Introduction to Signal Detection and Classification

Phani Chavali

Outline

- Detection Problem
- Performance Measures
- Receiver Operating Characteristics (ROC)
- F-Test
- R^2 - Test
- Linear Discriminant Analysis (LDA)

Statistical Detection Problem

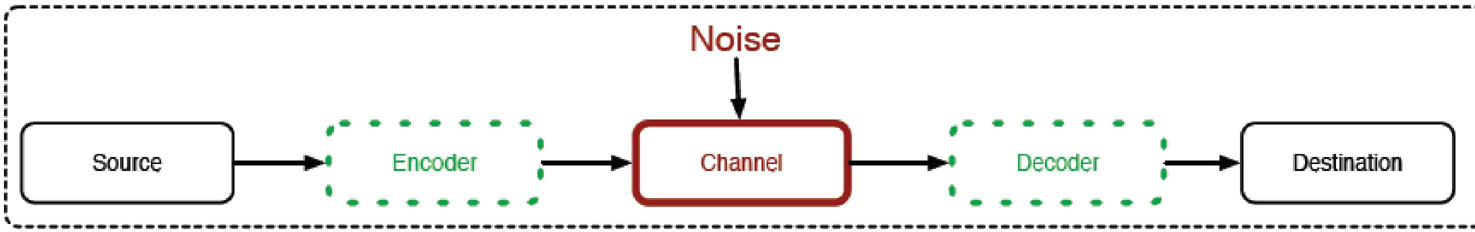
- The classical detection problem is to decide which of the many hypothesis is true using the noisy observations.

- Binary hypothesis testing

$$\mathcal{H}_0 \text{ or } \mathcal{H}_1?$$

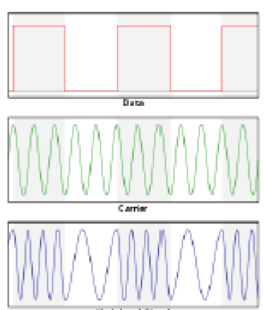
Examples

Detection example 1: digital communications



10001010100010

$$0 \leftrightarrow s_0(t) = \sin(\omega_0 t)$$
$$1 \leftrightarrow s_1(t) = \sin(\omega_1 t)$$



$$r(t) = \begin{cases} s_0(t) + n(t) & \text{if '0' sent} \\ s_1(t) + n(t) & \text{if '1' sent} \end{cases}$$

Detect?

Examples

Detection example 2: Radar communication

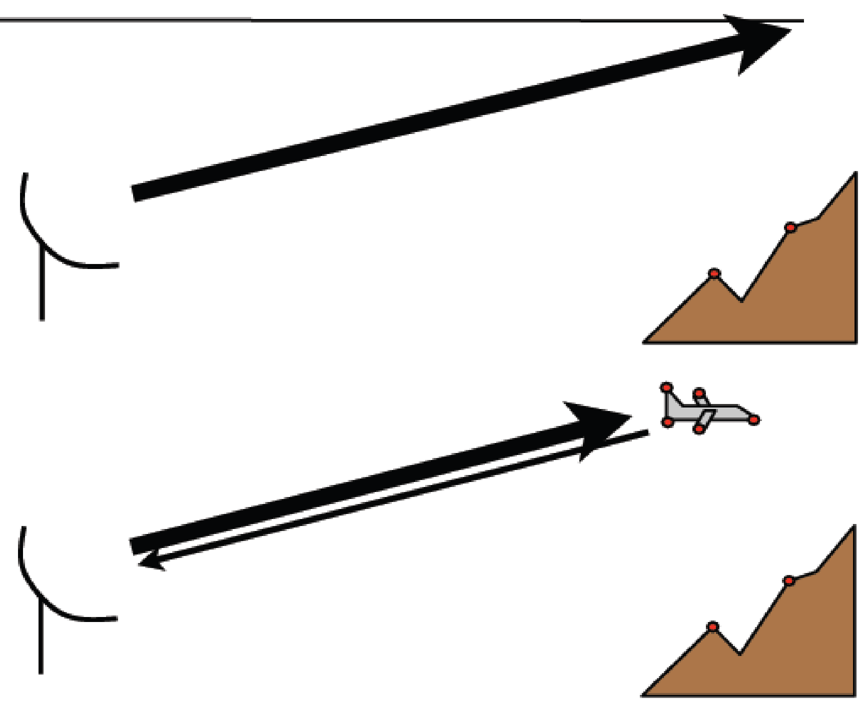
Send $s(t) = \sin(\omega_c t), 0 \leq t \leq T$

