Raied Aljadaany

# Pattern Recognition Theory

## Recitation 3: Decision Theory IV



Electrical & Computer ENGINEERING

# Minimum Error Classifier

$$\left\{P(\omega_1)f(x|\omega_1)-P(\omega_2)f(x|\omega_2)\right\}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\mathbf{0}$$

# Bayes Risk

- What if cost of misclassifying a class-1 sample is different than misclassifying a class-2 sample
- Introduce a loss that is a measure of the cost of making a decision that a sample belongs to class *i* when the true class is *j*.
- Bayes Decision rule now takes the action that minimizes the **Overall Risk (Risk is Expected Loss)**!!!
  - The objective function we try to minimize is also called the cost function

# Minimum Risk Classifier : 2 Class Example

$\lambda_{11} = \lambda(\alpha_1 | \omega_1)$ True class is $\omega_1$ -> loss of taking action $\alpha_1$ i.e. Deciding $\omega_1$ — 0 or small

$\lambda_{22} = \lambda(\alpha_2 | \omega_2)$ True class is $\omega_2$ -> loss of taking action $\alpha_2$ i.e. Deciding $\omega_2$ — 0 or small

$\lambda_{12} = \lambda(\alpha_1 | \omega_2)$ True class is $\omega_2$ -> loss of taking action $\alpha_1$ i.e. Deciding $\omega_1$ — large

$\lambda_{21} = \lambda(\alpha_2 | \omega_1)$ True class is $\omega_1$ -> loss of taking action $\alpha_2$ i.e. Deciding $\omega_2$ — large

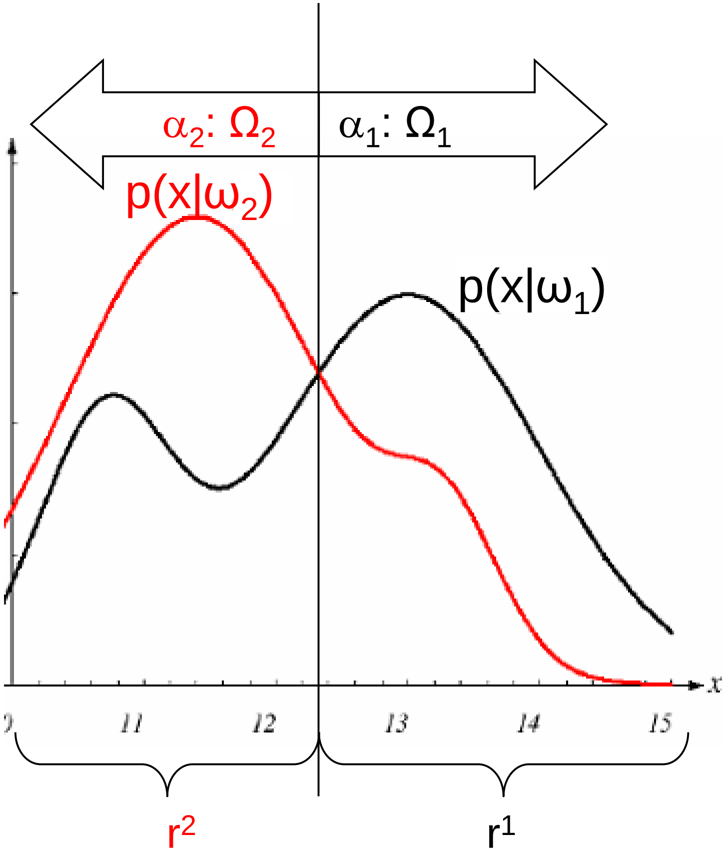
The conditional risk of deciding $\omega_1$ is:

$$l^1(x) = R\left(\alpha_1 | x\right) = \lambda_{11} P\left(x | \omega_1\right) P(\omega_1) + \lambda_{12} P\left(x | \omega_2\right) P(\omega_2)$$

The conditional risk of deciding $\omega_2$ is:

$$l^2(x) = R\left(\alpha_2 | x\right) = \lambda_{21} P\left(x | \omega_1\right) P(\omega_1) + \lambda_{22} P\left(x | \omega_2\right) P(\omega_2)$$

# Minimum (Bayes) Risk – Cont'd




The total risk over region $\Omega_1$ is :

$$r^1 = \int_{\Omega_1} \left[ \lambda_{11} P(x \mid \omega_1) P(\omega_1) + \lambda_{12} P(x \mid \omega_2) P(\omega_2) \right] dx$$

