EEE 598C: Statistical Pattern Recognition
Lecture Note 1: Bayesian Decision Theory

Darryl Morrell

August 27, 1996

1 Introduction

The state of nature is denoted Ω; it can be one of N_C values ω_1 through ω_{N_C}. $P(\omega_i)$ is the *a priori* probability that ω_i is the state of nature.

Our observation of nature (the feature vector) is denoted **X**; it is a random vector taking values denoted **x**. The feature vector is d dimensional. Given that the state of nature is ω_i, the feature vector is drawn from the probability density $f(\mathbf{x}|\omega_i)$.

The experiment (in a probability theory sense) underlying the pattern recognition problem is the following:

1. Nature selects a state ω_i according to the probability distribution $P(\omega_i)$.
2. Nature then selects a feature vector **x** according to the probability density function $f(\mathbf{x}|\omega_i)$.
3. Nature reveals the feature vector **x** to the pattern recognizer.
4. Given **x**, the pattern recognizer chooses a particular action α_j. Typically, this action is to decide that the state of nature is $\hat{\omega}$.

Note that this experiment is somewhat simpler than what usually happens in "real life", where a state of nature generates some sensor output, which must be processed to obtain the feature vector.

The pattern recognizer is thus a mapping (also called a *decision rule*) from the feature vector space to the space of possible actions; we denote this mapping as $\alpha(\mathbf{x})$. The fundamental problem of statistical pattern recognition is that of determining this decision rule. We will consider several different approaches: Bayesian, maximum likelihood, and sequential.

2 Bayesian Decision Rules

A Bayesian Decision Rule requires that the decision agent have the following information available:

1. Prior probabilities of the states of nature: $P(\omega_i)$.
2. Conditional probability densities: $f(\mathbf{x}|\omega_i)$.
3. A loss function: $\lambda(\alpha_j|\omega_i)$.

The *loss* is a function $\lambda(\alpha_j|\omega_i)$ which describes the cost associated with choosing action α_j given that ω_i is the true state of nature. We assume in the following that λ is nonnegative.

Example 1 *Suppose that there are two states of nature, ω_1 and ω_2. Suppose that there two actions are possible: α_1, which is to decide that ω_1 is the true state of nature, and α_2, which is to decide that ω_2 is the true state of nature. One loss function that is used quite often in practice is the following:*

$$\lambda(\alpha_1|\omega_1) = \lambda(\alpha_2|\omega_2) = 0$$
$$\lambda(\alpha_2|\omega_1) = \lambda(\alpha_1|\omega_2) = 1$$

Correct decisions are not penalized, while incorrect decisions are penalized equally.

Exercise 1 *Consider Example 1; suppose that we add an additional action α_3, which is to choose to suspend judgment between ω_1 and ω_2. When might this be a reasonable thing to do? How might you modify the loss function in Example 1 to include this possible action?*

Risk is expected loss. We define the *conditional risk* as

$$\begin{aligned} R(\alpha_j|\mathbf{x}) &= E_\Omega\left[\lambda(\alpha_j|\Omega)|\mathbf{x}\right] \\ &= \sum_{i=1}^{N_C} \lambda(\alpha_j|\omega_i)P(\omega_i|\mathbf{x}) \end{aligned}$$

For a given decision rule $\alpha(\mathbf{x})$, we define the *Bayes risk* as

$$\begin{aligned} R &= E_{\mathbf{X}}\left[R(\alpha(\mathbf{X})|\mathbf{X})\right] \\ &= \int_{\mathbf{X}} R(\alpha(\mathbf{x})|\mathbf{x})f(\mathbf{x})d\mathbf{x} \end{aligned}$$

To minimize the Bayes risk, for every value of $\mathbf{x}$, we choose the action that minimizes the conditional risk:

$$\alpha(\mathbf{x}) = \arg\min_j R(\alpha_j|\mathbf{x})$$

Thus, in principle at least, we can find our decision rule by computing the conditional risk for every possible value of $\mathbf{x}$ and, for each of these values, finding the action that minimizes the conditional risk.

Exercise 2 *Given the loss function in Example 1, what do the conditional risks and Bayes risks represent?*

3 Two-Category Classification

Suppose that there are only two states of nature ω_1 and ω_2. Also, suppose that there are only two possible actions: α_1 is to decide that the state of nature is ω_1 and α_2 is to decide that the state of nature is ω_2. Also, let $\lambda_{ji} = \lambda(\alpha_j|\omega_i)$. Under these conditions, the minimum risk decision rule is:

Choose α_1 if

$$R(\alpha_1|\mathbf{x}) \leq R(\alpha_2|\mathbf{x}) \tag{1}$$

Writing the conditional risks in terms of the loss function, we can simplify (1):

$$\lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x}) \leq \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$$

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

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

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

Assuming that $\lambda_{21} \geq \lambda_{11}$ and that $\lambda_{12} \geq \lambda_{22}$ (which is reasonable, since the loss of mistakes should be higher than the loss of correct decisions), we have

$$\frac{f(\mathbf{x}|\omega_1)}{f(\mathbf{x}|\omega_2)} \geq \frac{(\lambda_{12} - \lambda_{22})\,P(\omega_2)}{(\lambda_{21} - \lambda_{11})\,P(\omega_1)} \tag{2}$$

This is a likelihood ratio test, in which the ratio of the two likelihoods is compared to a threshold.

Exercise 3 *What is the likelihood ratio test for the case when $\lambda_{11} = \lambda_{22} = 0$ and $\lambda_{12} = \lambda_{21} = C$, where C is a positive constant? What important criterion of optimality does this likelihood ratio meet? This likelihood ratio is often called the* maximum a posteriori *(MAP) decision rule, since it is equivalent to the following decision rule: Choose ω_1 if*

$$P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$$

4 Probability of Error

In most pattern recognition applications, the possible actions are to decide on one of the N_C states of nature. An error is made when a state is chosen that is not the actual state of nature. In the following, we let α_i be the action to decide that the state of nature is ω_i.

To compute the probability of error P_E, we first define regions $\mathcal{R}_1$ through $\mathcal{R}_{N_C}$ as follows:

$$\mathcal{R}_i = \{\mathbf{x} : \alpha(\mathbf{x}) = \alpha_i\}$$

It is easier to compute the probability that the decision is correct P_C; then $P_E = 1 - P_C$. Given that ω_i is the true state of nature, the decision will be correct if $\mathbf{x} \in \mathcal{R}_i$; thus, $P_{C|\omega_i}$, the probability of correct given that the true state of nature is ω_i, is

$$P_{C|\omega_i} = \int_{\mathcal{R}_i} f(\mathbf{x}|\omega_i) d\mathbf{x}$$

The probability of correct is the conditional probability averaged over the probabilities of the states of nature:

$$\begin{aligned} P_C &= \sum_{i=1}^{N_C} P_{C|\omega_i} \\ &= \sum_{i=1}^{N_C} \int_{\mathcal{R}_i} f(\mathbf{x}|\omega_i) d\mathbf{x} \end{aligned}$$