Mid-term review
ECE 175
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos Spring 2008

1. In this problem we will consider the traditional probability scenario of coin tossing. However, we will consider two variations. First, the coin is not fair. Denoting by S the outcome of the coin toss we have

$$P_S(heads) = \alpha,\ \alpha \in [0,1].$$

Second, you do not observe the coin directly but have to rely on a friend that reports the outcome of the toss. Unfortunately your friend is unreliable, he will sometimes report heads when the outcome was tails and vice-versa. Denoting the report by R we have

$$P_{R|S}(tails|heads) = \theta_1 \quad (1)$$
$$P_{R|S}(heads|tails) = \theta_2 \quad (2)$$

where $\theta_1, \theta_2 \in [0,1]$. Your job is to, given the report from your friend, guess the outcome of the toss.

a) Given that your friend reports heads, what is the optimal decision function in the minimum probability of error sense. That is, when should you guess heads, and when should you guess tails?

b) Consider the case $\theta_1 = \theta_2$. Can you give an intuitive interpretation to the rule derived in **a)**?

c) You figured out that if you ask your friend to report the outcome of the toss various times, he will produce reports that are statistically independent. You then decide to ask him to report the outcome n times, in the hope that this will reduce the uncertainty. (Note: there is still only one coin toss, but the outcome gets reported n times). What is the new minimum probability of error decision rule?

d) Consider the case $\theta_1 = \theta_2$ and assume that the report sequence is *all heads*. Can you give an intuitive interpretation to the rule derived in **c)**?

2. Consider a two dimensional classification problem with two Gaussian classes

$$P_{\mathbf{X}|Y}(\mathbf{x}|i) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_i)}, \quad i \in \{0, 1\}$$

of identical covariance $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$. For all problems assume the "0-1" loss function.

a) If the classes have means

$$\mu_0 = -\mu_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

and equal prior probabilities, $P_Y(0) = P_Y(1)$, what is the Bayes decision rule for this problem?

b) What are the marginal distributions for the features x_1 and x_2 for each class? In particular

1. derive expressions for the class-conditional densities $P_{X_1|Y}(x_1|i)$ and $P_{X_2|Y}(x_2|i)$ for $i \in \{0, 1\}$, where $\mathbf{x} = (x_1, x_2)^T$.
2. plot a sketch of the two densities associated with class $Y = 0$ and a sketch of the two densities associated with class $Y = 1$.
3. determine which feature is most discriminant.

c) A linear transformation of the form

$$\mathbf{z} = \mathbf{\Gamma x}$$

was applied to the data, where $\mathbf{\Gamma}$ is a 2×2 matrix. The decision boundary associated with the BDR is now the hyperplane of normal $\mathbf{w} = (1/\sqrt{2}, -1/\sqrt{2})^T$ which passes through the origin.

1. determine the matrix $\mathbf{\Gamma}$
2. What would happen if the the prior probability of class 0 was increased after the transformation? Here it suffices to give a qualitative answer, i.e. simply say what would happen to the hyperplane.
3. what is the distance in the original space ($\mathbf{x}$) which is equivalent to the Euclidean distance in the transformed space ($\mathbf{z}$)?

3. Consider a classification problem with two Gaussian classes

$$P_{\mathbf{X}|Y}(\mathbf{x}|i) = \mathcal{G}(\mathbf{x}, \mu_i, \Sigma), \quad i \in \{0, 1\}$$

of equal probability

$$P_Y(i) = 1/2.$$

In class, we have considered the BDR solution to this problem. This consists of estimating the parameters of the Gaussian classes and then plugging on the BDR to obtain the decision boundary. Here we will consider an alternative solution, that works directly on the class posteriors.

a) Show that the posterior probability for class 1 is of the form (the posterior for class 0 is $1 - P_{Y|\mathbf{x}}(1|\mathbf{x})$)

$$P_{Y|\mathbf{X}}(1|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T\mathbf{t}}} \tag{3}$$

where $\mathbf{t}^T = [\mathbf{x}^T 1]$. What is the vector $\mathbf{w}$?

b) Show that an iid sample $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ has posterior probability

$$P_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}_y|\mathcal{D}_x) = \prod_{i=1}^{n} P_{Y|\mathbf{X}}(y_i|\mathbf{x}_i) \tag{4}$$

with

$$P_{Y|\mathbf{X}}(y_i|\mathbf{x}_i) = \left(\frac{1}{1 + e^{-\mathbf{w}^T\mathbf{t}_i}}\right)^{y_i} \left(\frac{e^{-\mathbf{w}^T\mathbf{t}_i}}{1 + e^{-\mathbf{w}^T\mathbf{t}_i}}\right)^{1-y_i}, \tag{5}$$

where $\mathcal{D}_y = \{y_1, \ldots, y_n\}$ and $\mathcal{D}_x = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

Note: We can now learn the classification boundary, by learning the posterior probabilities with standard maximum likelihood estimation. For example we can solve for $\mathbf{w}^\star$ such that

$$\mathbf{w}^\star = \arg\max_{\mathbf{w}} P_{\mathbf{Y}|\mathbf{X}}(\mathcal{D}_y|\mathcal{D}_x).$$