The total risk over region $\Omega_2$ is :

$$r^2 = \int_{\Omega_2} \left[ \lambda_{21} P(x \mid \omega_1) P(\omega_1) + \lambda_{22} P(x \mid \omega_2) P(\omega_2) \right] dx$$

The overall risk ($r$) is:

$$r = \int_{\Omega_1} \left\{ \lambda_{11} P(x \mid \omega_1) P(\omega_1) + \lambda_{12} P(x \mid \omega_2) P(\omega_2) \right\} dx \ldots$$

$$+ \int_{\Omega_2} \left\{ \lambda_{21} P(x \mid \omega_1) P(\omega_1) + \lambda_{22} P(x \mid \omega_2) P(\omega_2) \right\} dx$$

# Minimum (Bayes) Risk – Cont'd

The overall risk is

$$r = \int_{\Omega_1} \left\{ \lambda_{11} P\left(x \mid \omega_1\right) P\left(\omega_1\right) + \lambda_{12} P\left(x \mid \omega_2\right) P\left(\omega_2\right) \right\} dx \ldots$$

$$+ \int_{\Omega_2} \left\{ \lambda_{21} P\left(x \mid \omega_1\right) P\left(\omega_1\right) + \lambda_{22} P\left(x \mid \omega_2\right) P\left(\omega_2\right) \right\} dx$$

Using $\int_{\Omega_2} P(x \mid \omega_1) dx = 1 - \int_{\Omega_1} P(x \mid \omega_1) dx$

$$r = \lambda_{11} P\left(\omega_1\right) + \lambda_{12} P\left(\omega_2\right) + \int_{\Omega_2} \left(\lambda_{21} - \lambda_{11}\right) P\left(x \mid \omega_1\right) P\left(\omega_1\right) \ldots$$

$$+ \left(\lambda_{22} - \lambda_{12}\right) P\left(x \mid \omega_2\right) P\left(\omega_2\right)$$

Assign to $\Omega_2$ if negative

and $\Omega_1$ and if positive

Hence, we get $(\lambda_{21} - \lambda_{11}) P(x \mid \omega_1) P(\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (\lambda_{12} - \lambda_{22}) P(x \mid \omega_2) P\left(\omega_2\right)$

# Minimum Risk Classifier

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)} \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}$$

# Bayes Classifiers So Far…

$$\frac{P(x\,|\,\omega_1)}{P(x\,|\,\omega_2)} \mathop{\gtrless}_{\omega_2}^{\omega_1} \frac{P(\omega_2)}{P(\omega_1)} \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}$$

**From the likelihood functions, we can design a decision boundary** ➡ Maximum likelihood

⇓

**If we consider class priors, the decision boundary changes** ➡ Minimum Error

⇓

**If we consider class priors and risks, the decision boundary changes again** ➡ Minimum Risk

# Variations of Likelihood Ratio Test

**Bayes Criterion**: General Form

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)} \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}$$

**Zero-one loss function:**

$$\lambda_{ij} = \begin{cases} 0 & \mathrm{i} = \mathrm{j} \\ 1 & \mathrm{i} \neq \mathrm{j} \end{cases}$$

**Maximum Likelihood (ML) Criterion (equal priors and zero-one loss)**

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 1$$

**Maximum a Posteriori (MAP) Criterion**:

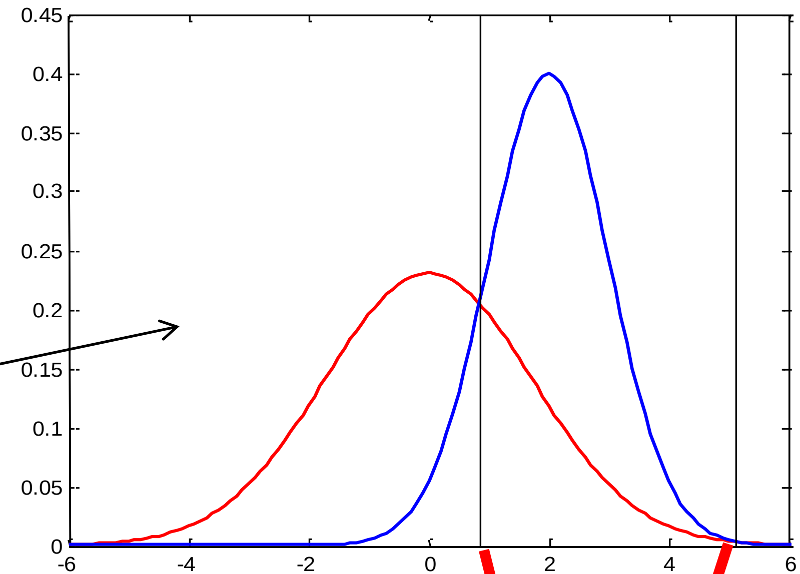
$$\frac{P(x|\omega_1)}{P(x|\omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)}$$

