EEE 598C: Statistical Pattern Recognition
Lecture Note 3: Sequential Decision Theory

Darryl Morrell

September 5, 1996

1 Introduction

We consider the two-category classification problem, in which there are two states of nature ω_1 and ω_2, and two possible actions α_1 and α_2; α_j is the action of choosing ω_j. For the sequential decision problem, we assume that the conditional densities $f(\mathbf{x}|\omega_1)$ and $f(\mathbf{x}|\omega_2)$ are known, but we do not use any prior probabilities nor do we use a loss function.

We will continue to use missed detection and false alarm to denote the two types of errors that can be made in the decision process; again, their interpretation depends on the particular application.

In sequential detection, we consider successively more elements of the feature vector until a decision can be made. Situations in which this might be the case include:

- We direct a sensor at an object and take repeated measurements until the object is recognized.
- We start attempting to do classification using basic, easily computed features. If a decision cannot be made, then other features, which typically are more computationally intensive, are included into the decision process.

Assuming that enough feature values are available, sequential detection allows us to achieve low probabilities of both false alarm and missed detection.

2 Sequential Probability Ratio Test

We will consider a Sequential Probability Ratio Test (SPRT) that is optimal in the sense that no other test that achieves the desired probabilities of error does so using, on average, fewer features. Denote by $\mathbf{x}_d$ the first d elements of the feature vector $\mathbf{x}$:

$$\mathbf{x}_d = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}^T$$

The SPRT is the following:

For two constant thresholds A and B, if

$$\frac{f(\mathbf{x}_d|\omega_1)}{f(\mathbf{x}_d|\omega_2)} > A$$

choose α_1; if

$$\frac{f(\mathbf{x}_d|\omega_1)}{f(\mathbf{x}_d|\omega_2)} < B$$

choose α_2; otherwise, get another feature x_{d+1} and repeat the test with $\mathbf{x}_{d+1}$.

Exercise 1 *Let the class-conditional densities of* $\mathbf{X}_d$ *be*

$$f(\mathbf{x}_d|\omega_1) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\sum_{i=1}^{d} x_i^2}$$

$$f(\mathbf{x}_d|\omega_2) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\sum_{i=1}^{d} (x_i-1)^2}$$

Find the SPRT test for this case, and interpret your results.

How do we choose the thresholds A and B to obtain the desired P_{FA} and P_{MD}? To find A, we use the structure of the decision rule to choose α_1:

$$f(\mathbf{x}_d|\omega_1) \geq A f(\mathbf{x}_d|\omega_2) \tag{1}$$

In order to choose α_1 for some value of d, (1) must be satisfied. If we integrate (1) over the region $\mathcal{R}_1$, we obtain the following inequality:

$$\int_{\mathcal{R}_1} f(\mathbf{x}_d|\omega_1) d\mathbf{x}_d \geq A \int_{\mathcal{R}_1} f(\mathbf{x}_d|\omega_2) d\mathbf{x}_d$$

$$(1 - P_{FA}) \geq A P_{MD}$$

$$A \leq \frac{1 - P_{FA}}{P_{MD}} \tag{2}$$

This equation gives an upper bound for A given the desired values of P_{FA} and P_{MD}. Typically, we assume that the inequality in (1) is approximately equal; this assumption will be valid if a large value of d is necessary to make the decision. In this case, then (2) can be replaced with an equality.

Exercise 2 *What is the effect on* P_{FA} *of assuming that (2) is an equality?*

Using a similar approach, we find the following expression for B:

$$B \geq \frac{P_{FA}}{1 - P_{MD}}$$

3 Recursive Computation of the SRPT

In many situations, the computation of the likelihood ratio in the SRPT can be made recursive. To see this, we use the chain rule for densities:

$$f(\mathbf{x}_d|\omega_i) = f(x_1|\omega_i) \prod_{j=2}^{d} f(x_j|\mathbf{x}_{j-1}, \omega_i)$$

The likelihood ratio becomes

$$\frac{f(x_1|\omega_1) \prod_{j=2}^{d} f(x_j|\mathbf{x}_{j-1}, \omega_1)}{f(x_1|\omega_2) \prod_{j=2}^{d} f(x_j|\mathbf{x}_{j-1}, \omega_2)} = \frac{f(x_1|\omega_1)}{f(x_1|\omega_2)} \prod_{j=2}^{d} \frac{f(x_j|\mathbf{x}_{j-1}, \omega_1)}{f(x_j|\mathbf{x}_{j-1}, \omega_2)}$$

Taking the logarithm of this likelihood ratio, we have

$$\ln \frac{f(\mathbf{x}_d|\omega_1}{f(\mathbf{x}_d|\omega_2} = \ln \frac{f(x_1|\omega_1)}{f(x_1|\omega_2)} + \sum_{j=2}^{d} \ln \frac{f(x_j|\mathbf{x}_{j-1}, \omega_1)}{f(x_j|\mathbf{x}_{j-1}, \omega_2)} \tag{3}$$

This log-likelihood ratio can be computed recursively; to incorporate feature x_d, the logarithm of the ratio of the conditional densities is added to the sum.

