7^{th} Homework

Chalkiopoulos Georgios | *p*3352124

March 4, 2022

Exercise 1.

(I) Assume that the classes are equiprobable.
(i) Depict graphically $P(\omega_i)p(x|\omega_i)$ and identify the respective regions.

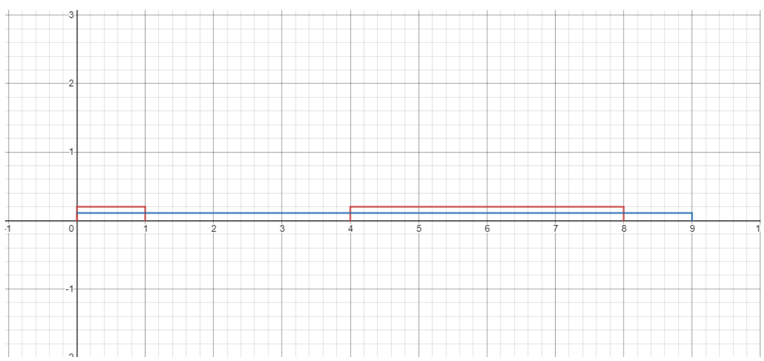

Figure 1: $P(\omega_i)p(x|\omega_i)$ for ω_1 (red) and ω_2 (blue).

The respective decision regions are:

$$R_{\omega_1} := x \in (0,1) \cup (4,8) \quad R_{\omega_2} := x \in (1,4) \cup (8,9)$$

(ii) Classify the point $x' = 3.5$, using the bayes classifier.

It is easy to prove that this point belongs to ω_2 since the corresponding probability for the ω_1 area is zero.

(II) Assume that the classes are not equiprobable. (i) Determine a set of values for the a priori probabilities of the two classes that guarantee that $x' = 5$ is assigned to class ω_2. Justify briefly your choice.

Normally (in the equiprobable case) at the point $x' = 5$ it is $P(\omega_1)p(x'|\omega_1) > P(\omega_2)p(x'|\omega_2)$. In order to classify the point at this point as ω_2 we have to solve the inequality:

$$
\begin{aligned}
P(\omega_1)p(x'|\omega_1) < P(\omega_2)p(x'|\omega_2) \Rightarrow \\
P(\omega_1)\frac{1}{5} < P(\omega_2)\frac{1}{9}
\end{aligned}
\tag{1.1}
$$

Any combination of $P(\omega_1), P(\omega_2)$ that satisfies (1.1) and $P(\omega_1) + P(\omega_2) = 1$ will classify the point $x' = 5$ to the ω_2 class. We can chose $P(\omega_1) = 0.1$ and $P(\omega_2) = 0.9$ which results in:

$$0.1\frac{1}{5} < 0.9\frac{1}{9} \Rightarrow 0.02 < 0.1 \rightarrow True$$

(ii) Is there any combination of the class priori probabilities that guarantees that $x' = 3$ will be assigned to ω_1? Explain.

$x' = 3$ can in no way be assigned to ω_1 since $P(\omega_1)p(x'|\omega_1) = 0$ thus the inequality $P(\omega_1)p(x' = 3.5|\omega_1) < P(\omega_2)p(x' = 3.5|\omega_2)$ cannot be satisfied for any $P(\omega_1), P(\omega_2)$.

Exercise 2.

We will start by plotting the two distributions:

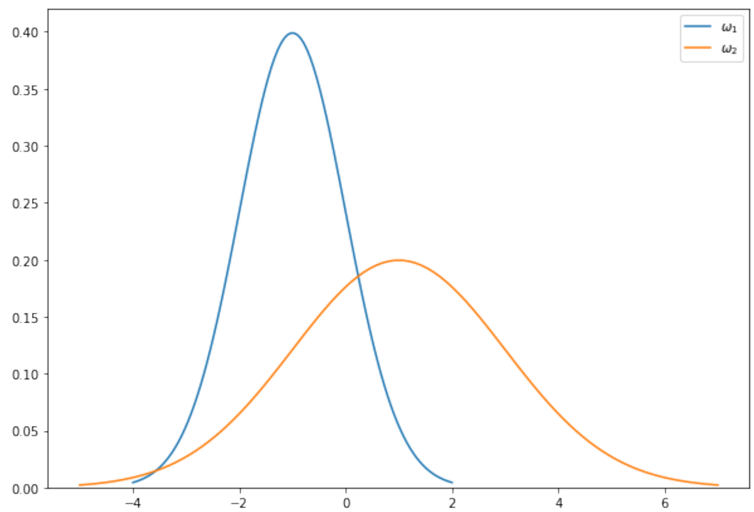

Figure 2: $\omega_1 \sim \mathcal{N}(-1, 1), \omega_2 \sim \mathcal{N}(1, 4)$.