# Minimum Risk : 2-Class Problem Example

$$P(x\,|\,\omega_1) = \frac{1}{\sqrt{2\pi 3}}\exp\left(-0.5\frac{x^2}{3}\right)$$

$$P(x\,|\,\omega_2) = \frac{1}{\sqrt{2\pi}}\exp\left(-0.5(x-2)^2\right)$$

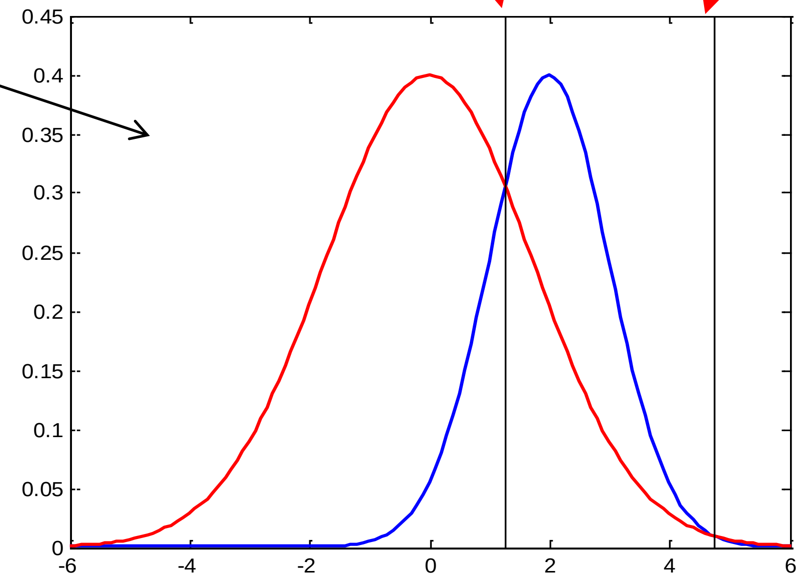
Assume equal priors, the minimum error decision rule is shown here:




Now if we assume these costs:

$\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 1$, $\lambda_{21} = \text{sqrt}(3)$

$$\frac{\frac{1}{\sqrt{2\pi 3}}\exp\left(-0.5\frac{x^2}{3}\right)}{\frac{1}{\sqrt{2\pi}}\exp\left(-0.5(x-2)^2\right)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{1}{\sqrt{3}}$$




Take ln on

both sides of eqn.

$$-\frac{1}{2}\frac{x^2}{3} + \frac{1}{2}(x-2)^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$x^2 - 12x + 12 \quad 0$$

# Minimum Bayes Risk Classifier – General Case

$\{\omega_1, \omega_2, \omega_3, \omega_c\}$ Set of Classes

$\{\alpha_1, \alpha_2, \alpha_3, \alpha_a\}$ Set of possible actions (assigning a region)

$\{\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{ac}\}$ Loss associated with each action

- Note that a need not equal c. We can make different number of actions than the number of classes. We could for example reject, then a = c + 1

- The Conditional Risk is given by:

$$l^i(x) = R\left(\alpha_i \mid x\right) = \sum_{j=1}^{c} \lambda_{ij} P\left(\omega_j \mid x\right) = \sum_{j=1}^{c} \lambda_{ij} \frac{P\left(x \mid \omega_j\right) P\left(\omega_j\right)}{P(x)}$$

  – It is the expected loss for taking action $\alpha_i$

- The overall risk over region $\Omega_i$ is :

$$r^i = \int_{\Omega_i} l^i(x) p(x) dx = \int_{\Omega_i} \sum_{j=1}^{c} \lambda_{ij} P\left(\omega_j \mid x\right) p(x) dx = \int_{\Omega_i} \sum_{j=1}^{c} \lambda_{ij} P\left(x \mid \omega_j\right) P\left(\omega_j\right) dx$$

What we must minimize

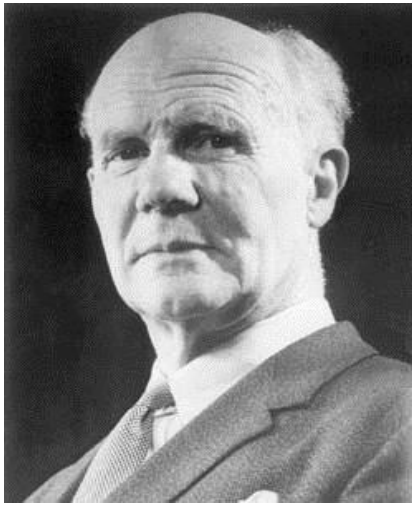
# So Far…

- Bayesian approach with overall risk and individual losses
- In practice it may be very hard to assign losses to classifications. Losses could be combination of several different factors measured in different units.
- What if we need to ensure that a certain of error has to be bounded? What is misclassifying class 1 is much worse than misclassifying class 2 and must be bounded?