In the special case where the elements of $\mathbf{x}_d$ are independent, (3) becomes

$$\ln \frac{f(\mathbf{x}_d|\omega_1}{f(\mathbf{x}_d|\omega_2} = \sum_{j=1}^{d} \ln \frac{f(x_j|\omega_1)}{f(x_j|\omega_2)} \tag{4}$$

Exercise 3 *For the densities in Exercise 1, find the recursive formulation (4) of the log-likelihood ratio.*

4 Performance Analysis

Let D_0 be the number of features needed to make a decision using the SPRT. Note that D_0 is a random variable. We characterize the performance of the SPRT with the quantities $E[D_0|\omega_1]$ and $E[D_0|\omega_2]$.

For this analysis, we will assume that the elements of $\mathbf{X}_d$ are independent and identically distributed. We use the following notation:

$$\delta^{(i)} = E\left[\ln \frac{f(X_j|\omega_1)}{f(X_j|\omega_2)}\middle|\,\omega_i\right]$$

Fixing the number of features in the log-likelihood ratio at d and taking the expected value gives

$$\begin{aligned} E\left[\ln \frac{f(\mathbf{x}_d|\omega_1}{f(\mathbf{x}_d|\omega_2}\middle|\,\omega_i\right] &= E\left[\sum_{j=1}^{d} \ln \frac{f(x_j|\omega_1)}{f(x_j|\omega_2)}\middle|\,\omega_i\right] \\ &= \sum_{j=1}^{d} E\left[\ln \frac{f(x_j|\omega_1)}{f(x_j|\omega_2)}\middle|\,\omega_i\right] \\ &= d\delta^{(i)} \end{aligned}$$

The number of features necessary to make a decision is D_0, which is a random variable.

$$E\left[\ln \frac{f(\mathbf{x}_{D_0}|\omega_1}{f(\mathbf{x}_{D_0}|\omega_2}\middle|\omega_i\right] = E\left[D_0|\omega_i\right]\delta^{(i)} \tag{5}$$

We now find $E\left[\ln \frac{f(\mathbf{x}_{D_0}|\omega_1}{f(\mathbf{x}_{D_0}|\omega_2}|\omega_i\right]$. Note that when the decision rule stops, the log likelihood ratio is either approximately equal to $\ln A$ or to $\ln B$. If ω_1 is the true state of nature, then the probability that the log-likelihood ratio is approximately $\ln A$ is the probability of a correct decision, which is $1-P_{FA}$, and the probability the the log-likelihood ratio is approximately $\ln B$ is P_{FA}. Similarly, if ω_2 is the true state of nature, then the probability that the log-likelihood ratio is approximately $\ln A$ is P_{MD}, and the probability that the log-likelihood ratio is approximately $\ln B$ is $1 - P_{MD}$. Thus, the expected value of the log-likelihood ratio is

$$E\left[\ln \frac{f(\mathbf{x}_{D_0}|\omega_1}{f(\mathbf{x}_{D_0}|\omega_2}\middle|\omega_1\right] = (1 - P_{FA})\ln A + P_{FA}\ln B$$

$$E\left[\ln \frac{f(\mathbf{x}_{D_0}|\omega_2}{f(\mathbf{x}_{D_0}|\omega_2}\middle|\omega_1\right] = P_{MD}\ln A + (1 - P_{MD})\ln B$$

Using these results to solve for $E\left[D_0|\omega_i\right]$ in (5) gives the following:

$$E\left[D_0|\omega_1\right] = \frac{(1 - P_{FA})\ln A + P_{FA}\ln B}{\delta^{(1)}}$$

$$E\left[D_0|\omega_2\right] = \frac{P_{MD}\ln A + (1 - P_{MD})\ln B}{\delta^{(2)}}$$

Exercise 4 *Find $E\left[D_0|\omega_0\right]$ and $E\left[D_0|\omega_1\right]$ for the densities in Exercise 1.*