In order to determine the decision regions, we will solve the equality: $P(\omega_1)p(x|\omega_1) = P(\omega_2)p(x|\omega_2)$, with $P(\omega_1) = P(\omega_2) = 0.5$. It is:

$$0.5\frac{1}{1\sqrt{2\pi}}e^{-\left(\frac{x+1}{2}\right)^2} = 0.5\frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-1}{2}\right)^2} \Rightarrow$$

$$x = \begin{cases} -3.57 \\ 0.24 \end{cases}$$

The above points result in the following decision regions, shown in Figure 3:

$$R_1 = (-\infty, -3.57) \cup (0.24, +\infty)$$
$$R_2 = (-3.57, 0.24)$$

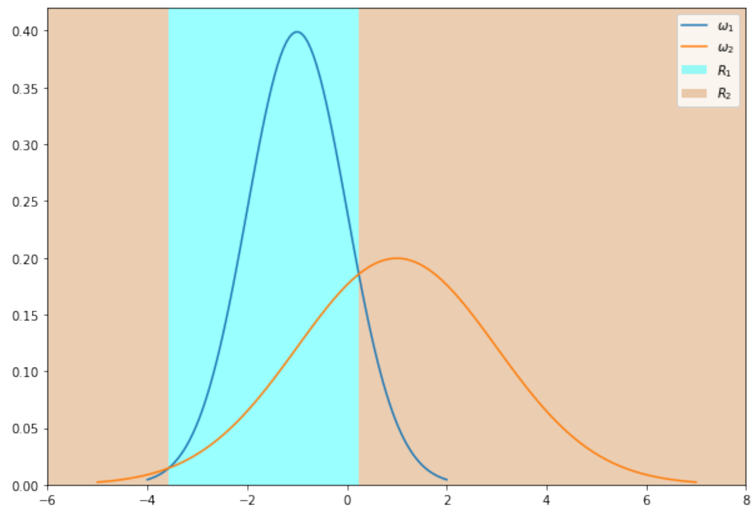

Figure 3: $\omega_1 \sim \mathcal{N}(-1,1), \omega_2 \sim \mathcal{N}(1,4)$ with decision regions R_1, R_2.

Exercise 3.

It is:

$$\begin{aligned}
&p(\boldsymbol{x}|\omega_1) = N(\boldsymbol{\mu_1}, \Sigma_1), p(\boldsymbol{x}|\omega_2) = N(\boldsymbol{\mu_2}, \Sigma_2)\\
&\boldsymbol{\mu_1} = [6,0]^T, \boldsymbol{\mu_1} = [0,6]^T\\
&\Sigma_1 = \Sigma_2 = 2 \cdot I
\end{aligned}$$

(a) Utilizing the Bayes decision rule, classify each one of the data points $\boldsymbol{x}_1 = [2,4]^T$, $\boldsymbol{x_2} = [4,2]^T$ and $\boldsymbol{x}_3 = [2,2]^T$ one out of the three classes. For each point we will calculate the a posteriori probabilities, considering that $P(\omega_1) = P(\omega_2) = 0.5$.

For the first point, $x_1 = [2,4]^T$ it is:

$$p(\boldsymbol{x}_1|\omega_j) = \frac{1}{(2\pi)^{j/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{(\boldsymbol{x}_1 - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_j)}{2}\right)$$

Similar to the previous assignments we calculate the estimations of $p(\boldsymbol{x}_i|\omega_j)$ of the various x_is:

Looking at Table 1 we observe that x_1 is assigned the the ω_2 class, x_2 is assigned the the ω_1 class while x_3 is probably on the decision surface since the estimation is the same for both classes.

