Machine Learning 521289S
Spring 2019
Exercise 2: Multidimensional Classification,
Discriminant Functions & Performance Criteria
Exercises 3 and 5 give exam points!

General Instructions

Use the upper case $P(\cdot)$ to denote the *probability mass function* (of a discrete random variable), and lower case $p(\cdot)$ to denote a *probability distribution* (of a continuous random variable). Vectors are written in bold lowercase letters ($\mathbf{x}$) and matrices in bold uppercase letters (**X**). In handwriting, use bars over the letters to denote vectors ($\vec{x}$). Variables are written in upper case letters and their values in lower case letters ($X = x$, $C = 1$).

A d-dimensional multivariate normal density has the PDF

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

where the expected value μ is a vector with d components, i.e. $\mu \in \mathbb{R}^d$, and $\mathbf{\Sigma}$ is a d-by-d covariance matrix, i.e. $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$.

Exercises

1. *Discriminant functions*

   (a) Figure 1 shows a dataset gathered in a two class scenario. Class 1 is represented by the red triangles and class 2 by the blue boxes. The figure is two dimensional meaning that (at least) two features have been measured. Let us denote the features by X_1 and X_2. The points in the dataset are pairs $\mathbf{x} = (x_1, x_2)$, realisations of the two-dimensional random variable $\mathbf{X} = (X_1, X_2)$. Using the figure, design a simple discriminant function $g(\mathbf{x})$ that is able to correctly classify all the shown samples

   (b) Now the situation is similar to the part (a) but there are three classes meaning that a single discriminant function cannot be used. Again using Figure 2, find three simple discriminant functions $g_1(\mathbf{x})$, $g_2(\mathbf{x})$ and $g_3(\mathbf{x})$ that can be used to correctly classify this learning data

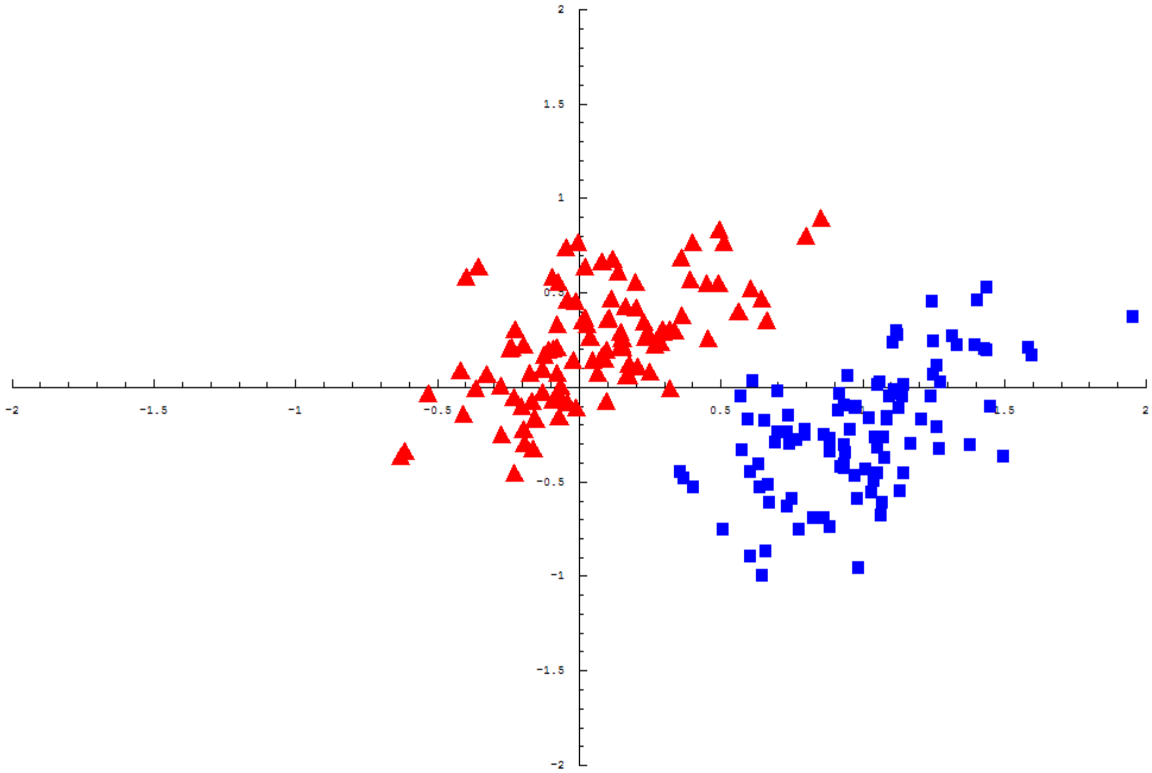

Figure 1: A learning dataset in a two-class scenario with two features.

