$$P(A|B) = \frac{P(A,B)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

X, Y *independent iff* $p(x,y) = p(x)p(y)$

$$E(X) = \sum_x x p(x)$$

$$P(B) = \sum_{j=1}^{n} P(B|A_j)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{j=1}^{n} P(B|A_j)P(A)}$$

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - \bar{x})(y_i - \bar{y})$$

$$Var(X) = E((X - E(X)^x)$$
where $E(X) = \mu$

$$P(A) = P(A|B)P(B) + P(A|B')P(B')$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$\lambda(\alpha_i|\omega_j) = Cost\ of\ taking\ \alpha\ when\ nature\ is\ \omega_j$

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)p(\omega_j|\boldsymbol{x})$$

$$R = \sum_{i=1}^{c} R(\alpha_i|\boldsymbol{x})$$

Mahalonobis Distance $\quad r = (\boldsymbol{x}_0 - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu})$

A refined measurement of a test point's distance from the distribution. Takes into account the shape of the distribution.

$$g_i(x) = \quad +\boldsymbol{w}_i^t \boldsymbol{x} \quad \frac{1}{\sigma^2}\boldsymbol{\mu}_i \boldsymbol{x}$$
$$+w_{i0} \quad \frac{1}{2\sigma^2}\boldsymbol{\mu}_i^t \boldsymbol{\mu}_i + \ln(P(\omega_i))$$

$R(a_1|\boldsymbol{x}) = \lambda_{1,1}P(\omega_1|\boldsymbol{x}) + \lambda_{1,2}P(\omega_2|\boldsymbol{x})$

$R(a_2|\boldsymbol{x}) = \lambda_{2,1}P(\omega_1|\boldsymbol{x}) + \lambda_{2,2}P(\omega_2|\boldsymbol{x})$

Where $\lambda_{1,1}$ is the cost of taking action 1 when the state of nature is ω_1

Decide ω_1 if
$$\frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)} > \frac{(\lambda_{1,2} - \lambda_{2,2})}{(\lambda_{2,1} - \lambda_{1,1})}\frac{P(\omega_2)}{P(\omega_1)}$$

$$g_i(x) = \quad \boldsymbol{x}^t \boldsymbol{W}_i \boldsymbol{x} \quad \boldsymbol{x}^t \frac{-1}{2}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}$$
$$+\boldsymbol{w}_i \boldsymbol{x} \quad +(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i)^t \boldsymbol{x} \quad (2)$$
$$-\frac{1}{2}\boldsymbol{\mu}_i^t \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i \quad (2)$$
$$+w_{i0} \quad -\frac{1}{2} ln|\boldsymbol{\Sigma}_i| \quad (2)$$
$$+\ln(P(\omega_i)) \quad (2)$$

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)p(\omega_j|\boldsymbol{x}) = \sum_{i \neq j} p(\omega_j|\boldsymbol{x}) = 1 - p(\omega_i|\boldsymbol{x})$$

Case 1:	Diagonal, equal covariances	linear
Case 2:	Equal, not diagonal	
Case 3:	Anything goes	quadratic

Adjust x^* to adjust ROC/FAR and FRR

Bayesian Networks
Arcs represent causal influence
$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|\pi_i)$$

If missing information, use the sum of all possible values in place of that term:
$$p(c_?|x_1)p(x_1|a_?)p(a_?)$$
$$\Rightarrow \sum_{i=1}^{n} p(c_i|x_1) * \sum_{i=1}^{n} p(x_1|a_i)\, p(a_i)$$

Naïve Bayesian Network: Assume that all features are conditionally independent given the class:
$$p(x_1, \ldots, x_n|\omega) = \prod_{i=1}^{n} p(x_i|\omega)$$

$$P(\mathbf{X}/e) = \frac{P(\mathbf{X},e)}{P(e)} = \alpha P(\mathbf{X},e)$$

Decision Boundary: $\boldsymbol{w}^t(\boldsymbol{x} - \boldsymbol{x}_0) = 0$

Case 1:
$$\boldsymbol{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$
$$\boldsymbol{x}_0 = \frac{(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)}{2} - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} * \ln(\frac{P(\omega_i)}{P(\omega_j)})(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

Case 2:
$$\boldsymbol{w} = \Sigma^{-1}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$
$$\boldsymbol{x}_0 = \frac{(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)}{2} - \frac{\ln(\frac{P(\omega_i)}{P(\omega_j)})}{(\boldsymbol{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i)} * (\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

Chernoff Bound $\quad P(error) \leq P^{\beta}(\omega_1)P^{1-\beta}(\omega_2) * e^{-\kappa(\beta)}$

$$\kappa(\beta) = \frac{\beta(1-\beta)}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t * [\beta\boldsymbol{\Sigma}_1 + (1-\beta)\boldsymbol{\Sigma}_2]^{-1} * (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2}\ln(\frac{|\beta\boldsymbol{\Sigma}_1 + (1-\beta)\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|^{\beta}|\boldsymbol{\Sigma}_2|^{1-\beta}})$$

Special: Battacharyya (β=0.5), $P(\omega_1) = P(\omega_2)$ $\quad P(error) \leq 0.5 * e^{-\kappa(\beta)}$

$$\kappa(\beta) = \frac{1}{8}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t * \left[\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right]^{-1} * (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2}\ln(\frac{\left|\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right|}{\sqrt{|\boldsymbol{\Sigma}_1|}\sqrt{|\boldsymbol{\Sigma}_2|}})$$

Both distributions must be Gaussian for meaningful results

MLE mu $\quad \hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k$

MLE sigma $\quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t$

$$\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k = \boldsymbol{\mu} =$$
$$E[\hat{\boldsymbol{\Sigma}}] = \frac{n-1}{n}\sigma^2 \neq \sigma^2 \quad |||$$
$$\boldsymbol{C} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{x}_k - \boldsymbol{\mu})(\boldsymbol{x}_k - \boldsymbol{\mu})^t$$

MLE	Maximize $p(\mathcal{D}|\boldsymbol{\theta})$
MAP	Maximize $p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})$
BE	

BE
$$p(\boldsymbol{x}|\mathcal{D}) = \int p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$
where $p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$
and $p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{n} p(\boldsymbol{x}_k|\boldsymbol{\theta})$

data $(N \to \infty)$

$p(x|D) = \int p(x|\theta)\, p(\theta|D)\, d\theta$ where $p(\theta|D) = \frac{P(D|\theta)\, p(\theta)}{P(D)}$

As $N \to \infty$, $P(D|\theta)$ reaches a sharp peak at $\theta = \hat{\theta}$, thus, $P(\theta|D)$ peaks at this point (assuming $p(\theta)$ is flat around $\hat{\theta}$

(Remember: ML maximizes $P(D|\theta)$!!)

$$\max \frac{|U^T S_b U|}{|U^T S_w U|} \qquad \mathbf{y} = U^T \mathbf{x}$$

$$S_B u_k = \boldsymbol{\lambda}_k S_w u_k$$