Table 1: Estimations of $p(\boldsymbol{x}_i|\omega_j)$.

	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$
$p(\boldsymbol{x}_i\|\omega_1)$	3.78e-05	1.52e-02	7.58e-04
$p(\boldsymbol{x}_i\|\omega_2)$	1.52e-02	3.78e-05	7.58e-04

We can go one step further and calculate $P(\omega_j|x_i) = \dfrac{p(\boldsymbol{x}_1|\omega_j)\cdot p(\omega_j)}{\sum_{k=1}^{m} p(\boldsymbol{x}_1|\omega_k)\cdot p(\omega_k)}$.

By performing the calculation (similar to assignment 6) we get the following table for the various points: As expected, the same conclusions are made for

Table 2: a posteriori probabilities $p(\omega_j|\boldsymbol{x}_i)$.

	x_1	x_2	x_3
$p(\omega_j\|\boldsymbol{x}_i)$	0.0025	0.9975	0.5000
$p(\omega_j\|\boldsymbol{x}_i)$	0.9975	0.0025	0.5000

the 3 points, but it is now easier to understand, since we computed probabilities.

(b) Determine the line that separates the class regions of the two classes.

We will solve the equation:

$$P(\omega_1)p(x|\omega_1) = P(\omega_2)p(x|\omega_2) \qquad \Rightarrow$$

$$0.5\frac{1}{(2\pi)|\Sigma_1|^{1/2}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\right) =$$

$$0.5\frac{1}{(2\pi)|\Sigma_2|^{1/2}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\right) \qquad \Rightarrow$$

$$(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) \qquad \Rightarrow$$

$$\begin{bmatrix} x_1-6 & x_2 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1-6 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2-6 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2-6 \end{bmatrix} \qquad \Rightarrow$$

$$x_1^2 - 12x_1 + x_2^2 + 36 = x_1^2 + x_2^2 - 12x_2 + 36 \qquad \Rightarrow$$
$$x_1 = x_2$$

The results is expected, since the point $\boldsymbol{x}_3 = [2,2]^T$ had the same a posteriori probability and was not assigned to any Class, having $x_1 = x_2$.

Exercise 4.

(a) Show that the Bayesian classifier borders the decision regions by the perpendicular bisector of the line segment whose endpoints are $\mu_1 \ \ \mu_2$

Having worked in the previous exercise with a two-dimensional two-class problem we found that the Bayesian classifier decision line is:

$$
\begin{aligned}
&P(\omega_1)p(x|\omega_1) = P(\omega_2)p(x|\omega_2) && \Rightarrow \\
&0.5\frac{1}{(2\pi)|\Sigma_1|^{1/2}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}{2}\right) = \\
&0.5\frac{1}{(2\pi)|\Sigma_2|^{1/2}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}{2}\right) \\
&(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) && (4.1)
\end{aligned}
$$

Having $\Sigma_1 = \Sigma_2 = \sigma^2 I$ (4.1) can now be written as:

$$
\begin{aligned}
&(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) && \Rightarrow \\
&(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T\sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) && \Rightarrow \\
&(\boldsymbol{x}-\boldsymbol{\mu}_1)^T(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T(\boldsymbol{x}-\boldsymbol{\mu}_2) && \Rightarrow \\
&\|\boldsymbol{x}-\boldsymbol{\mu}_1\|^2 = \|\boldsymbol{x}-\boldsymbol{\mu}_2\|^2 && (4.2)
\end{aligned}
$$

(4.2) according to the Exercise matched the equation that describes the perpendicular bisector of a line segment whose endpoints are $\mu_1 \ \ \mu_2$.

(b) What would be the border in the case where $\Sigma \neq \sigma^2 I$.

In this case, looking at (4.1) we can see that each x_i, μ_i would be multiplied by a quantity of σ thus, in the solution, the various x_is would have different weights. In this case, the resulting border would be a curve in the space, which would be the result of solving (4.1).

Exercise 5.

It is:

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \text{ and } p(x|\omega_2) \begin{cases} \frac{1}{2\sqrt{2\pi}}, & x \in [-\sqrt{2\pi}, \sqrt{2\pi}] \\ 0, \end{cases}$$

(a) Determine the classifier that minimizes the probability of classification error and the write down the decision regions associated with each class.

