Pattern Recognition Winter Term 2019

Institute of Neural Information Processing

PD Dr F. Schwenker

Assignment 3 (Submission until November 19, 2019)

Exercise 1 (3 points): Decision Surfaces via Gaussian Density Functions [Pen and Paper]

In the Bayes classification process, we try to find decision surfaces which separate the classes in the feature domain. These surfaces define a region for each class so that new data points can easily be assigned to one of the classes. In this exercise, we want to analyse different decision surfaces based on multivariate Gaussian densities.

We operate in the two-dimensional feature space and use a two-class problem, i.e. a point $x \in \mathbb{R}^2$ is either assigned to the class ω_1 or the class ω_2. With the Bayes rule, the resulting surface is defined by the corresponding posterior probabilities of the two classes

$$
\begin{aligned}
P(\omega_1|x) &= P(\omega_2|x) \\
\frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} &= \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \\
p(x|\omega_1) \cdot P(\omega_1) &= p(x|\omega_2) \cdot P(\omega_2)
\end{aligned}
$$

and reduces to the likelihoods $p(x|\omega_i)$ and prior-probabilities $P(\omega_i)$. For the sake of simplicity, we assume equiprobable classes ($P(\omega_1) = P(\omega_2) = 0.5$) so that the decision is solely based on the Gaussian density functions

$$
p(x|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right)
$$

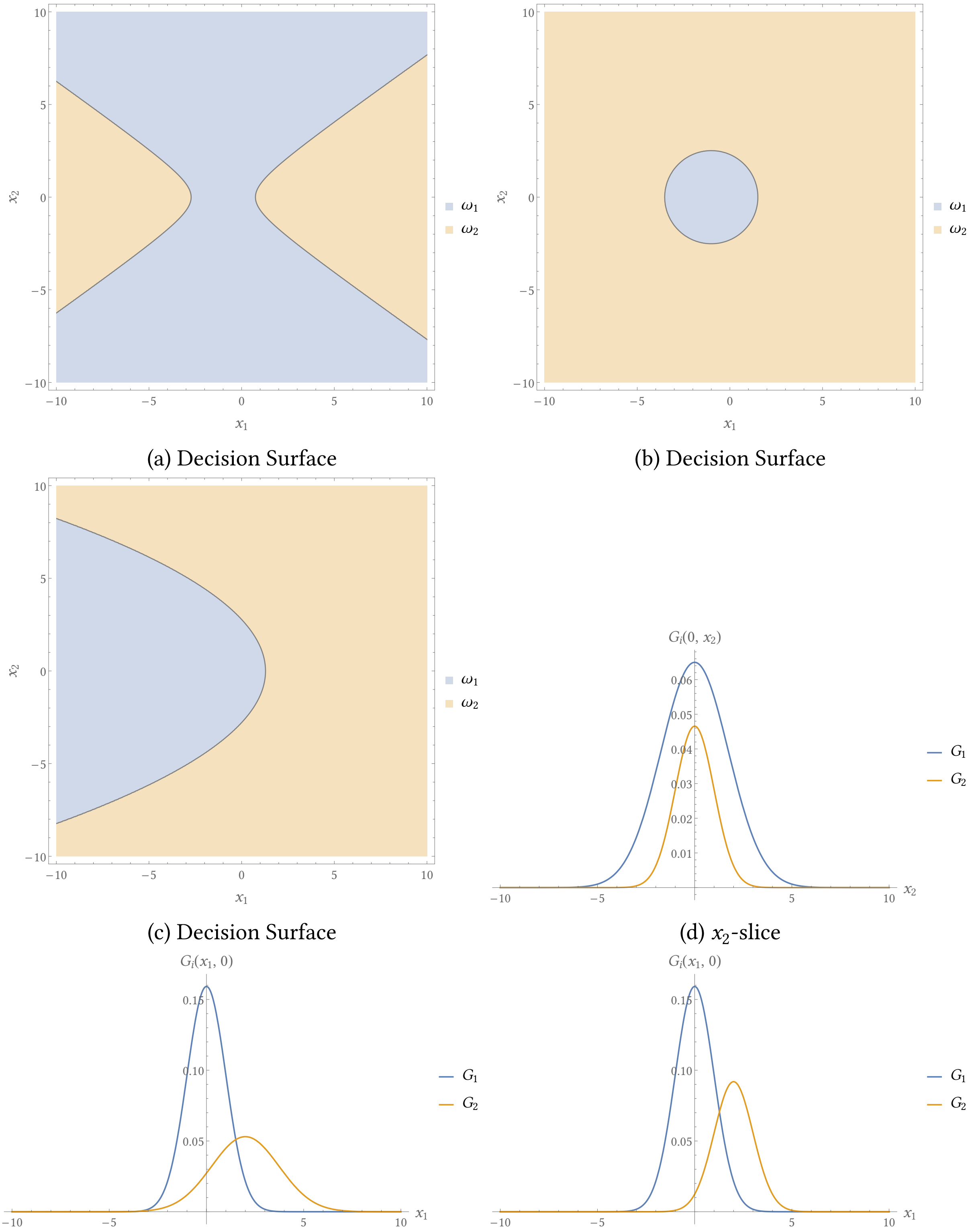
with the mean vector μ_i and the covariance matrix Σ_i for each class i.

