Maura Pintor, PhD Student

maurapintor

https://maurapintor.github.io

@maurapintor

https://www.linkedin.com/in/maura-pintor

maura.pintor@unica.it

Survey: https://forms.gle/HfNfRJbi7EkMg5u46 (until Friday)

Exercise 1

Given the following patterns belonging to three different classes A, B, and C

A	1.1 1.3	1.7 1.4	1.2 2.0	1.6 1.9
B	2.7 1.4	2.6 1.2	2.2 2.0	2.2 1.3
C	1.4 2.5	1.2 2.4	1.8 2.6	1.5 2.9

We want to classify the unknown pattern:

x_t=(2; 2)'

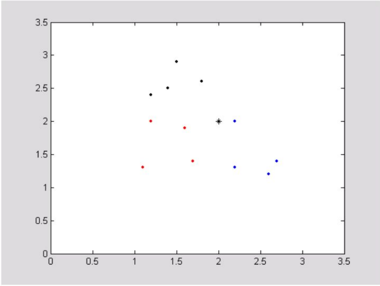
but we do not know from which probability distribution the pattern has been generated. Then, we can use a non-parametric method like the *k*-nn pattern classifier.

A) Classify the pattern x_t with values of *k*= 1, …, 4 using the *Euclidean* and the *Manhattan* distance.
Manhattan distance: |x1 – x2| + |y1 – y2|.

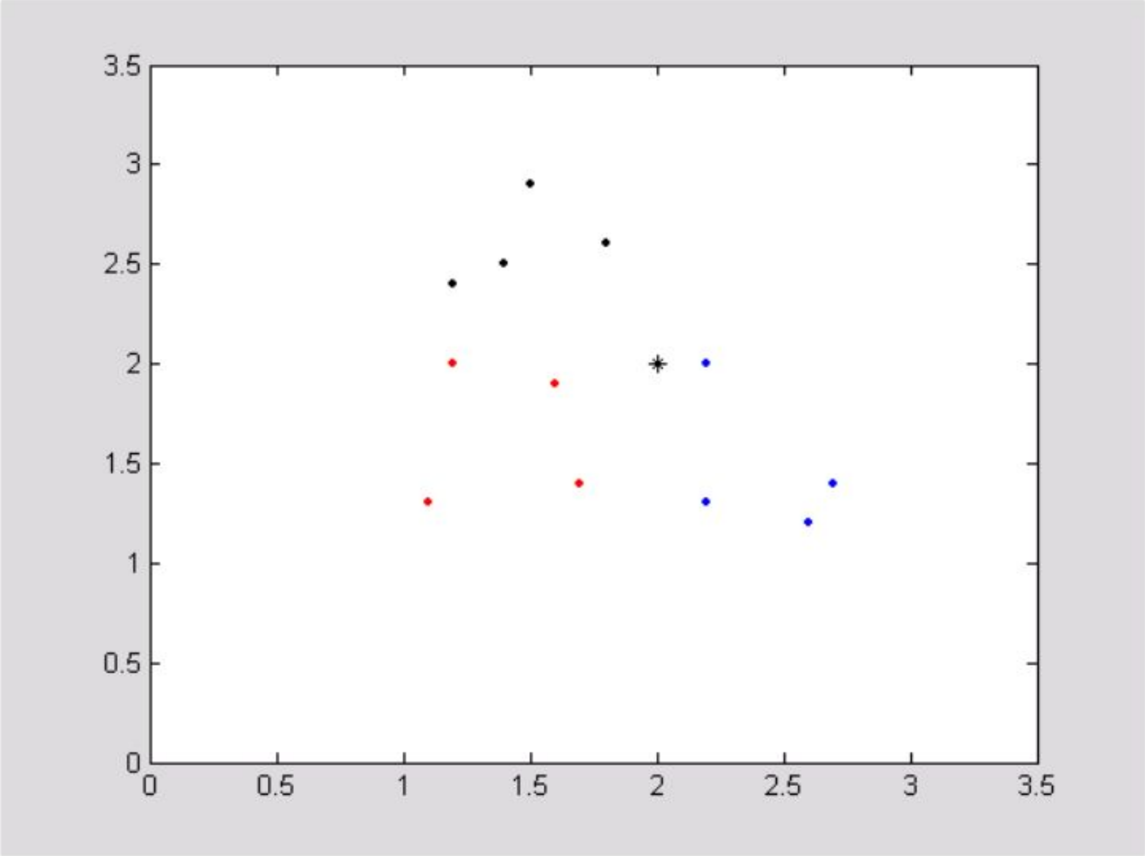
B) Use the "*leave-one-out*" method to select the best value of the "k" parameter between k=1 and k=4, using the *Euclidean* distance. The "*leave-one-out*" method works as follows:

1) Given the training set D with *n* patterns (12 patterns in this exercise)
2) for i=1, …, *n,* use the training set {D -{x_i}} and then classify the pattern x_i left out.
3) Repeat the point (2)
4) Compute the error probability (number of errors for the classifications of the *n* patterns left out)

You should use the above "*leave-one-out*" method for k=1 and k=4 and then select the value of the k parameter that provides the minimum error.