Before proceeding we will plot the two pdfs.

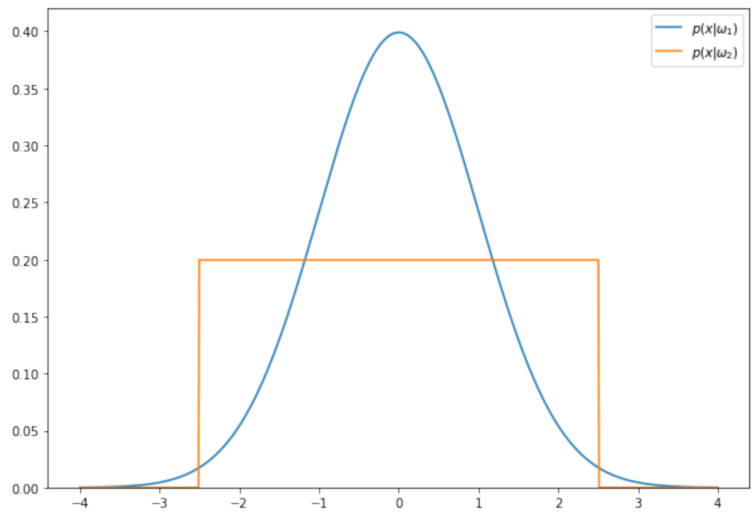

Figure 4: $p(x|\omega_i)$ for ω_1 and ω_2.

As we now know from the previous exercises we have to solve the equation $P(\omega_1)p(x|\omega_1) = P(\omega_2)p(x|\omega_2)$, and with the classes being equiprobable, for $x \in [-\sqrt{2\pi}, \sqrt{2\pi}]$ it is:

$$\begin{aligned} & p(x|\omega_1) = p(x|\omega_2) \Rightarrow \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = \frac{1}{2\sqrt{2\pi}} && \Rightarrow \\ & \exp\left(-\frac{x^2}{2}\right) = \frac{1}{2} \Rightarrow \frac{x^2}{2} = \ln 2 && \Rightarrow \\ & x_{1,2} = \begin{cases} -1.177 \\ 1.177 \end{cases} , x \in [-\sqrt{2\pi}, \sqrt{2\pi}] && \end{aligned} \tag{5.1}$$

Finally we can determine the decision regions for the two classes, ω_1 and ω_2

$$R_1 = (-\infty, -\sqrt{2\pi}) \cup (-1.177, 1.177) \cup (\sqrt{2\pi}, +\infty)$$
$$R_2 = (-\sqrt{2\pi}, -1.177) \cup (1.177, \sqrt{2\pi})$$

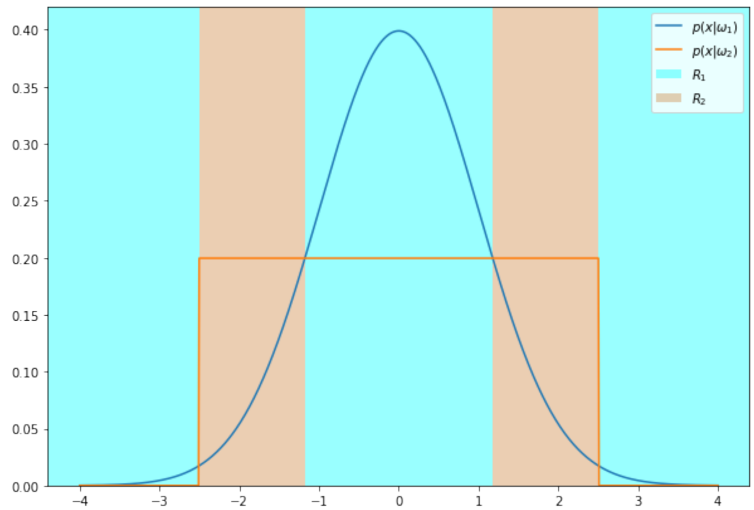

Figure 5: ω_1 and ω_2 with decision regions R_1, R_2.

(b) Assume that the errors on the data that stem from ω_2 are twice more serious than those that stem from ω_1. Write explicitly the average risk for this case by defining explicitly the loss matrix Λ and determine the decision regions for the classes in this case.

