$

**Marginal PDF:**  $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$

**Cond. PMF:**  $P_{X|Y}(x|Y) = P[X = x|Y = y] = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

**Cond. PDF:**  $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

**Bayes' Rule**

$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y') f_Y(y') dy'}$

\*  $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{P_{X|Y}(x|y) P_Y(y)}{\sum_{j=1}^{\infty} P_{X|Y}(x|y_j) P_Y(y_j)}$

**Ind.:**  $f_{X|Y}(x|y) = f_X(x) \ \forall y \Leftrightarrow f_{X,Y}(x, y) = f_X(x) f_Y(y)$

\* If independent, then uncorrelated.

**Uncorrelated:**  $\text{Cov}[X, Y] = 0 \Leftrightarrow \rho_{X,Y} = 0$

**Orthogonal:**  $E[XY] = 0$

**Cond. Exp.:**  $E[Y] = E[E[Y|X]]$  or  $E[E[h(Y)|X]]$

\*  $E[E[Y|X]]$  w.r.t.  $X \mid E[Y|X]$  w.r.t.  $Y$ .

**Estimation:** Estimate unknown parameter  $\theta$  from  $n$  i.i.d. measurements  $X_1, X_2, \dots, X_n$ ,  $\hat{\Theta}(\underline{X}) = g(X_1, X_2, \dots, X_n)$

**Estimation Error:**  $\hat{\Theta}(\underline{X}) - \theta$ .

**Unbiased:**  $\hat{\Theta}(\underline{X})$  is unbiased if  $E[\hat{\Theta}(\underline{X})] = \theta$ .

\* **Asymptotically unbiased:**  $\lim_{n \rightarrow \infty} E[\hat{\Theta}(\underline{X})] = \theta$ .

**Consistent:**  $\hat{\Theta}(\underline{X})$  is consistent if  $\hat{\Theta}(\underline{X}) \rightarrow \theta$  as  $n \rightarrow \infty$  or  $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P[|\hat{\Theta}(\underline{X}) - \theta| < \epsilon] \rightarrow 1$ .

**Sample Mean:**  $M_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

\* Given a sequence of i.i.d. RVs,  $X_1, X_2, \dots, X_n$ ,  $M_n$  is unbiased and consistent.

**Chebychev's Inequality:**  $P[|X - E[X]| \geq \epsilon] \leq \frac{\text{Var}[X]}{\epsilon^2}$

**Weak Law of Large #s:**  $\lim_{n \rightarrow \infty} P[|M_n - \mu| < \epsilon] = 1 \ \forall \epsilon > 0$ .

**ML Estimation:** Choose parameter  $\theta$  that is most likely to generate the obs.  $x_1, x_2, \dots, x_n$ .

\* Disc:  $\hat{\Theta} = \arg \max_{\theta} P_{\underline{X}}(\underline{x}|\theta) \xrightarrow{\log} \hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \log P_X(x_i|\theta)$

\* Cont:  $\hat{\Theta} = \arg \max_{\theta} f_{\underline{X}}(\underline{x}|\theta) \xrightarrow{\log} \hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \log f_X(x_i|\theta)$

**Maximum A Posteriori (MAP) Estimation:**

\* Disc:  $\hat{\theta} = \arg \max_{\theta} P_{\Theta|\underline{X}}(\theta|\underline{x}) = \arg \max_{\theta} \frac{P_{\underline{X}|\Theta}(\underline{x}|\theta) P_{\Theta}(\theta)}{P_{\underline{X}}(\underline{x})}$

\* Cont:  $\hat{\theta} = \arg \max_{\theta} f_{\Theta|\underline{X}}(\theta|\underline{x}) = \arg \max_{\theta} \frac{f_{\underline{X}|\Theta}(\underline{x}|\theta) f_{\Theta}(\theta)}{f_{\underline{X}}(\underline{x})}$

\*  $f_{\Theta|\underline{X}}(\theta|\underline{x})$ : Posteriori,  $f_{\underline{X}|\Theta}(\underline{x}|\theta)$ : Likelihood,  $f_{\Theta}(\theta)$ : Prior

**Bayes' Rule:**  $P_{\Theta|\underline{X}}(\theta|\underline{x}) = \begin{cases} \frac{P_{\underline{X}|\Theta}(\underline{x}|\theta) P_{\Theta}(\theta)}{P_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ disc.} \\ \frac{f_{\underline{X}|\Theta}(\underline{x}|\theta) f_{\Theta}(\theta)}{f_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ cont.} \end{cases}$

$f_{\Theta|\underline{X}}(\theta|\underline{x}) = \begin{cases} \frac{P_{\underline{X}|\Theta}(\underline{x}|\theta) f_{\Theta}(\theta)}{P_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ disc.} \\ \frac{f_{\underline{X}|\Theta}(\underline{x}|\theta) f_{\Theta}(\theta)}{f_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ cont.} \end{cases}$

