II. Bayesian Decision Theory
Bayes formula

- Two-category classification:
  - Model
    - *Two categories:* ω_1 - sedan, ω_2 - truck;
    - $P(\omega_1)$, $P(\omega_2)$ a priory probabilities;
  - Classification based only on a priori probabilities:
    - Decision rule: ω_1 if $P(\omega_1) > P(\omega_2)$; ω_2 otherwise;

II. Bayesian Decision Theory
Bayes formula

- Classification based on features:
  - Feature x : continuous;
  - Likelihood functions - class conditional probabilities: $p(x|\omega_1)$, $p(x|\omega_2)$;
  - A priori probabilities: $P(\omega_1)$, $P(\omega_2)$
  - Bayes formula

$$P(\omega_i \mid x) = \frac{p(x \mid \omega_i)P(\omega_i)}{p(x)}, \quad i = 1,2$$

$$p(x) = \sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)$$

  - Observation x.

II. Bayesian Decision Theory
Bayes formula

- Decision rule:

  If $P(\omega_1|x) > P(\omega_2|x)$ then $x \in \omega_1$

  If $P(\omega_2|x) > P(\omega_1|x)$ then $x \in \omega_2$

- Probability of error x:

$$P(Err\,|\,x) = \begin{cases} P(\omega_1\,|\,x)\ \text{nÕu lùa chän}\ \omega_2 \\ P(\omega_2\,|\,x)\ \text{nÕu lùa chän}\ \omega_1 \end{cases}$$

- Decision rule: minimize $P(Err|x)$.
- Average probability of error:

$$P(Err) = \int_{-\infty}^{\infty} P(Err, x)dx = \int_{-\infty}^{\infty} P(Err\,|\,x)\,p(x)dx$$

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Model of the problem:
  - Object: feature vector **x** – d-D vector;
  - Object space - feature space $\boldsymbol{\Omega}$: n partitions - classes $\boldsymbol{\Omega} = \{\omega_1, \ldots, \omega_n\}$ – n state of the space;
  - Decision actions $\boldsymbol{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$;
  - $\lambda(\alpha_i|\omega_j)$ – loss incurred for taking an action α_i when the state of nature is ω_j;

II. Bayesian Decision Theory
General Bayesian Decision Theory

- For each class ω_j, probability densities: $p(\mathbf{x}|\ \omega_j)$ – likelihood function: a probability of an event when an object has feature vector **x** belong to class ω_j ;
- $P(\omega_j)$ –*a priori* probability of class ω_j .

- Bayes rule:

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})}, \quad i = 1,\ldots,n$$

$$p(\mathbf{x}) = \sum_{j=1}^{n} p(\mathbf{x} \mid \omega_j)P(\omega_j) \quad , \sum_{i=1}^{n} P(\omega_i \mid \mathbf{x}) = 1$$

II. Bayesian Decision Theory
General Bayesian Decision Theory

- We have observation **x**, take an action α_k, and If the true state of nature is ωj , the loss will be $\lambda(\alpha_k|\omega_j)$;
- The expected loss associated with taking action α_k :

$$R(\alpha_k \mid \mathrm{x}) = \sum_{i=1}^{n} \lambda(\alpha_k \mid \omega_i) P(\omega_i \mid \mathrm{x})$$

- Decision rule: select the action, that minimize $R(\alpha_k|\mathbf{x})$ - *conditional risk* with observation **x**.

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Problem formulation: find decision rule:
  - Based on prior probabilities $P(\omega_i)$;
  - Minimize the over all risk $\boldsymbol{R}$;

$$R = \int_{\Omega} R(\alpha(\mathbf{x}) \mid \mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$$

II. Bayesian Decision Theory General Bayesian Decision Theory

- Decision function $\alpha(\mathbf{x})$;
- Over all risk $\boldsymbol{R}$:

$$R = \int_{\Omega} R(\alpha(\mathbf{x}) \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

- Select $\alpha(\mathbf{x})$: $R(\alpha_i(\mathbf{x})|\mathbf{x})$ min for every $\mathbf{x}$ => over all risk $\boldsymbol{R}$ min;

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Bayes decision rule:
  - Compute the conditional risks for each action:

$$R(\alpha_k \mid \mathrm{x}) = \sum_{i=1}^{n} \lambda(\alpha_k \mid \omega_i) P(\omega_i \mid \mathrm{x}), \quad k = 1, \ldots, m$$