We will start by defining Λ:

$$\Lambda = \begin{bmatrix} 0 & 0.5\lambda \\ \lambda & 0 \end{bmatrix}$$

Similar to (a) the decision area now becomes:

$$\lambda_{12} p(x|\omega_1) P(\omega_1) = \lambda_{21} p(x|\omega_2 1) P(\omega_2) \Rightarrow 0.5\lambda p(x|\omega_1) = \lambda p(x|\omega_2)$$
$$\exp\left(-\frac{x^2}{2}\right) = 1 \Rightarrow x = 0$$

The corresponding areas now become:

$$R_1 = (-\infty, -\sqrt{2\pi}) \cup (\sqrt{2\pi}, +\infty)$$
$$R_2 = (-\sqrt{2\pi}, \sqrt{2\pi})$$

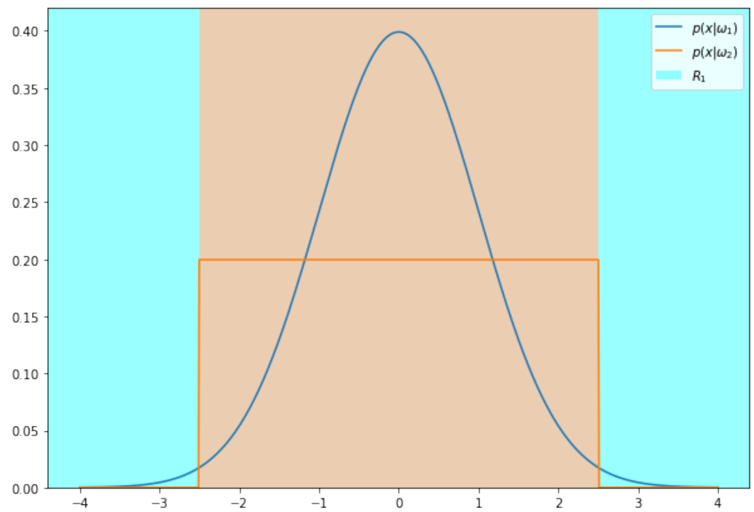

Figure 6: ω_1 and ω_2 with decision regions R_1, R_2 ($\lambda_{21} \neq \lambda_{12}$).

We can now define the average risk, which corresponds to the area under the normal distribution for $x \in [-\sqrt{2\pi}, \sqrt{2\pi}]$:

$$
\begin{aligned}
r &= \sum_{k=1}^{M} \left(\sum_{k=1}^{M} \lambda_{ki} \int_{R_i} p(\boldsymbol{x}|\omega_k)\ dx \right) P(\omega_k) \\
&= 0.5\lambda \int_{-\sqrt{2\pi}}^{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\ dx\ 0.5 \\
&= 0.247\lambda
\end{aligned}
$$

(c) Compare the results obtained from (a) and (b) and comment briefly on them.

As seen from Figures 5 and 6 we can see that when we consider one Class more important than the other, the decision areas change drastically. In the first case, the error was basically the integral of the normal distribution, where the uniform is larger than the normal plus the integral of the uniform from -1.177 to -1.177. In the second case the error is the integral of the normal distribution in the entire area of the uniform, which, for equal s would be higher. Finally the average risk is a function of lambda and the absolute value depends on the values we chose.

Exercise 6.

(a) Assume that we have our disposal a set of N data points, with $N \to \infty$, that stem from the 2-dimensional uniform distribution in [0,1]x[0,1]. Let us pretend that we have at our disposal only the data set consisting of the N data points and no additional information (regarding the form of the distribution from which they stem). Give an argumentation for supporting the fact that the pdf estimate based on the kNN density estimation method at any point in the square [0,1]x[0,1] is approximately constant (which, of course, is in line with the fact that the data points stem from a uniform distribution).

