Learning Objectives:

- Hypothesis testing
- The receiver operator characteristic (ROC) curve
- Bayes’s Theorem, positive and negative predictive value (PPV, NPV)
- Basic principles of supervised networks

Supplementary Slides: “Probability.”
Supplementary Reading: “Fundamental of Neural Networks.”
http://www.greylodge.org/occultreview/gl0_008/Fundamentals_of_Neural_Networks.pdf

I. Hypothesis Testing

A. Reference (or “Parent”) distribution

1. Typically a Gaussian distribution assuming sufficient sampling of an unbiased random variable:

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \text{ and standard deviation,} \\
\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x - \bar{x})^2}.
\]
Lecture 18: Noise modeling and introduction to decision theory

C. Test or experiment must define a threshold for significance.

1. Typically state a Null hypothesis: $H_0$.

$H_0 \equiv$ Presume that the sampled distribution and the reference distribution are one and the same.

Only reject the Null hypothesis when the sample mean, $\bar{x}$, exceeds the significance level of the test, $\alpha$.

For a two-tailed test, the threshold could be met by $\bar{x}$ being greater than or less than the threshold. Thus, the value of $\alpha$ defines the probabilities falsely rejecting the Null hypothesis:

$$\Pr(\bar{x} \leq \phi_{\frac{\alpha}{2}}) = \int_{-\infty}^{\phi_{\frac{\alpha}{2}}} p(\phi)d\phi$$

(1a)

$$\Pr(\bar{x} \geq \phi_{\frac{\alpha}{2}}) = \int_{\phi_{\frac{\alpha}{2}}}^{\infty} p(\phi)d\phi$$

(1b)

For a one-tailed test,

$$\Pr(\bar{x} \geq \phi_{-\alpha}) = \int_{\phi_{-\alpha}}^{\infty} p(\phi)d\phi$$

(2)

D. Type I and Type II errors.

Type I error $\equiv$ The probability that $H_0$ is rejected if it is correct - defined by Eqs. 1a and b.

Type II error $\equiv$ The probability that $H_0$ is accepted when it is false.

Need to identify the underlying parent distribution of the sampled, or measured, data:

$$\beta = \int_{-\infty}^{\phi_{-\alpha}} p(\phi)d\phi \equiv \text{Type II error}.$$ 

Power $\equiv 1-\beta$; a measure of our ability to resolve two distinct underlying distributions.
Lecture 18: Noise modeling and introduction to decision theory

D. The error function:

\[ \text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy \]

For a normalized Gaussian, then perform a change of variable to a new variable, \( z \), such that,

\[ z = \frac{x - \mu}{\sigma} \]

The distribution of \( z \) is zero mean with standard deviation, \( \sigma = 1 \).

Then for the one-tailed case,

\[ \text{erf}(z_{1-\alpha}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{1-\alpha}} e^{-y^2/2} \, dy = 1 - \Pr(\bar{x} > z_{1-\alpha}) \]

\( \Rightarrow \) Type I error = \( 1 - \text{erf}(z_{1-\alpha}) \)

\[ \beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{1-\alpha}} e^{-y^2/2} \, dy = 1 - \Pr(\bar{x} > z_{1-\alpha}) \]

\( \Rightarrow \) Type II error = \( \text{erf}(z_{1-\alpha} - d) \).

\( \text{Power} = 1 - \beta = \text{erf}(d - z_{1-\alpha}) \)
Lecture 18: Noise modeling and introduction to decision theory

II. Receiver operator characteristic (ROC) curve

A. Constructing the curve

1. Pr(n) and Pr(s):

   \[ k_o = d \text{ in units of standard deviations} \]

   \[ k_i, \text{ or “k-threshold”} = Z_{1-\alpha} \]

2. Type I error – Probability of detecting signal when signal is absent = \( P(S|n) \)

   \[ = 1 - Pr(N|n) = 1 - \text{Specificity} = 1 - \text{erf}(k_i) = \text{erf}(-k_i). \]

3. Type II error – Probability of detecting noise when signal is present = \( P(N|s) = \beta \)

   \[ = 1 - Pr(S|s) = 1 - \text{Sensitivity} \]

   \[ \Rightarrow \text{Sensitivity} = 1 - \beta = \text{Power} = \text{erf}(k_o - k_i) \]


   Sensitivity \( \equiv \) True Positive Fraction = \( \frac{TP}{TP + FN} = Pr(S|s) \)

   Specificity \( \equiv \) True Negative fraction = \( \frac{TN}{TN + FP} = Pr(N|n) \)
Lecture 18: Noise modeling and introduction to decision theory

4. Positive and Negative Predictive Values.

Positive Predictive Value \( \equiv \) The probability that a patient has the disease given a positive test results \( = \Pr(s|S) \)
Negative Predictive Value \( \equiv \) The probability that a patient has the disease given a positive test results \( = \Pr(n|N) \)

Use Bayes’s Theorem:

\[
\Pr(S|s) = \frac{\Pr(S|s) \Pr(s)}{\Pr(S|s) \Pr(s) + \Pr(S|n) \Pr(n)} = PPV.
\]

III. Introduction to Networks.

1. Neural Networks: Concept based on basic features of biological nervous systems.

Interconnecting identical nodes, or processing elements (PE’s). The PE’s are organized into layers. Number of PE’s per layer is a design choice based on how many inputs are desired and their correlation with each other. Interconnections between PE’s is also a design choice. The final layer is the output layer and other layers are “hidden” layers.

![Figure 2.2: Basic Elements of an Artificial Neuron.](http://www.greylodge.org/occultreview/glor_008/Fundamentals_of_Neural_Networks.pdf)


2. Processing Elements

Output, \( O \), of each PE:

\[
O = g[X^T \cdot W] = g \left[ \sum_{i=1}^{N} x_i w_i \right] = g[S]
\]

\( O \) is a scaler, \( X \) is the input vector, and \( W \) is the weight vector associated with a given PE. The weights are adjusted by the training process. \( g[\cdot] \) is a nonlinear activation function, generally with a sigmoid dependence, as shown above in the Figure 2.2 from Zhilouchian’s chapter.
During training, feature vectors of known objects are presented in random order to the network. The interconnection weights of the PE’s are adjusted for each iteration of the training. Presumably, the training converges on a set of weights that identifies the appropriate set of measures and their relative importance to making the decision, i.e. malignant or benign in the cancer setting.

3. Bayesian Networks

Weighting factors derived from the predictive value of a set of given variables. Recognition that many factors contribute to diagnosis:

(Burnside et al, “A Bayesian Network for Mammography”)

![](Image)