Figure 1 shows three different decision surfaces as well as either a corresponding x_1- or x_2-slice. Table 1 summarizes the parameters used to create the figures. Note that we have two Gaussian functions in each case (one for each class).

Your task is to map each line in Table 1 to the corresponding decision surface and slice. Please explain your choices, e.g. via sketches.

Exercise 2 (7 points): Euclidean Distance, Standardization, Mahalanobis Distance

For many applications, it is helpful when we deal with homogeneous data. This is especially important when points should be compared in one way or another. Here, we want to measure the similarity between two points by calculating the corresponding distance. We start with the Euclidean distance, move forward to standardized variables and end up with the Mahalanobis distance.



(a) Decision Surface

(b) Decision Surface

(c) Decision Surface

(d) x_2-slice

(e) x_1-slice

(f) x_1-slice

Figure 1: Decision surfaces and slices of different Gaussian functions.

Table 1: Parameters for the Gaussian functions which define the decision surfaces. Each line corresponds to two plots of Figure 1.

μ_1	Σ_1	μ_2	Σ_2	Decision Surface (**a**, **b** or **c**)	Slice (**d**, **e** or **f**)
$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}3 & 0\\0 & 3\end{pmatrix}$		
$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}1 & 0\\0 & 3\end{pmatrix}$		
$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}2 & 0\\0 & 3\end{pmatrix}$	$\begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}6 & 0\\0 & 1\end{pmatrix}$		

We are using the example dataset in Table 2 throughout this exercise. It comprises a toy example which consists of $n = 7$ data points $x_i \in \mathbb{R}^2 = (x_i, y_i)$ relating education to monthly income. The dataset is also visualized in Figure 2.

Table 2: Seven observations which relate education to income.

Years of education x	Monthly income y
1	1200
1.5	1000
1.9	1300
3	1900
5	3200
5.5	2400
7	4200

1. [Pen and Paper] In matrix notation, the *Euclidean distance* is defined as

$$\mathrm{d}_E(x_i, x_j) = \sqrt{(x_i - x_j)^T(x_i - x_j)} \tag{1}$$

with the data points x_i being represented by column vectors.

a) When you look at Figure 2, what would you expect is the nearest neighbour of the point $(1, 1200)$?

b) Figure 3 shows graphs where for each data point the nearest neighbour is depicted using different distance types. The graph in Figure 3 (a) corresponds to the Euclidean distance. As you can see, the nearest point to $(1, 1200)$ is $(1.9, 1300)$. Does this result match with your previous expectation? If not, what could be the problem here and why does Figure 2 trick you?