# Neyman-Pearson Criterion

Jerzy Neyman

(1894-1981)



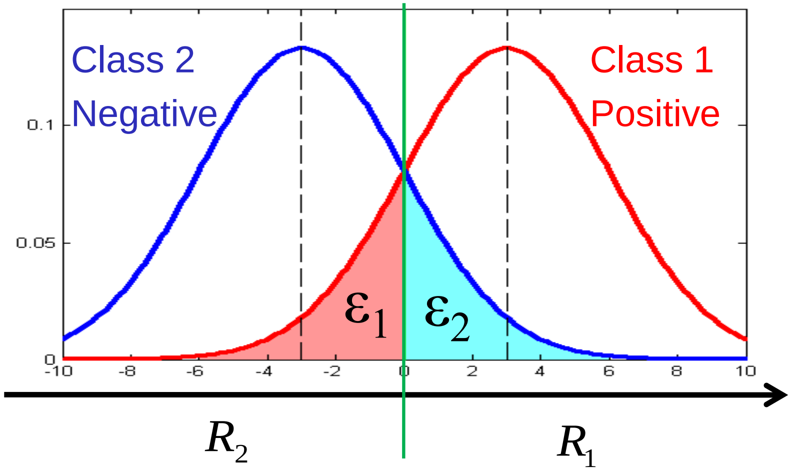
Egon Pearson

(1895-1980)

This criterion fixes one class error and seeks to minimize the other. Extensively used in Detection and Estimation Theory, where the problem is to detect a signal in the presence of noise.

For example, there may be a government regulation that not more than 1% of salmon should be misclassified as sea bass

# Neyman-Pearson Criterion




$$\varepsilon_1 = \int_{R_2} P(x \mid \omega_1)dx = \text{error probability of Type I}$$ **False Rejection (Negative) Rate (FRR)**

$$\varepsilon_2 = \int_{R_1} P(x \mid \omega_2)dx = \text{error probability of Type II}$$ **False Acceptance (Positive) Rate (FAR)**

Let's fix $\varepsilon_2 = \varepsilon_0$ and aim at minimizing $\varepsilon_1$

Assuming Equal Priors For Both Classes

# Neyman-Pearson Criterion (contd.)

We want to minimize $\int_{R_2} P(x \mid \omega_1)dx$ subject to the constraint $\int_{R_1} P(x \mid \omega_2)dx = \varepsilon_0$

Lagrange Multiplier - Regularization

$$r = \int_{R_2} P(x \mid \omega_1)dx + \mu\left\{\int_{R_1} P(x \mid \omega_2)dx - \varepsilon_0\right\}$$

$$= 1 - \int_{R_1} P(x \mid \omega_1)dx + \int_{R_1} \mu P(x \mid \omega_2)dx - \mu\varepsilon_0$$

$$= (1 - \mu\varepsilon_0) + \int_{R_1} \{\mu P(x \mid \omega_2)dx - P(x \mid \omega_1)dx\}$$