Class A: red points; Class B: blue points; Class C: black points;



Class A: red points; Class B: blue points; Class C: black points;

Linear Decision Functions

- Given a regression problem and the following data samples

$$\mathbf{X} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 2 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{bmatrix},$$

- find the linear discriminant function via ordinary least squares (OLS), i.e., by minimizing:

$$L_r(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(\mathbf{w}^T\boldsymbol{x}_i + b - y_i)^2$$

- Initialize $\boldsymbol{w} = [0.1, 0.1]^T,\ b = 0.1,\ \eta = 0.1,\ \theta = 0.06$

iter	$f(\boldsymbol{x})$	L_r	$\nabla_{\boldsymbol{w}} L_{\mathrm{r}}, \nabla_b L_r$	$\boldsymbol{w}, b$	t	θ=0.06
1.	0.1*[0 0 1 3 4 4]	0.240	[1. 1.4], 0.6	[0. -0.04], 0.04	0.3	
2.	0.01*[4 8 4 -4 -4 0]	0.057	[-0.28 -0.64], -0.52	[0.028 0.024], 0.092	0.14	
3.	0.01*[6 7 9 14 17 17]	0.014	[0.25 0.32], 0.104	[0.0032 -0.008], 0.0816	0.07	
4.	0.01*[8 9 8 7 7 8]	0.003	[-0.05 -0.14], -0.136	[0.008 0.006], 0.09	0.03	

Exercise 4

Let us suppose that we want to diagnose a disease of which we know the prior probability:

$P(\omega_{sane=healthy})=0.85$, $P(\omega_{AFFECTED})=0.15$

$P(\omega_{AFFECTED})$ is the prior probability that a person within a given population is affected by this disease.

Sane=Healthy.

The disease can be diagnosed by the amount of a certain substance in the blood. The amount of this substance is higher for people affected by the disease.

Let $\mu_s=4$ and $\mu_a=8$ be the average amount of this substance, respectively, for people *not affected* (**healthy** people) and *affected* by the disease. The amount of substance in the two cases is Gaussian distributed around the average value, with $\sigma=1$

$p(x|\omega_i) = N(\mu_i, \sigma^2)$; i=1 healthy, i=2 affected

Continues in next slide...

Continues from prev. slide ...

How to solve this?

The error is now lower than the Bayesian one without reject option. However, we do NOT classify (we reject) 15 patients every 1000.

$$\int_{x_{S1}}^{x_{S2}} p(x/\omega_{AFFECTED})\, P(\omega_{Affected})\, dx + \int_{x_{S1}}^{x_{S2}} p(x/\omega_{SANE})\, P(\omega_{SANE})\, dx \cong 15.22 \times 10^{-3}$$

It is worth noting the difference of the Chow's criterion with respect to the empirical criterion we initially used. In the latter case, we misclassify 7 patients every 1000, we reject 24 patients every 1000.

Exercise 3

Let us suppose that we want to discriminate between normal and intrusive network traffic, namely, two data classes ω_N, normal traffic, and ω_{INT}, intrusive network traffic. We suppose to use a single *feature* x to characterize traffic data (one-dimensional feature space), and we assume that the model of the network traffic is the following:

$$P(\omega_N) = \frac{1}{2}; P(\omega_{INTR}) = \frac{1}{2}$$

$$p(x/\omega_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2\right];$$

$$\mu_N = 0; \mu_{INTR} = 4; \ \sigma_N = \sigma_{INTR} = 1;$$

Let the cost of missing the detection of intrusion be ten times higher than the opposite error (a normal traffic is wrongly recognized as an intrusion).

a) Determine the decision regions using the likelihood ratio, <u>without</u> considering the costs of errors.

b) Specify the loss (cost) matrix that satisfies the above assumption.

c) Determine the decision regions that minimize the risk, and compute the related classification error.

Links used today

- MachineLearningCheatSheet.pdf
- ML-tutor-03-whiteboard
- https://colab.research.google.com/drive/1pt3eLWoqeRTHHv5IyJeSnLhu7HbX-ikX?usp=sharing