Receive

Hypothesis $\mathcal{H}_0$

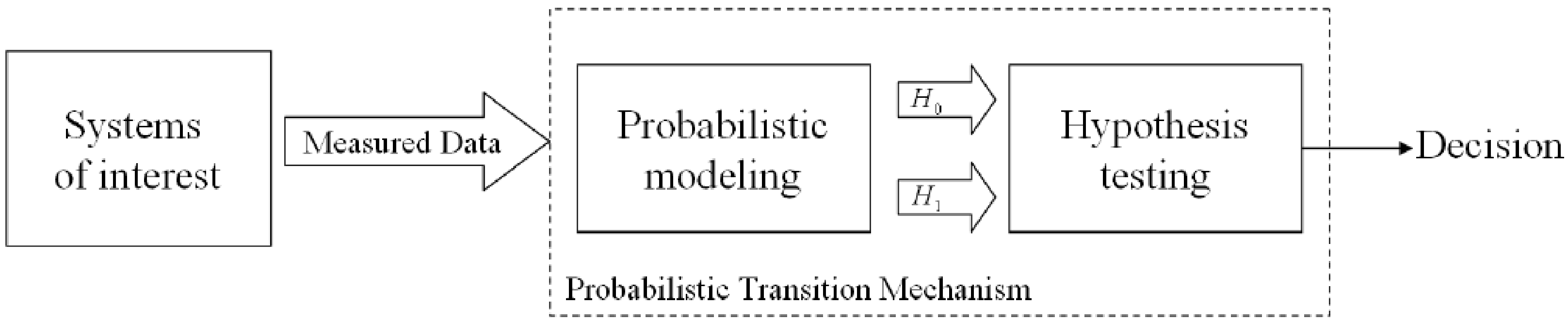
$$r(t) = n(t),\ 0 \leq t \leq T$$

Detect?



Hypothesis $\mathcal{H}_1$

$$r(t) = V_r \sin((\omega_c + \omega_d)(t - \tau) + \theta_r) + n(t),\ \tau \leq t \leq t + \tau$$



Block diagram of statistical detection.

Likelihood ratio test

- Compute the likelihood ratio under the two hypothesis.
- Compare the likelihood ratio to a threshold.
- If the ratio is greater than the threshold, choose the alternate hypothesis; else choose the null hypothesis.

$$\underbrace{\Lambda(\boldsymbol{x})}_{\text{likelihood ratio}} = \frac{p(\boldsymbol{x}\,|\,\theta_1)}{p(\boldsymbol{x}\,|\,\theta_0)} \overset{\mathcal{H}_1}{\gtrless} \frac{\pi_0\,\mathrm{L}(1|0)}{\pi_1\,\mathrm{L}(0|1)} \equiv \tau$$

- Likelihood ratio is called as test-statistic.

Performance Analysis

- In binary hypothesis testing, the following events are possible:

Correct Decision	*Error Decision*
Decide H_0 when H_0 is true	Decide H_1 when H_0 is true
Decide H_1 when H_1 is true	Decide H_0 when H_1 is true

- Accordingly, we assign four probabilities:

$P(\mathrm{H}_0 \vert \mathrm{H}_0)$	$P(\mathrm{H}_1 \vert \mathrm{H}_0)$
$P(\mathrm{H}_1 \vert \mathrm{H}_1)$	$P(\mathrm{H}_0 \vert \mathrm{H}_1)$