Must be made as small as possible

$$\mu P(x \mid \omega_2) - P(x \mid \omega_1) < 0$$

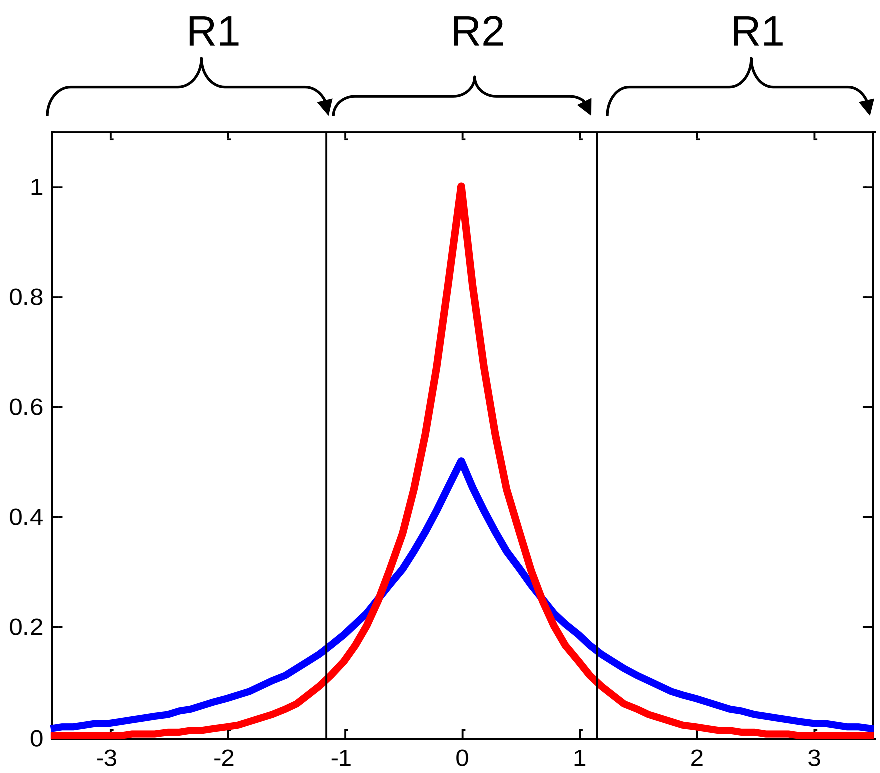
$$\rightarrow \boxed{\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu}$$

# Neyman-Pearson Criterion Example

Note that there is no dependence on the a priori probabilities

In general, μ cannot be determined analytically and requires numerical computation

For example, consider this 2 class problem and hold $\varepsilon_2 = \varepsilon_0 = 0.1$

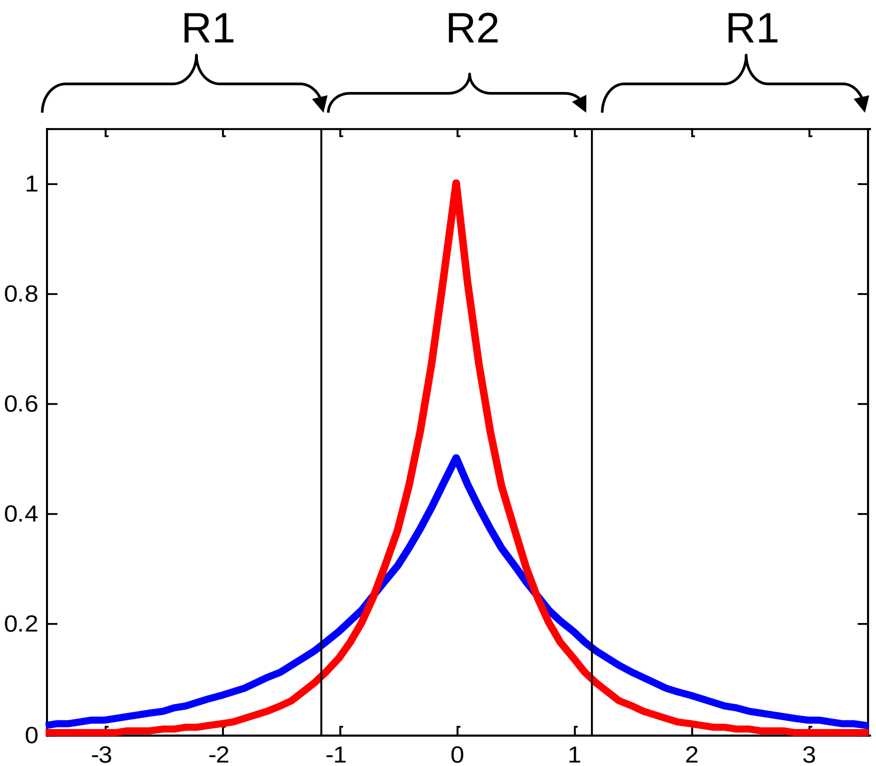



$$\int_{R_1} P(x \mid \omega_2)dx = \varepsilon_0$$

$$f\left(x \mid \omega_1\right) = 0.5e^{-|x|}$$

$$f\left(x \mid \omega_2\right) = e^{-2|x|}$$

# Neyman-Pearson Criterion Example




$$f(x\,|\,\omega_1) = 0.5e^{-|x|}$$

$$f(x\,|\,\omega_2) = e^{-2|x|}$$

$$\int_{R_1} P(x\,|\,\omega_2)dx = \varepsilon_0$$

$$\frac{f(x\,|\,\omega_1)}{f(x\,|\,\omega_2)} = 0.5e^{|x|} \mathop{\gtrless}_{\omega_2}^{\omega_1} \mu$$