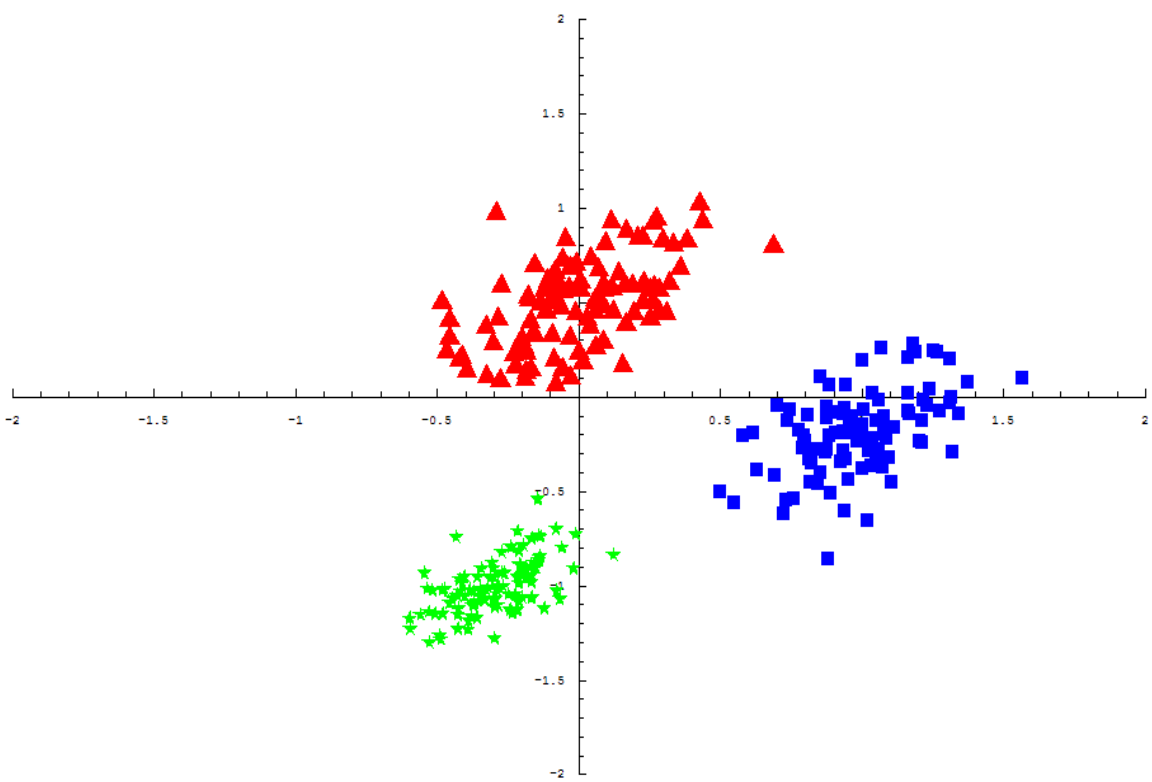

Figure 2: A learning dataset in a three-class scenario with two features.

2. *Discriminant Functions for Multivariate Normal Distributions*

   Show that for k multivariate normal distributions with arbitrary means $\mu_i \in \mathbb{R}^n$ and **same** covariance matrices $\Sigma_i = \Sigma$, $i = 1, \ldots, k$, the discriminant functions can have the linear form

   $$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0},$$

   where $\mathbf{w}_i = \mathbf{\Sigma}^{-1}\mu_i$ and $w_{i0} = -\frac{1}{2}\mu_i^T \mathbf{\Sigma}^{-1}\mu_i + \ln P(\omega_i)$. In other words, complete the steps to derive the form given in the course book on page 39 Section 2.6.2. (Case $\mathbf{\Sigma_i} = \mathbf{\Sigma}$) or on page 33 in Chapter 2 of the old lecture notes (in Finnish).

   Hint: You will probably need to use the symmetry of the covariance matrix ($\mathbf{\Sigma} = \mathbf{\Sigma}^T$) and the following identities familiar from matrix algebra: $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$, $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ that hold for certain matrices $\mathbf{A}$ and $\mathbf{B}$. (Check that the conditions required by these identities are satisfied when using them!)

3. *Classifying with Multivariate Normal Distributions*

   Let us consider a two-class scenario with a two-dimensional feature vector. The basic setting is somewhat similar to the question 1 a). Suppose that we have a two-dimensional feature $\mathbf{X} \in \mathbb{R}^2$, and $f_1(\mathbf{x}) = p(\mathbf{x}|C = 1) \sim N(\mu_1, \mathbf{\Sigma}_1)$ and $f_2(\mathbf{x}) = p(\mathbf{x}|C = 2) \sim N(\mu_2, \mathbf{\Sigma}_2)$ are the class conditional probability distributions , where

   $$\mu_1 = \mu_2 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T,$$

   $$\mathbf{\Sigma}_1 = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

   The situation differs from the assignment 1 a) in that in this case we know the exact distributions which happen to overlap (same means) and do not have equal covariance matrices, i.e. the point clouds would be oriented differently.