  - Select an action α_k : $R(\alpha_k|\mathbf{x})$ min;

$$\alpha_k(\mathbf{x}) = \arg\min_{1 \le i \le m}(R(\alpha_i \mid \mathbf{x}))$$

  - Bayes risk: over all risk $\boldsymbol{R}$;

II. Bayesian Decision Theory General Bayesian Decision Theory

- Bayes decision scheme:
  - Compute a posteriori probabilities $P(\omega_i|\mathbf{x})$
    - Based on likelihood functions: $p(\mathbf{x}|\omega_i)$.
  - Compute conditional risks $R(\alpha_k|\mathbf{x})$
  - Select decision action α_i : $R(a_i|\mathbf{x})$ min.
  - Problem: $p(\mathbf{x}|\omega_i)$??

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Example: Two classes classification
  - Feature space: $\boldsymbol{\Omega} = \{\omega_1, \omega_2\}$; $n = 2$
  - Set of decision actions $\boldsymbol{A} = \{\alpha_1, \alpha_2\}$: $m = n$
  - The decision making risk $\lambda(\alpha_k|\omega_i) = \lambda_{ki}$;
    - For example: $\lambda_{kk} = \lambda_{ii} = 0$, $\lambda_{ki} = \lambda_{ik} = 1$
  - The conditional risks:

$$R(\alpha_1 \mid \mathbf{x}) = \lambda_{11} P(\omega_1 \mid \mathbf{x}) + \lambda_{12} P(\omega_2 \mid \mathbf{x})$$
$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x})$$

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Selection rules:
  - Select ω_1 iff $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$, that means:

$$(\lambda_{21} - \lambda_{11})P(\omega_1 \,|\, \mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2 \,|\, \mathbf{x})$$

  - As usual: $\lambda_{ki} > \lambda_{ii}$ with $k \neq i$, so:

$$(\lambda_{21} - \lambda_{11})P(\omega_1 \,|\, \mathrm{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2 \,|\, \mathrm{x}) \quad \Leftrightarrow$$

$$(\lambda_{21} - \lambda_{11})\,p(\mathrm{x} \,|\, \omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})\,p(\mathrm{x} \,|\, \omega_2)P(\omega_2) \quad \Leftrightarrow$$

$$\theta = \underbrace{\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \bullet \frac{P(\omega_2)}{P(\omega_1)}}_{\text{h»ng sè kh«ng phô thuéc x}} < \frac{p(\mathrm{x} \,|\, \omega_1)}{p(\mathrm{x} \,|\, \omega_2)}$$

Constant, which doesn't depend on **x**

  - Bayes's rule states that: select class ω_1 if $\dfrac{p(x \,|\, \omega_1)}{p(x \,|\, \omega_2)}$ exceed predetermined threshold θ.

II. Bayesian Decision Theory
General Bayesian Decision Theory

- Example: P(RD) = 0.3 and P(SD)=0.7.
  - Cloudy when rain : 0.9
  - Cloudy when without rain: 0.6
  - Today, when go out, cloudy: do we need bring raincoat ?
  - Tính P(M|m), P(KM|m)
  - $P(M|m) = p(m|M)P(M)/p(m)=0.9*0.3/p(m)=0.27/p(m)$
  - $P(KM|m)=p(m|KM)P(KM)/p(m)=0.6*0.7/p(m)=0.42/p(m)$

II. Bayesian Decision Theory Minimum Error rate Classification

- Minimum error rate classification
- Minimax Criterion
- Neyman-Pearson Criterion

II. Bayesian Decision Theory
Minimum Error rate Classification

- Special Case: minimum error rate classification
  - Number of decision rules equal the number of classes;
  - Decision $\alpha_i(\mathbf{x})$: state of the observation $\mathbf{x}$ is ω_i;
  - Zero - one loss fuction $\lambda(\alpha_k|\omega_i)$:

$$\lambda(\alpha_k|\omega_i) = \begin{cases} 0, & \text{when } k = i \\ 1, & \text{when } k \neq i \end{cases} \qquad i, k = 1, \ldots, n$$

  - The loss function assigns:
    - No loss to a correct classification;
    - Unit loss to any error: all errors are equally costly;

II. Bayesian Decision Theory
Minimum Error rate Classification

- When: $m = n$: the loss will be:

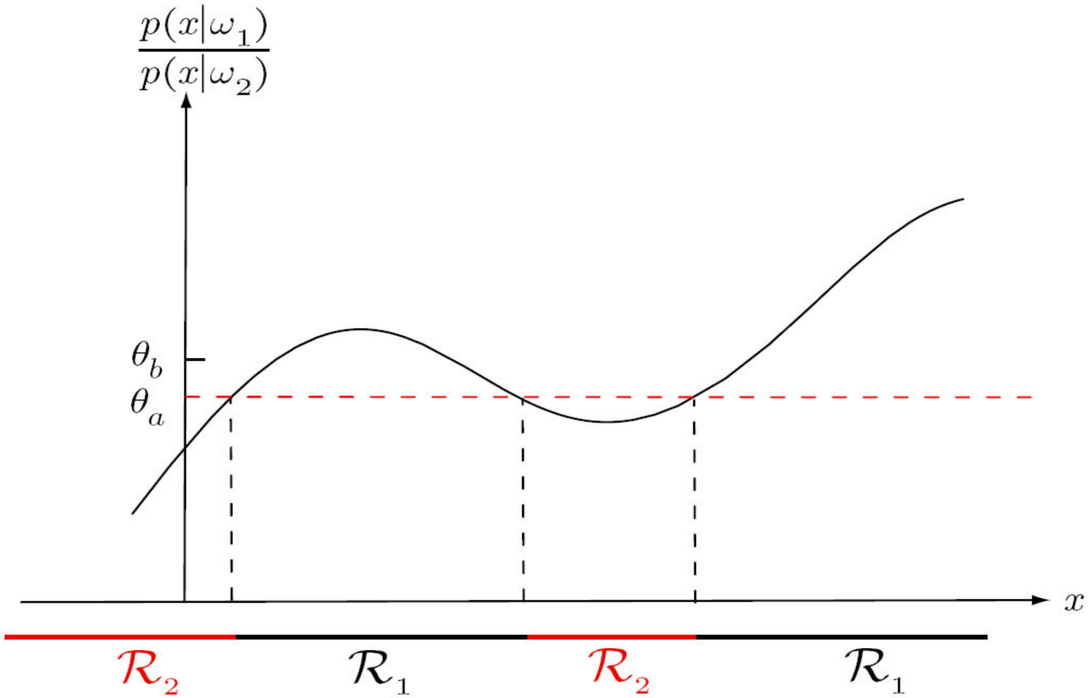
$$R(\alpha_k \mid \mathrm{x}) = \sum_{i=1}^{n} \lambda(\alpha_k \mid \omega_i) P(\omega_i \mid \mathrm{x}) = \sum_{i \neq k} P(\omega_i \mid \mathrm{x}) = 1 - P(\omega_k \mid \mathrm{x})$$

- $P(\omega_i|\mathbf{x})$ – conditional probability that decision action α_i is correctly selected.
- Bayes decision rules minimize the risk of selecting action, that minimizes conditional risks.
  - So, to minimize average value of error probability, one must select an action *i*, that maximizes *aposteriori* probability $P(\omega_i|\mathbf{x})$

II. Bayesian Decision Theory Minimum Error rate Classification

- Decision rule: In the case of Minimum error rate classification:
  - Select i , that maximize $P(\omega_i|\mathbf{x})$ or
  - Choose ω_i if $P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})\ \forall j \neq i$
- Denote subspace corresponding with subclass ω_i by R_i. This space can be unconnected.

II. Bayesian Decision Theory Minimum Error rate Classification



The likelihood ratio $p(\mathbf{x}|\omega_1)/p(\mathbf{x}|\omega_2)$. In case of classification loss function λ, decision boundaries are determined by threshold θ_a

II. Bayesian Decision Theory Minimum Error rate Classification

- Minimax Criterion
  - When we must design our classifier to perform well over a *range* of prior probabilities.
  - Design classifier so that the worst overall risk for any value of the priors is as small as possible: minimize the maximum possible overall risk