$$|x| \mathop{\gtrless}_{\omega_2}^{\omega_1} \ln(2\mu) = T$$

Determine $T$ such that $\varepsilon_2 = 0.1$:

$$\varepsilon_2 = \int_{R_1} f(x\,|\,\omega_2)dx = 2\int_T^{\infty} e^{-2x}dx = -e^{-2x}\Big|_T^{\infty} = e^{-2T}$$

$$T = 0.5\ln(10) = 1.15$$

Then, the desired decision strategy is:

$$\boxed{|x| \mathop{\gtrless}_{\omega_2}^{\omega_1} 1.15}$$

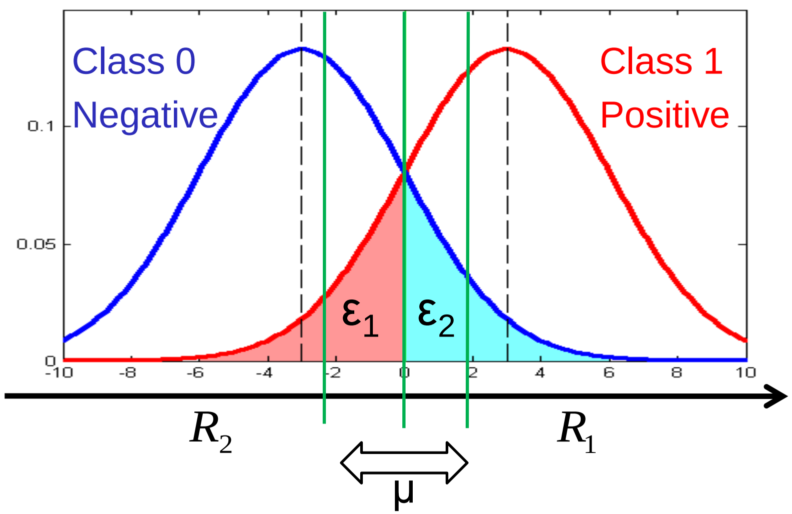
We can also calculate $\varepsilon_1 = \int_{R_2} f(x\,|\,\omega_1)dx$

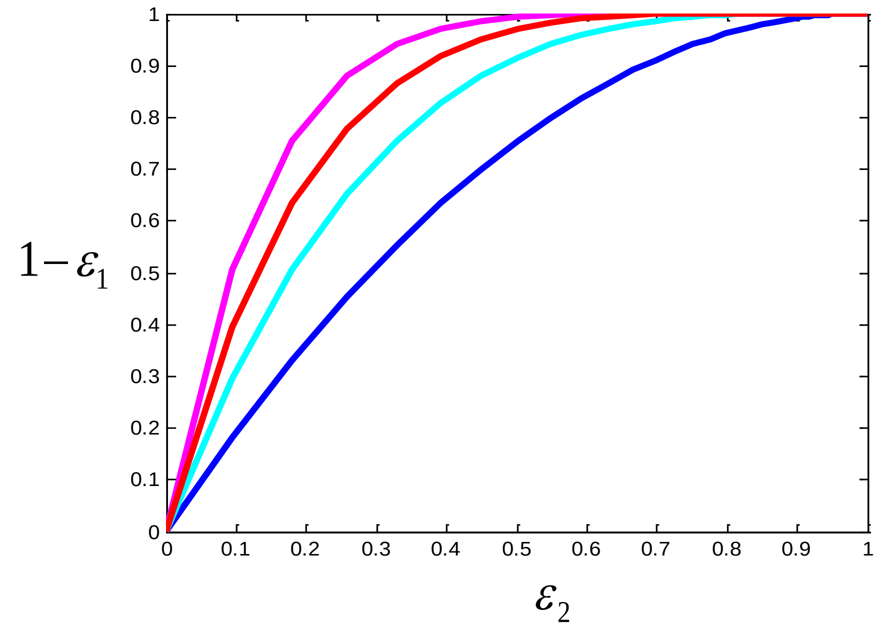
$$= 2\int_0^T 0.5e^{-x}dx = 1 - e^{-T} = 0.68$$

Electrical & Computer ENGINEERING

# Receiver Operating Characteristic Curve

The performance of the decision rule is summarized in an ROC curve, which plots the true positive against the false positive as the threshold (μ) is varied.