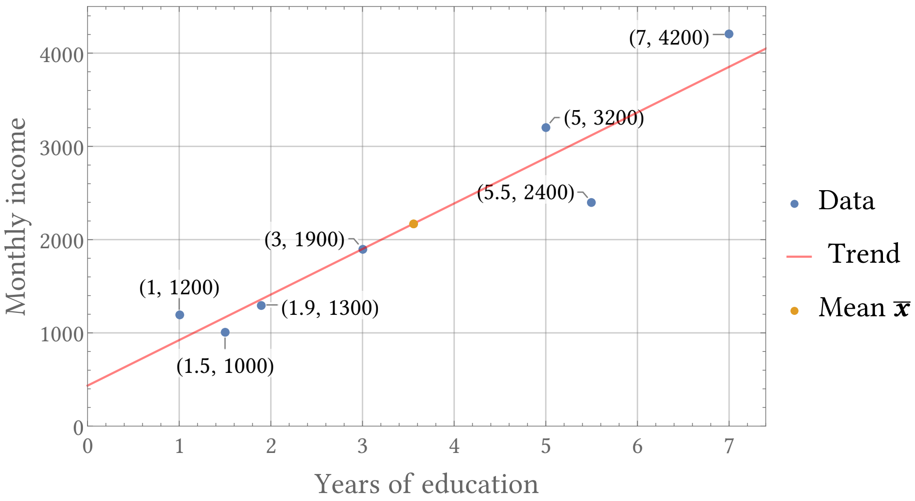


Figure 2: Example dataset which relates education to income (Table 2).

2. [Pen and Paper] One way to overcome the previous issue is by *standardizing the variables*

$$\tilde{x}_i = \frac{x_i - \mu_x}{\sigma_x} \tag{2}$$

with respect to the mean and the variance

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)^2.$$

The same procedure is applied to the y-dimension.

a) What are the units of the dimensions $\tilde{x}$ and $\tilde{y}$ and which term of Equation 2 is responsible for this change? As a reminder, the unit of x is years of education and y denotes monthly income.

b) After standardization, we can use the Euclidean distance again to compare standardized points $\tilde{\boldsymbol{x}}_i = (\tilde{x}_i, \tilde{y}_i)$. The result is visualized in Figure 3 (b). We see that using standardized variables mitigates the previous issue as, for example, the nearest neighbour of the point (1, 1200) is now (1.5, 1000). Explain why this is the case, i.e. why does it help to standardize the variables?

c) Besides the two-step procedure (first standardizing, then calculating the Euclidean distance), it is also possible to use a direct formula which operates on the original points $\boldsymbol{x}_i$ and standardizes the variables on the fly. For this, show that the equation

$$\mathrm{d}_E(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T S^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)} = \mathrm{d}_S(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3}$$

holds. The matrix

$$S = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

contains the variances on the diagonal.

3. [Pen and Paper] Even though we compensated for the effect of different variances in each dimension, there is still room for improvement. The trend line in Figure 2 clearly indicates a linear dependency between the variables. This means that points which are located more on the right (longer years of education) have naturally a higher income.

   If we also want to account for this effect, standardization is not enough anymore since it only compensates for the variance *inside* each dimension but the effect we observe here arises from the variance *across* two dimensions. Namely, the dimension of years of education correlates positively with the monthly income. To compensate for this correlation, we can use the *Mahalanobis distance*

   $$\mathrm{d}_M(x_i, x_j) = \sqrt{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)} \tag{4}$$

   where Σ denotes the covariance matrix

   $$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} \quad \text{and} \quad \sigma_{xy} = \sigma_{yx} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)$$

   is the covariance between the two feature dimensions.

   a) We first want to know how the distance pattern looks like when using the Mahalanobis distance. For this, sketch some isocurves in Figure 2 around the mean $\bar{x}$. All points p on such a curve have the same distance $d \in \mathbb{R}$ to the mean $\left\| \bar{x} - p \right\| = d$ (cf. script page 52).

   b) The nearest neighbour graph for this distance type is shown in Figure 3 (c). Explain with the help of isocurves why the nearest neighbour of $(5, 3200)$ is $(7, 4200)$ and not $(5.5, 2400)$ as it was the case in Figure 3 (b). Hint: you are free to move the isocurves in the coordinate system, i.e. they are translation-invariant.

4. [Python] Finally, we want to implement the different distance types and compute some of the distances between points of Table 2.