II. Bayesian Decision Theory Minimum Error rate Classification

- The model:
  - R_i - feature space for class ω_i, in the two class case:

$$R = \int_{R_1} [\lambda_{11} P(\omega_1) p(\mathbf{x} \mid \omega_1) + \lambda_{12} P(\omega_2) p(\mathbf{x} \mid \omega_2)] d\mathbf{x} +$$

$$+ \int_{R_2} [\lambda_{21} P(\omega_1) p(\mathbf{x} \mid \omega_1) + \lambda_{22} P(\omega_2) p(\mathbf{x} \mid \omega_2)] d\mathbf{x}$$

  - We have: $P(\omega_2) = 1 - P(\omega_1)$ and

$$\int_{C_1} p(\mathbf{x} \mid \omega_1) d\mathbf{x} = 1 - \int_{C_2} p(\mathbf{x} \mid \omega_1) d\mathbf{x}$$

II. Bayesian Decision Theory Minimum Error rate Classification

- We have:

$$R(P(\omega_1)) = \overbrace{\lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x} \mid \omega_2)] d\mathbf{x}}^{= R_{m,m}\text{, minimax risk}} +$$

$$+ P(\omega_1) \underbrace{\left[(\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x} \mid \omega_1) d\mathbf{x} - (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x} \right]}_{=0\text{, for minimax solution}}$$

- The overall risk is linear in $P(\omega_1)$.
- If decision boundary will be defined so scale of $P(\omega_1)$ is 0, then minimal risk is free term and independent of prior $P(\omega_1)$ - *minimax solution*

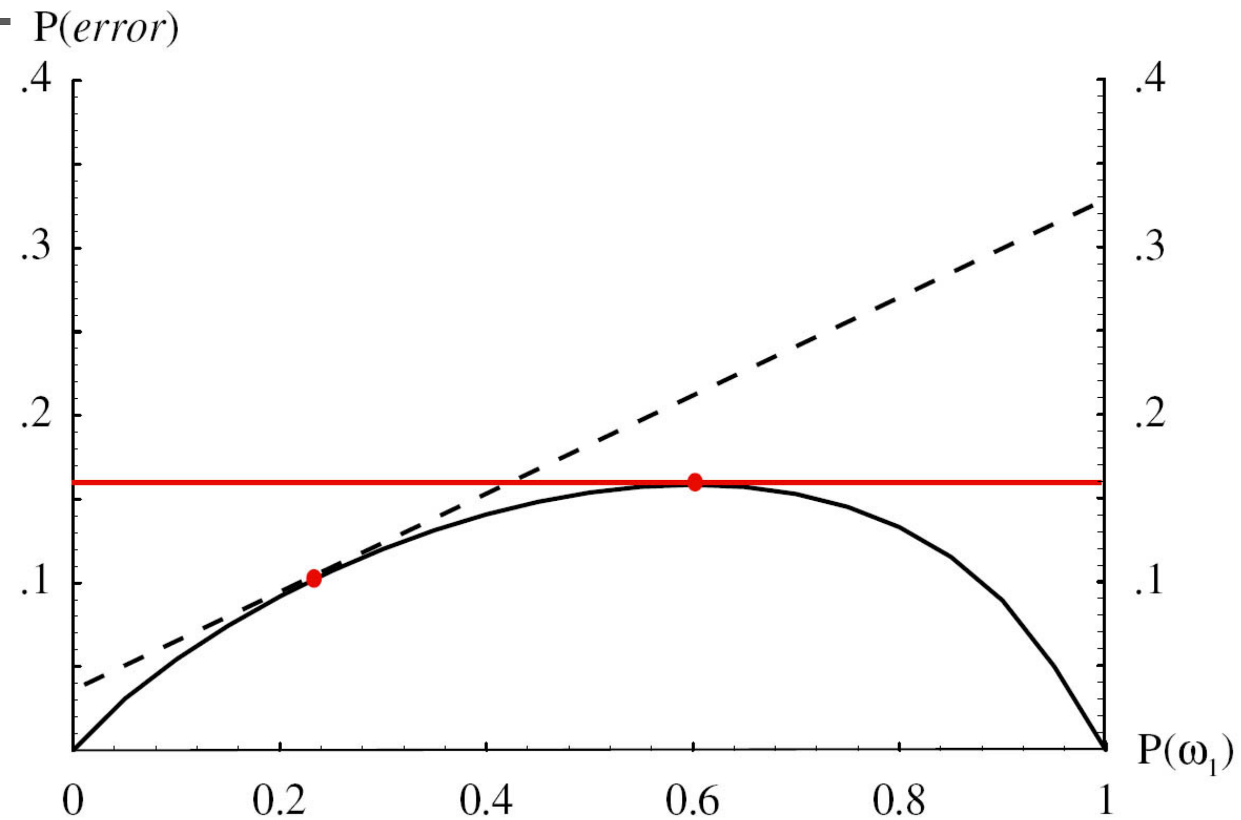
II. Bayesian Decision Theory Minimum Error rate Classification