$$\varepsilon_1 = \int_{R_2} P(x \mid \omega_1)\,dx = \text{error probability of Type I}$$

False Rejection (Negative) Rate (FRR)

$$\varepsilon_2 = \int_{R_1} P(x \mid \omega_2)\,dx = \text{error probability of Type II}$$

False Acceptance (Positive) Rate (FAR)

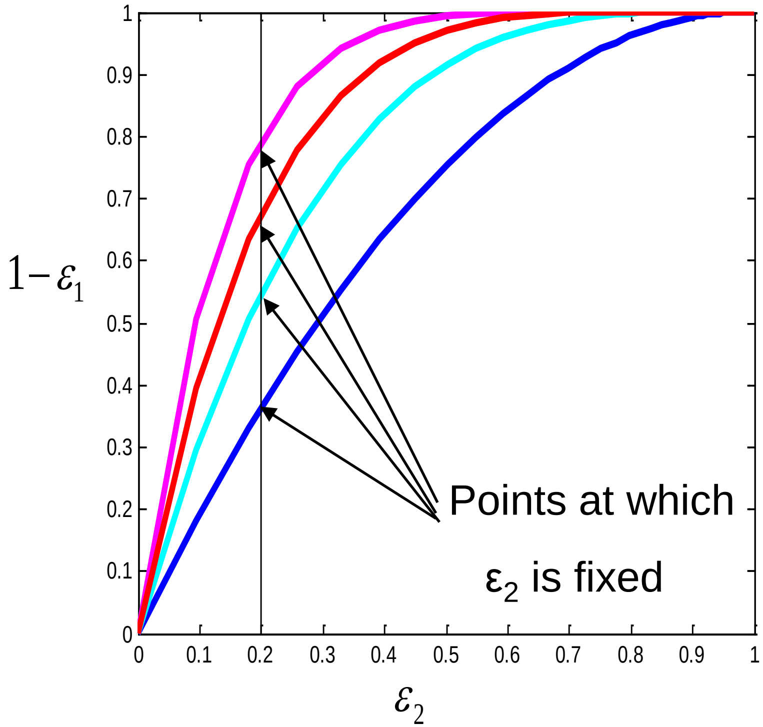
True Rejection (Negative) Rate (TRR) = 1-FAR = 1-$\varepsilon_2$

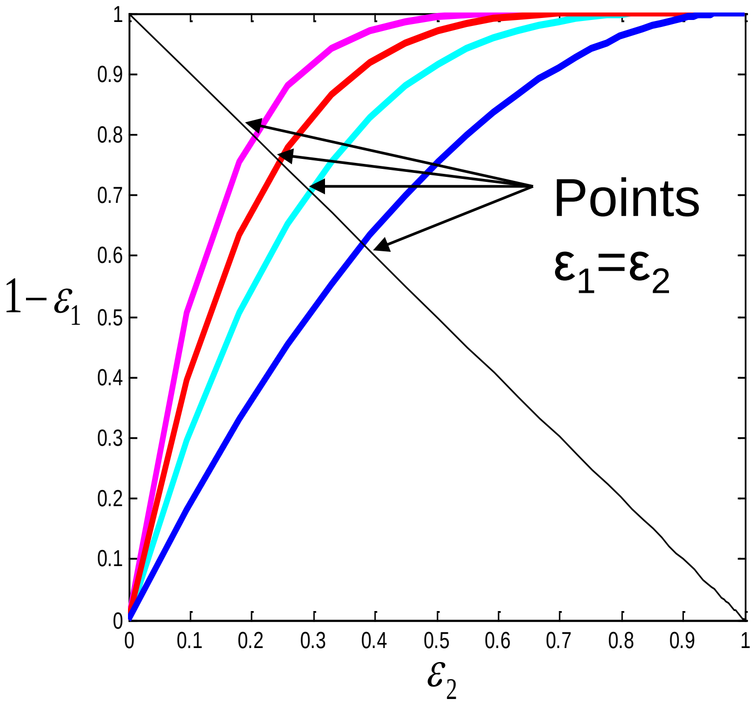
True Acceptance (Positive) Rate = 1-FRR = 1-$\varepsilon_1$

Electrical & Computer ENGINEERING

# Receiver Operating Characteristic Curve

The performance of the decision rule is summarized in an ROC curve, which plots the true positive against the false positive as the threshold μ is varied.