Performance Analysis

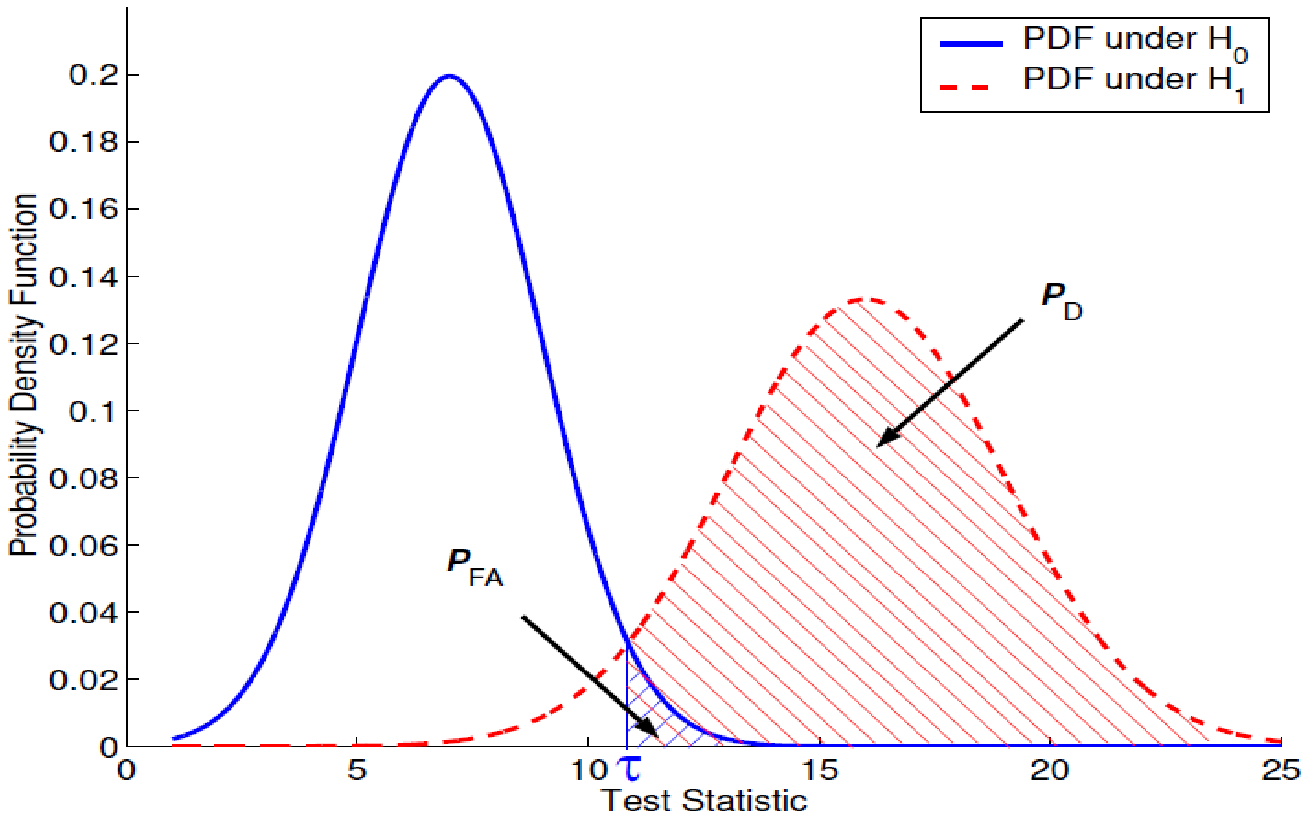
- Performance is determined by two kinds of probabilities.
  - Probability of false alarm

$$P_{\text{FA}} = P[\overbrace{\underbrace{\text{test statistic} > \tau}_{\boldsymbol{X} \in \mathcal{X}_1}}^{\Lambda(x)} \mid \theta = \theta_0]$$