- Minimax risk:

$$R_{m,m} = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(\mathbf{x} \mid \omega_2)]d\mathbf{x} = \lambda_{11} + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(\mathbf{x} \mid \omega_1)]d\mathbf{x}$$

- Minimax criterion defines:
  - Search for the prior for which the Bayes risk is maximum;
  - The corresponding decision boundary gives the minimax solution;
  - The value of minimax risk R_{mm} is equal the worst Bayes risk.

II. Bayesian Decision Theory
Minimum Error rate Classification



Relation of P(error) and prior $P(\omega_i)$

II. Bayesian Decision Theory Minimum Error rate Classification

- Neyman-Pearson criterion
  - Minimize overall risk under constraints.
  - Minimize total risk subject to the constraint for some *i*:.

$$\int R(\alpha_i \mid \mathbf{x}) d\mathbf{x} < Const$$

  - Neyman-Pearson criterion.
    - The Neyman-Pearson criterion says that we should construct our decision rule to have maximum probability of detection while not allowing the probability of false alarm to exceed a certain value

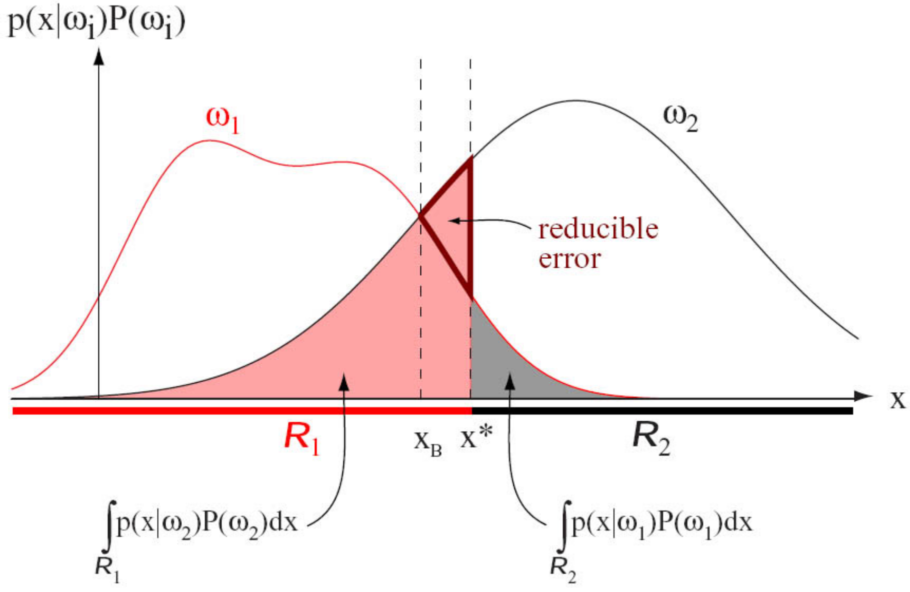
II. Bayesian Decision Theory
Error probabilities and Integrals

- Two-category classification
  - The classifier has devided the space into two regions R_1, R_2;
  - Two ways, that a classification error can occur:
    - Observation x falls in region R_2 and the true state of nature is ω_1;
    - Observation x falls in region R_1 and the true state of nature is ω_2;
    - These events are mutually exclusive and exhaustive.

II. Bayesian Decision Theory
Error probabilities and Integrals

- The probability of error is:

$$P(Error) = P(\mathbf{x} \in R_2, \omega_1) + P(\mathbf{x} \in R_1, \omega_2) =$$
$$= P(\mathbf{x} \in R_2 \mid \omega_1)P(\omega_1) + P(\mathbf{x} \in R_1 \mid \omega_2)P(\omega_2) =$$
$$= \int_{R_2} p(\mathbf{x} \mid \omega_1)P(\omega_1)d\mathbf{x} + \int_{R_1} p(\mathbf{x} \mid \omega_2)P(\omega_2)d\mathbf{x}$$