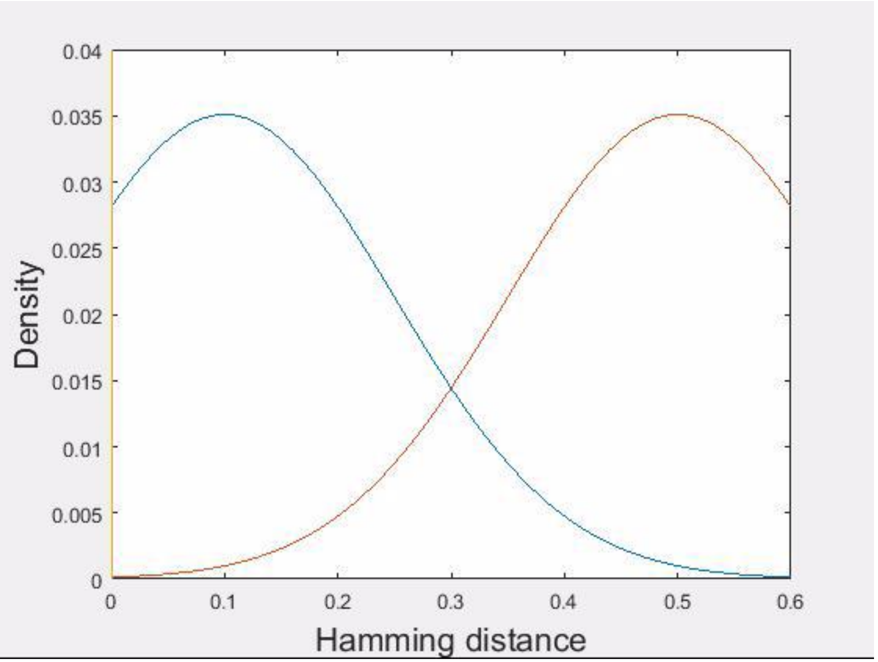


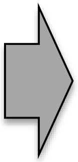



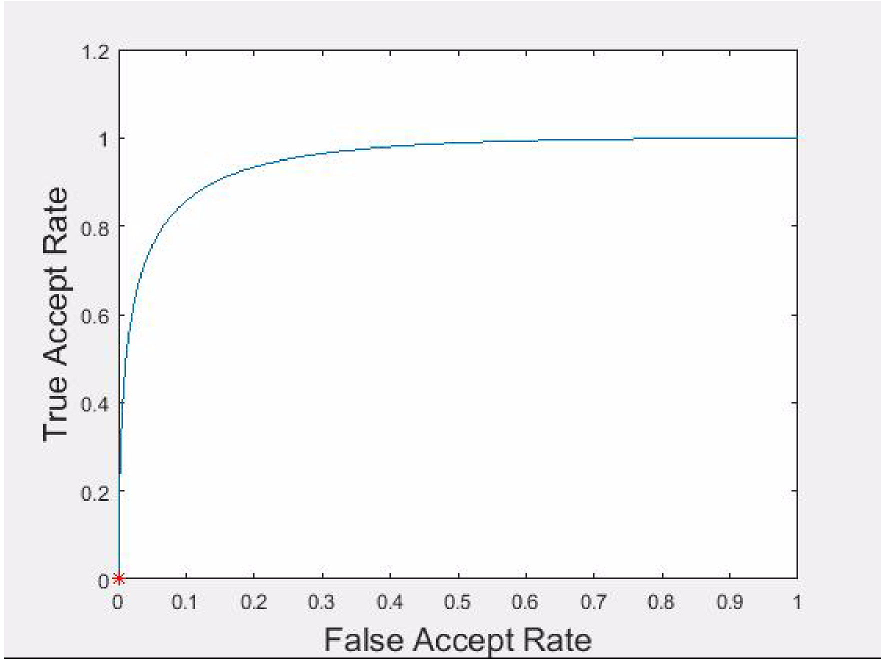

Equal Error Rate (EER)

Most popular way to measure performance

# As long as we move the decision threshold from the right to the left, the true accept rate will increase besides the false accept rate





# Summary of Decision Strategies

- Bayes Decision Rule / Minimum Error Classification

$$P(\omega_1)P(x|\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2)P(x|\omega_2) \quad \text{or} \quad P(\omega_1|x) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2|x)$$

- Minimum Cost/Risk Classification

$$(\lambda_{21}-\lambda_{11})P(\omega_1)P(x|\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (\lambda_{12}-\lambda_{22})P(\omega_2)P(x|\omega_2)$$

- Neyman-Pearson Criterion
  – Fixes one class error probability and seeks to minimize the other.