\* Independent of  $\theta$ :  $f_{\underline{X}}(\underline{x}) = \int_{-\infty}^{\infty} f_{\underline{X}|\Theta}(\underline{x}|\theta) f_{\Theta}(\theta) d\theta$

**Beta Prior**  $\Theta$  is a Beta R.V. w/  $\alpha, \beta > 0$

$f_{\Theta}(\theta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} & \text{if } 0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}$

\*  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$

**Prop.:** 1.  $\Gamma(x+1) = x\Gamma(x)$ . For  $m \in \mathbb{Z}^+$ ,  $\Gamma(m+1) = m!$ .

2.  $\beta(\alpha, \beta) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} = \beta \binom{\alpha+\beta-1}{\alpha-1}$

3. Expected Value:  $E[\Theta] = \frac{\alpha}{\alpha+\beta}$  for  $\alpha, \beta > 0$

4. Mode (max of PDF):  $\frac{\alpha-1}{\alpha+\beta-2}$  for  $\alpha, \beta > 1$

**Drawing Beta Dist.** 1. Label x-axis from 0 to 1. 2. Identify mode.

3. Determine shape based on  $\alpha$  and  $\beta$ :  $\alpha = \beta = 1$  (uniform),  $\alpha = \beta > 1$  (bell-shaped, peak at 0.5),  $\alpha = \beta < 1$  (U-shaped w/ high density near 0 and 1),  $\alpha > \beta$  (left-skewed),  $\alpha < \beta$  (right-skewed).

**Least Mean Squares (LMS) Estimation:** Assume prior  $P_{\Theta}(\theta)$  or  $f_{\Theta}(\theta)$  w/ obs.  $\underline{X} = \underline{x}$ .

\*  $\hat{\theta} = g(\underline{x}) = E[\Theta|\underline{X} = \underline{x}] \mid \hat{\Theta} = g(\underline{X}) = E[\Theta|\underline{X}]$

**Uniform PDF**  $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

\*  $E[X] = \frac{a+b}{2}$ ,  $\text{Var}[X] = \frac{(b-a)^2}{12}$

**Conditional Exp.**  $E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$

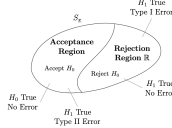
**Binary Hyp. Testing:**  $H_0$ : Null Hyp.,  $H_1$ : Alt. Hyp.

**TI Err. (False Rejection):** Reject  $H_0$  when  $H_0$  is true.

\* $\alpha(R) = P[\underline{X} \in R \mid H_0]$

**TII Err. (False Accept.):** Accept  $H_0$  when  $H_1$  is true.

\* $\beta(R) = P[\underline{X} \in R^c \mid H_1]$



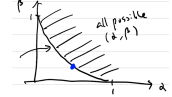
**Likelihood Ratio Test:** For each value of  $\underline{x}$ ,

\* $L(\underline{x}) = \frac{P_{\underline{X}}(\underline{x} | H_1)}{P_{\underline{X}}(\underline{x} | H_0)} \geq 1$  or  $\xi$

\*MLT: 1, LRT:  $\xi$

**Neyman-Pearson Lemma:** Given a false rejection prob. ( $\alpha$ ), the LRT offers the smallest possible false accept. prob. ( $\beta$ ), and vice versa.

\*LRT produces  $(\alpha, \beta)$  pairs that lie on the efficient frontier.



**Bayesian Hyp. Testing: MAP Rule:**

$L(\underline{x}) = \frac{p_{\underline{X}}(\underline{x} | H_1)}{p_{\underline{X}}(\underline{x} | H_0)} \geq \frac{P[H_0]}{P[H_1]}$

**Min. Cost Bayes' Dec. Rule:**  $C_{ij}$  is cost of choosing  $H_j$  when  $H_i$  is true. Given obs.  $\underline{X} = \underline{x}$ , the exp. cost of choosing  $H_j$  is  $A_j(\underline{x}) = \sum_{i=0}^1 C_{ij} P[H_i | \underline{X} = \underline{x}]$ .

**Min. Cost Dec. Rule:**  $L(\underline{x}) = \frac{P_{\underline{X}}(\underline{x} | H_1)}{P_{\underline{X}}(\underline{x} | H_0)} \geq \frac{(C_{01} - C_{00})P[H_0]}{(C_{10} - C_{11})P[H_1]}$ .

\* $C_{01}$ : False accept. cost,  $C_{10}$ : False reject. cost.

**Naive Bayes Assumption:** Assume  $X_1, \dots, X_n$  (features) are ind., then  $p_{\underline{X}}(\underline{x} | \theta) \prod_{i=1}^n p_{X_i}(x_i | \theta)$ .