   (a) Sketch the shapes of the class conditional probability distributions using equipotential curves. In other words, you have to solve the equations $f_1(x) = c$ and $f_2(x) = c$ for a range of suitable values of c.

   In the following, we assume that the classes have equal prior probabilities.

   (b) Use discriminant functions to classify the sample with the feature vector $\mathbf{x} = \begin{bmatrix} 1 & 4 \end{bmatrix}^T$. See page 41 Section 2.6.3. in the course book.

   (c) Solve the decision boundary between the classes when the classification is made according to the Bayes decision rule.

4. *Bayes Formula Redux*

You have designed a new, portable cancer screening device for a major health care company. A study has shown that the prevalence of the cancer is 0.8 per cent in the overall population. The device is pencil shaped and lits up a green led when there is no cancer detected and a red led when cancer is found. However, the device is not perfectly reliable. In fact, it detects cancer only in 98 per cent of people with cancer. What is more, the device has been found to show the red light to 3 per cent of test subjects that did not have cancer.

(a) What is the probability that a patient has cancer when the test is positive?

(b) After a positive result, the test is run once more. If the result of the second test were also positive, what would be the probability of cancer?

5. *Sensitivity, Specificity and Predictive Values*

When we have a two class problem, it is many times useful to examine Sensitivity and Specificity which resemble the classifiers ability to separate the classes from each other. Based on a four-fold table known as *the confusion matrix* (see Table 1), we can determine following:

(1) **Sensitivity**, true positive rate (TPR), the proportion of positive samples that are correctly identified as such

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

(2) **Specificity**, true negative rate (TNR), the proportion of negative samples that are correctly identified as such

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True negatives} + \text{False Positives}}$$

(3) Predictive value of a positive test, **Precision**

$$\text{Positive Predictive Value} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

(4) Predictive value of a negative test

$$\text{Negative Predictive Value} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}$$

Table 1: Confusion matrix.

	Predicted Condition Positive	**Predicted Condition Negative**	**Total**
Condition Positive	True Positives (TP)	False Negatives (FN)	TP + FN
Condition Negative	False Positives (FP)	True Negatives (TN)	FP + TN
Total	TP + FP	FN + TN	Total Amount of Samples

(a) To evaluate the performance of your new diagnostic test, you have checked it out on 100 known cases of the disease for which the test was designed, and on 200 controls known to be free of the disease. 95 of the known cases yield positive tests, as do 35 of the controls. Based on these data, what is the sensitivity and specificity of the test?

(b) Another diagnostic test is 90% sensitive and 85% specific. The test group is comprised of 700 people known to have the disease and 600 people known to be free of the disease. How many of the known positives would actually test positive? How many of the known negatives would actually test negative?

(c) A test with 98, 9% sensitivity and 99, 1% specificity is used to screen a population of 1 000 000 people for a disease with 1% prevalence. What would be the positive predictive value of this test?

(d) A test with 98% sensitivity and 99% specificity is used to screen a population of people for a disease with a 5% prevalence rate. What is the positive predictive value of this test?

Answers:
1. Infinite number of correct solutions exist.
2. Hint: Start with the logarithmic form of the discriminant function on page 30 in the course book Section 2.4.1., $g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$.
3. Equipotential curves are of the form: $\frac{x^2}{Ca_i} + \frac{y^2}{Cb_i} = 1$, where the constant C depends on the equipotential value c: $-2\left(\ln c + \ln 2\pi + \ln \sqrt{|\mathbf{\Sigma}_i|}\right)$
4. Let $C = 1$ for cancer and 0 for no cancer, and $D = 1$ when it is detected and 0 when not. a) We get the probability of $0.208\,51 \approx 21\%$, b) Let us denote the result of the first test with D_1 and the result of the second test with D_2. Further, we assume conditional independence: $P(D_1 = d_1, D_2 = d_2|C = c) = P(D_1 = d_1|C = c)\,P(D_2 = d_2|C = c)$ for any $d_1, d_2, c \in \{1, 2\}$. We get the probability of $0.895\,90 \approx 90\%$
5. a) Sensitivity 95%, Specificity 82.5%, b) $TP = 630$, $FN = 70$, $TN = 510$, and $FP = 90$, c) $PPV \approx 52.6\%$, d) $PPV \approx 84\%$.