II. Bayesian Decision Theory
Error probabilities and Integrals

- Multicategory case:
  - Probability of being correct:

$$P(correct) = \sum_{i=1}^{n} P(\mathbf{x} \in R_i, \omega_i) =$$

$$= \sum_{i=1}^{n} P(\mathbf{x} \in R_i \mid \omega_i) P(\omega_i) =$$

$$= \sum_{i=1}^{n} \int_{R_i} p(\mathbf{x} \mid \omega_i) P(\omega_i) d\mathbf{x}$$

  - Probability of error:

P(Error) = 1 - P(correct)

II. Bayesian Decision Theory
Error probabilities and Integrals

- Error bounds for Normal Densities:
  - Chernoff bound

$$P(Error) \le P^{\beta}(\omega_1)P^{1-\beta}(\omega_2)\int p^{\beta}(\mathbf{x}\,|\,\omega_1)\,p^{1-\beta}(\mathbf{x}\,|\,\omega_2)\,d\mathbf{x}$$

$$0 \le \beta \le 1$$

  - Normal densities:

$$\int p^{\beta}(\mathbf{x}\,|\,\omega_1)\,p^{1-\beta}(\mathbf{x}\,|\,\omega_2)\,d\mathbf{x} = e^{-k(\beta)}$$

$$k(\beta) = \frac{\beta(1-\beta)}{2}(\mu_2-\mu_1)^t[\beta\Sigma_1+(1-\beta)\Sigma_2]^{-1}(\mu_2-\mu_1) + \frac{1}{2}\ln\frac{|\beta\Sigma_1+(1-\beta)\Sigma_2|}{|\Sigma_1|^{\beta}|\Sigma_2|^{1-\beta}}$$

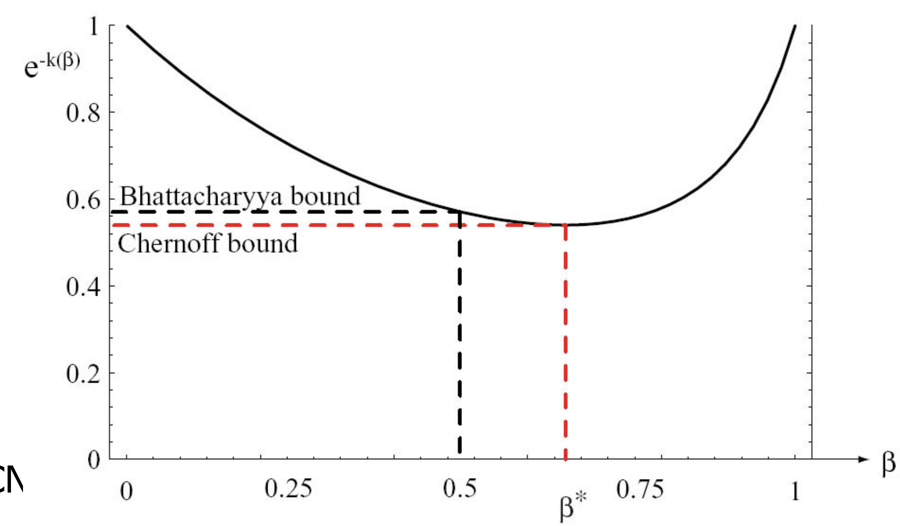
II. Bayesian Decision Theory
Error probabilities and Integrals

- Bhattacharyya Bound

$$P(Error) \le \sqrt{P(\omega_1)P(\omega_2)} \int \sqrt{p(\mathbf{x} \mid \omega_1)\, p(\mathbf{x} \mid \omega_2)}\, d\mathbf{x}$$

$$= \sqrt{P(\omega_1)P(\omega_2)}\, e^{-k(1/2)}$$

$$k(1/2) = \frac{1}{8}(\mu_2 - \mu_1)^t \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \ln \frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$



II. Bayesian Decision Theory
Bayes rules for Normal Densities

- Discriminant functions:

$$g_i(\mathbf{x}) = \ln(p(\mathbf{x} \mid \omega_i)) + \ln(P(\omega_i))$$

  - Multivariate normal densities:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu)^T \sum{}^{-1} (\mathbf{x} - \mu) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

III. Parametric methods Supervised Learning

- Maximum likelihood;
- Bayesian Estimation
- EM(Expectation Maximization)
- HMM