**Notation:**  $P_{\underline{X}}(\underline{x} | \theta)$ , only put RVs in subscript, not values.

$P_{\underline{X}}(\underline{x} | H_i)$ , didn't put  $H$  in subscript b/c it's not a RV.

**Binomial #** of successes in  $n$  trials, each w/ prob.  $p$

$b(x | n, p) = \binom{n}{x} p^x (1-p)^{n-x}$ ,  $x = 0, 1, 2, \dots$

\* $E[X] = \mu = np$  |  $Var(X) = \sigma^2 = np(1-p)$

**Multinomial #** of  $x_i$  successes in  $n$  trials, each w/ prob.  $p_i$

$f(x_i | p_i \forall i, n) = \frac{n!}{x_1! \dots x_m!} p_1^{x_1} \dots p_m^{x_m}$

\* $\sum_i x_i = n$ , and  $\sum_{i=1}^m p_i = 1$

\* $E[X_i] = \mu_i = np_i$  |  $Var(X_i) = \sigma_i^2 = np_i(1-p_i)$

**Hypergeometric #** of successes in  $n$  draws from  $N$  items,  $k$  of which are successes

$h(x | N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$

\* $\max\{0, n - (N - k)\} \leq x \leq \min\{n, k\}$

\* $E[X] = \mu = \frac{n k}{N}$  |  $Var(X) = \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \cdot \left(1 - \frac{k}{N}\right)$

**Negative Binomial #** of trials until  $k$  successes, each w/ prob.  $p$

$b^*(x | k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$

\* $x \geq k$ ,  $x = k, k+1, \dots$

\* $E[X] = \mu = \frac{k}{p}$  |  $Var(X) = \sigma^2 = \frac{k(1-p)}{p^2}$

**Geometric #** of trials until 1st success, each w/ prob.  $p$

$g(x | p) = p(1-p)^{x-1}$

\* $x \geq 1$ ,  $x = 1, 2, 3, \dots$

\* $E[X] = \mu = \frac{1}{p}$  |  $Var(X) = \sigma^2 = \frac{1-p}{p^2}$

**Poisson #** of events in a fixed interval w/ rate  $\lambda$

$p(x | \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$

\* $x \geq 0$ ,  $x = 0, 1, 2, \dots$

\* $E[X] = \mu = \lambda t$  |  $Var(X) = \sigma^2 = \lambda t$

**Gaussian to Q Fcn:** 1. Find  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ .

2. Use table to find  $Q(x)$  for  $x \geq 0$ .

**Random Vector:**  $\underline{X} = [X_1, \dots, X_n]^T$

**Mean Vector:**  $\underline{m}_X = E[\underline{X}] = [\mu_1, \dots, \mu_n]^T$

**Corr. Mat.:**  $R_{\underline{X}} = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] & \dots & E[X_1 X_n] \\ E[X_2 X_1] & E[X_2^2] & \dots & E[X_2 X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_n X_1] & E[X_n X_2] & \dots & E[X_n^2] \end{bmatrix}$

\* $R$  is real, symmetric, and PSD ( $\forall \underline{a}, \underline{a}^T R \underline{a} \geq 0$ ).

**Covar. Mat.:**  $K_{\underline{X}} = \begin{bmatrix} \text{Var}[X_1] & \dots & \text{Cov}[X_1, X_n] \\ \text{Cov}[X_2, X_1] & \dots & \text{Cov}[X_2, X_n] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_n, X_1] & \dots & \text{Var}[X_n] \end{bmatrix}$

\* $K_{\underline{X}} = R_{\underline{X}} - \underline{m}_X \underline{m}_X^T = R_{\underline{X}} - \underline{m} \underline{m}^T$

\*Diagonal  $K_{\underline{X}} \iff X_1, \dots, X_n$  are (mutually) uncorrelated.

**Lin. Trans.**  $\underline{Y} = A \underline{X}$  ( $A$  rotates and stretches  $\underline{X}$ )

**Mean:**  $E[\underline{Y}] = A \underline{m}_X$

**Covar. Mat.:**  $K_{\underline{Y}} = A K_{\underline{X}} A^T$

**Diag. Covar. Mat.:** For any  $\underline{X}$ , if  $\underline{Y} = P^T \underline{X}$ , then

$K_{\underline{Y}} = P^T K_{\underline{X}} P = \Lambda$  (i.e.  $\underline{Y}$  is uncorrelated)

\* $K_{\underline{X}} = P \Lambda P^T$  |  $P = [\underline{e}_1, \dots, \underline{e}_n]$  of  $K_{\underline{X}}$