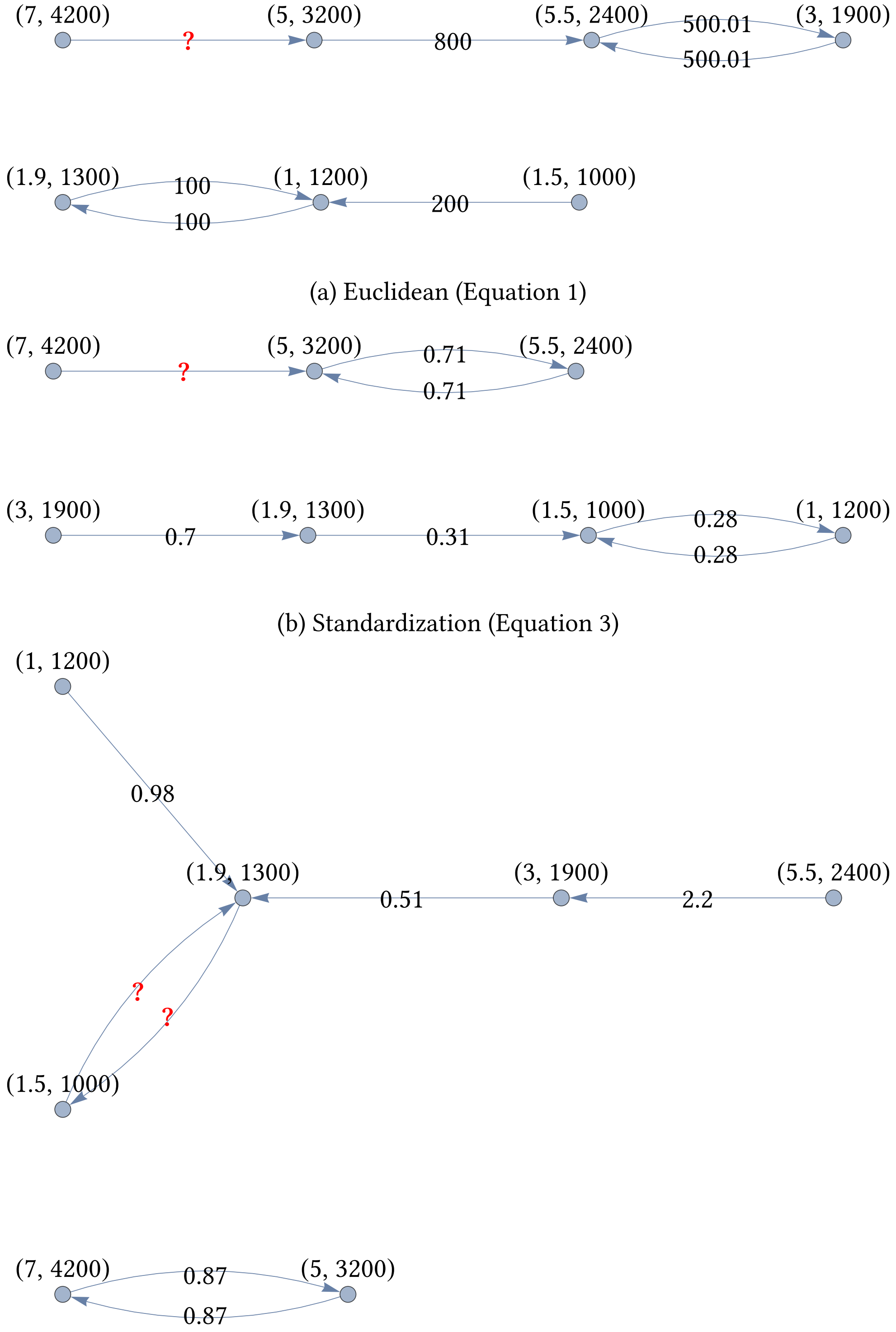
   a) Write a function `distance(xi, xj, matrixA)` which implements

   $$\mathrm{d}(x_i, x_j) = \sqrt{(x_i - x_j)^T A^{-1} (x_i - x_j)},$$

   i.e. a general version of the above distance equations with the distance type (Euclidean, standardization, Mahalanobis) being defined by the matrix A.

   b) Define the matrix for each distance type for the data listed in Table 2.

   c) Calculate the missing distances of Figure 3. Round your results to two decimal places.



(a) Euclidean (Equation 1)

(b) Standardization (Equation 3)

(c) Mahalanobis (Equation 4)

Figure 3: Nearest neighbour graphs. Each vertex in the graph corresponds to a data point from Table 2 and has exactly one outgoing edge pointing to its nearest neighbour.