> Suppose we chose two points x_1, x_2 inside the square [0,1]x[0,1]. For these points we need to calculate the distance of a k-th nearest neighbor. By assuming that $k = 1$ we can write:
>
> $$p(x) = \frac{k}{NV(x)}$$
>
> In order to calculate the distance from x_1's and x_2's nearest neighbors we would have to find the closest point in [0,1]x[0,1]. We can plot some points in the finite space to get a better understanding of the data, and then generalize for infinite points. In this case the value of $p(x)$ depends on the position, but what happens

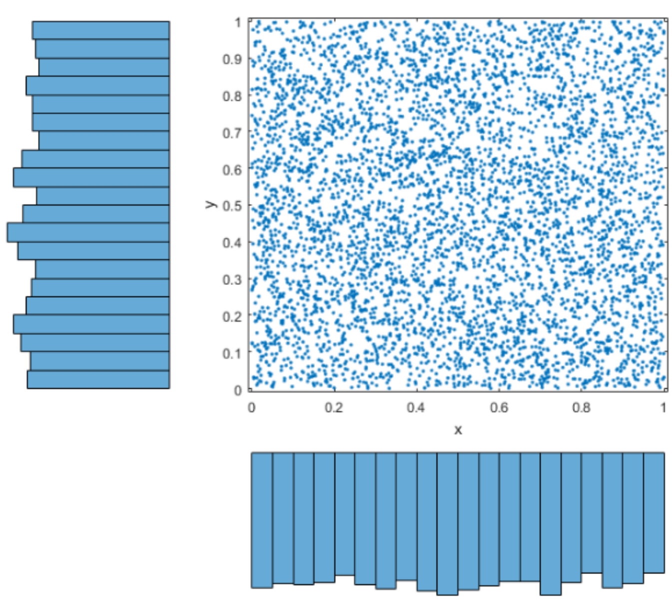

Figure 7: Finite case.

when $N \to \infty$? It is easy to understand that for any point, we can draw a circle which will be completely filled with points, since the whole space will be filled with points. The radius of that will include the same number of points, no matter where x is positioned in the space.

(b) Assume that we have our disposal a set of N data points, with $N \to \infty$, that stem from the 2-dimensional zero mean normal distribution, with identity covariance matrix, $\mathcal{N}(\mathbf{0}, I)$ Let us pretend that we have at our disposal only the data set consisting of the N data points and no additional information (regarding the form of the distribution from which they stem). Give an argumentation for supporting the fact that the pdf estimate based on the kNN density estimation method at the point $\mathbf{0}$ is greater than that at the point $[1, 1]^T$ (which, of course, this is what is expected, since the data points stem from a normal distribution).

We will approach the problem using a graph and a broad explanation. Looking

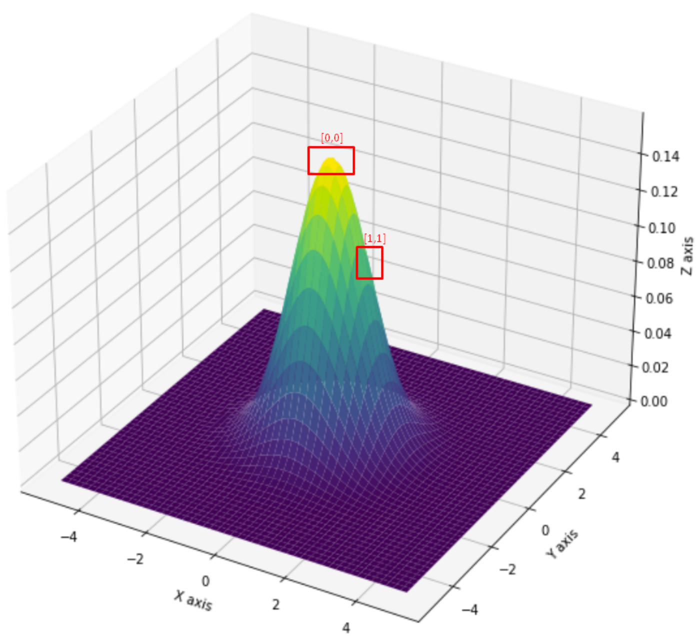

Figure 8: 2-d normal distribution.

at Figure 8 we can see that at $[0, 0]^T$ there are neighbor in all directions, thus there will be more points inside a circle (whose radius will be a chosen k) going in all directions. On the other hand, at $[1, 1]^T$ the points will be less for two reasons. First of all, the slope is larger, and most importantly the are not points in all directions, rather there are only points on one side of the 2-d normal.