**Find  $K_{\underline{Y}}$**  1. Find eigenvalues, norm. eigenvectors of  $K_{\underline{X}}$ .

2. Set  $\underline{Y} = P^T \underline{X}$ ,  $K_{\underline{Y}} = \Lambda$ .

**PDF of L.T.:** If  $\underline{Y} = A \underline{X}$  w/  $A$  not singular, then

$f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}(\underline{x})}{|\det A|} \Big|_{\underline{x}=A^{-1}\underline{y}}$

**Find  $f_{\underline{Y}}(\underline{y})$**  1. Given  $f_{\underline{X}}(\underline{x})$ , define transformation  $A$

2. Determine  $|\det A|$ ,  $A^{-1}$ , then  $f_{\underline{Y}}(\underline{y})$ .

**Gaussian RVs: Analytic Tractability:** PDF of jointly Gaussian  $X_1, \dots, X_n$  is Gaussian vector.

\* $f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} |\det \Sigma|} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1}(\underline{x}-\underline{\mu})}$

\* $\underline{\mu} = \underline{m}_X, \Sigma = K_X$  (if  $\Sigma$  is not singular)  
**Properties of Guassian Vector:**  $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$   
1. PDF is completely determined by  $\underline{\mu}, \Sigma$ .  
2.  $\underline{X}$  uncorrelated  $\implies \underline{X}$  independent.  
3. Any L.T.  $\underline{Y} = A\underline{X} + \underline{b}$  is Gaussian vector w/  $\underline{\mu}_Y = A\underline{\mu}_X$ ,  
 $\Sigma_Y = A\underline{\Sigma}_X A^T$ .  
4. Any subset of  $\{X_i\}$  are jointly Gaussian.  
5. Any cond. PDF of a subset of  $\{X_i\}$  given the other elements is Gaussian.

**Diag. of Gaussian Covar.** Eigen decomp. of  $\Sigma_X$ :  $\{\lambda_i\}, \{e_i\}$   
 $A = [\underline{a}_1, \dots, \underline{a}_n]^T$ , then  $\underline{Y} = A\underline{X}$  has  $\Sigma_Y = \Lambda$ .  
\* $\underline{Y}$ : Vector of indep. Gaussian RVs.  
**How to go from  $Y$  to  $X$ ?** 1. Given,  $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$ , then find  $\Sigma = P\Lambda P^T$ .  
2.  $\underline{Y} \sim \mathcal{N}(\underline{0}, I)$  3.  $\underline{W} = \sqrt{\Lambda}\underline{Y}$  4.  $\underline{Y} = P\underline{W}$  4.  $\underline{X} = \underline{Y} + \underline{\mu}$   
**Gaussian Discriminant Analysis:** Obs:  $\underline{X} = \underline{x} = (x_1, \dots, x_D)$   
Hyp:  $\underline{x}$  is gen. by  $\mathcal{N}(\underline{\mu}_c, \Sigma_c), c \in C$   
Dec: Which "Gaussian bump" generated  $\underline{x}$ ?  
Prior:  $P[C = c] = \pi_c$  (Gaussian Mixture Model)

**MAP Rule:**  $\hat{c} = \arg \max_c P_C[c | \underline{X} = \underline{x}] = \arg \max_c f_{\underline{X}|C}(\underline{x} | c) \pi_c$

**LGD:**  $\Sigma_c = \Sigma \forall c$ , find  $c$  w/ best  $\underline{\mu}_c$   
 $\hat{c} = \arg \max_c \underline{\beta}_c^T \underline{x} + \gamma_c$   
\* $\underline{\beta}_c^T = \underline{\mu}_c^T \Sigma^{-1} \underline{x} \mid \gamma_c = \log \pi_c - \frac{1}{2} \underline{\mu}_c^T \Sigma^{-1} \underline{\mu}_c$   
\*Bin. hyp. dec. boundary:  $\underline{\beta}_0^T \underline{x} + \gamma_0 = \underline{\beta}_1^T \underline{x} + \gamma_1$  (lin. in space of  $\underline{x}$ )  
**QGD:**  $\Sigma_c$  are diff., find  $c$  w/ best  $\underline{\mu}_c, \Sigma_c$   
\*Bin. hyp. dec. boundary: Quadratic in space of  $\underline{x}$   
**How to find  $\underline{x}_c, \underline{\mu}_c, \Sigma_c$ :** Given  $n$  points gen. by GMM, then  $n_c$  points  $\{\underline{x}_1^c, \dots, \underline{x}_{n_c}^c\}$  come from  $\mathcal{N}(\underline{\mu}_c, \Sigma_c)$   
 $\hat{\pi}_c = \frac{n_c}{n}, \hat{\mu}_c = \frac{1}{n_c} \sum_{i=1}^n \underline{x}_i^c$ ,  
 $\Sigma_c = \frac{1}{n_c} \sum_{i=1}^{n_c} (x_i^c - \hat{\mu}_c)(x_i^c - \hat{\mu}_c)^T$