  - Probability of miss

$$P_M = P[\text{test statistic} < \tau \mid \theta = \theta_1]$$

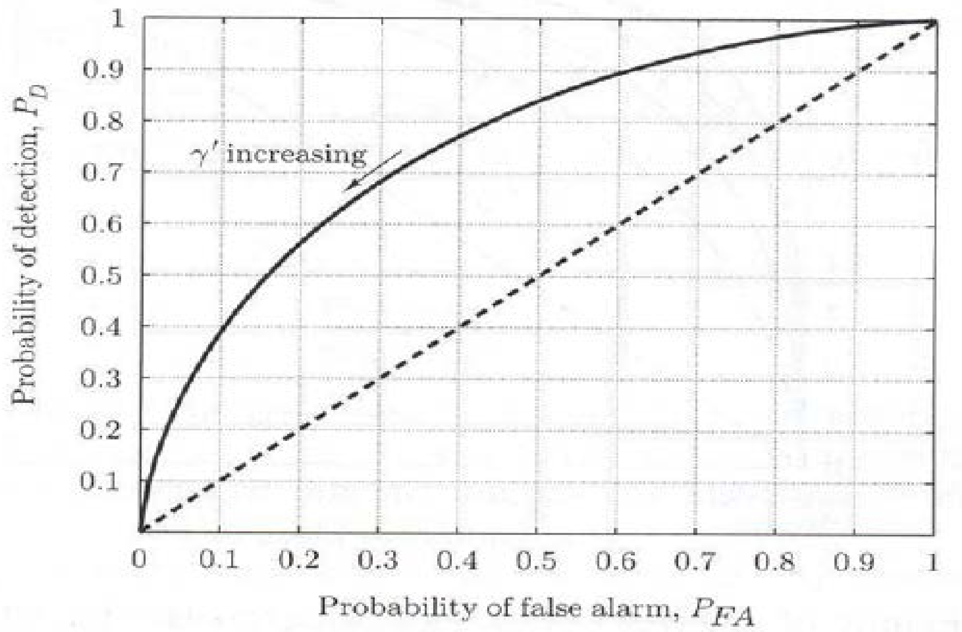
$$P_{\text{D}} = P[\underbrace{\text{test statistic} > \tau}_{\boldsymbol{X} \in \mathcal{X}_0} \mid \theta = \theta_1].$$



We cannot minimize both probabilities simultaneously

Receiver operating characteristic

- Graph of probability of detection Vs the probability of false alarm.



- In this class, we plot the graphs as a function of the threshold.

F - test

- F-Test is a statistical test where the test statistic follows an F-distribution.
- Most commonly used scenario:
  - The observations come from several distributions.
  - Null hypothesis corresponds to the case where means of all the distributions are same.
  - Alternate hypothesis corresponds to the case where the at least one mean is different.

How to do an F-test

- Compute the F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}}$$

- Compare with a threshold.

$$F > \tau, \text{ choose } \mathcal{H}_1$$

$$F \leq \tau, \text{ choose } \mathcal{H}_0$$

- Between group variability refers to the variability between various groups.
- Within-group variability refers to variability in each group

R^2 - Test

- R^2 - Test is similar to the F-test and it is used in models whose purpose is to predict the future values using the observed values.

$$R^2 = \frac{\text{between-group variability}}{\text{total variability}}$$

Classification

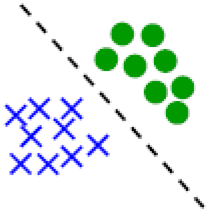
- Classification is the assignment of the data to one of pre-defined classes.
- Given noisy observations, we need to classify them into one of the several classes.
- Statistically, classification is equivalent to M-ary hypothesis test.

Methods

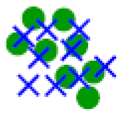
- There are many procedures to do this.
  - Linear Discriminant Analysis (LDA)
  - K-NN
  - Logistic and probit regression
  - Support Vector Machines and many more…
- We will learn about LDA today.

Features

- Given the observation, find a feature (similar to the test-statistic or the likelihood ratio) that separates various classes.
- The goal is to find which is the best feature.
  - Examples from the same class should have similar feature values
  - Examples from different classes have different feature values



“Good” features



“Bad” features

More Properties

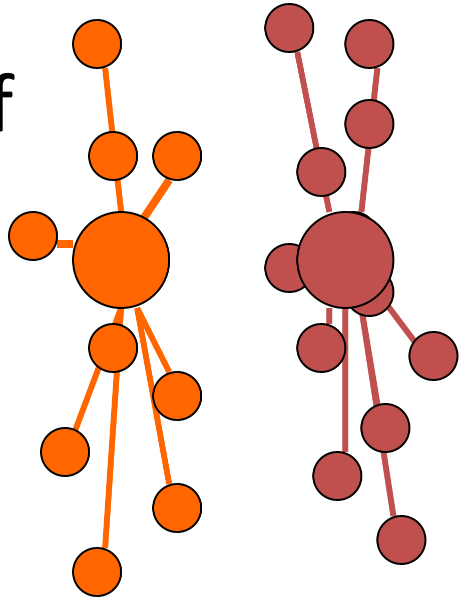
Linear separability *Non-linear separability* *Highly correlated features* *Multi-modal*

Some times it is better to select several features, instead of just one.

Linear discriminant analysis

What class information may be useful?

- Between-class distance
  - Distance between the centroids of different classes
- Within-class distance
  - Accumulated distance of an instance to the centroid of its class



How it works ?

- Linear discriminant analysis (LDA) finds most discriminant projection by maximizing between-class distance and minimizing within-class distance

Some Theory (1)

- Assume we have a set of D-dimensional samples $(x1, x2, \dots xN)$, $N1$ of which belong to class $\omega1$ and $N2$ to class $\omega2$
- We seek to obtain a scalar y by projecting the samples x onto a line

$$y = w^T x$$

- Of all the possible lines we would like to select the one that maximizes the separability of the scalars

Some Theory(2)

- In order to find a good projection vector, we need to define a measure of separation
- The mean vector of each class in x-space and

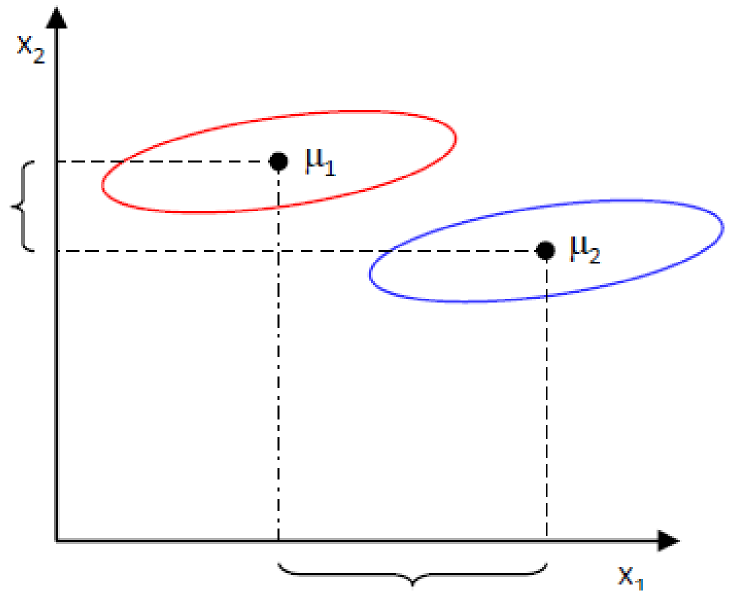
$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x$ and $\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in \omega_i} y = \frac{1}{N_i} \sum_{x \in \omega_i} w^T x = w^T \mu_i$

- We could then choose the distance between the projected means as our objective function

$$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T(\mu_1 - \mu_2)|$$

Some Theory(3)

- However, the distance between projected means is not a good measure since it does not account for the standard deviation within classes



Some Theory(4)

- We use the difference between the means, normalized by a measure of the within-class variance
- For each class we define the scatter, an equivalent of the variance, as

$$\tilde{s}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2$$

- The linear discriminant is defined as the linear function that maximizes the criterion function

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Some Theory(5)

- Therfore, we are finding most discriminant projection by maximizing between-class distance and minimizing within-class distance