Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966

Ex1: (Introduction) Find an <u>example</u> in order to explain each of these concepts:

a) **Feature:**
   i) The variables which the algorithm use. It should be done some pre-processing. Ex: image pixel values
b) **Supervised learning:**
   i) An algorithm that optimizes their output trying to fit the test input → output pairs. Ex: Linear regression
c) **Classification**
   i) Discriminating the data into different classes.
d) **Regression**
   i) Obtaining a formula that models the data. For some input vector it should output some numerical value(s)
e) **Unsupervised Learning**
   i) An algorithm that finds patterns/ characteristics inside the data without any required output given.
f) **Clustering**
   i) Creating groups in the data.

<u>Gaussian Mixture Model</u>

Ex2: (Bayes Theorem) Walking through the jungle, you see a large feline shadow. However, you can not identify what animal it is. You know that, in this area of the jungle, 90% of felines are not dangerous cats whereas the other 10% are pumas. In addition, you know that 80% of cats are smaller than the animal you are seeing whereas 85% of pumas has a similar size. Would you run?

P(puma | large) = P(large | puma) * P(puma) / P(large)

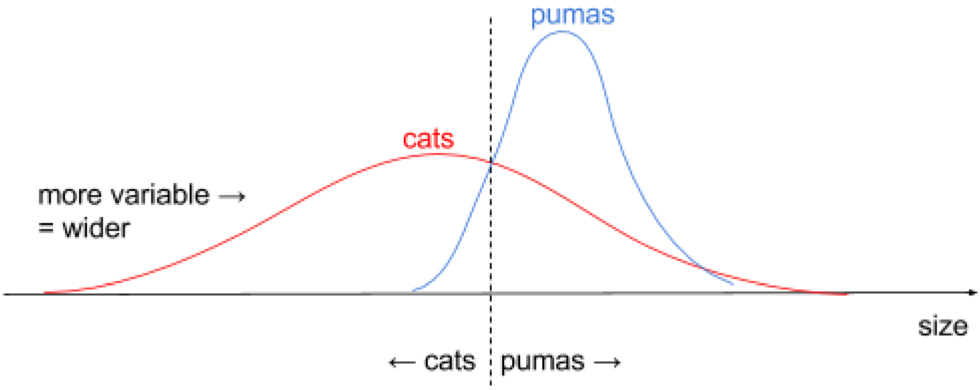
P(puma) = 1 - P(cat) = 1 - 0.9 = 0.1
P(small | cat) = 0.8
P(large | cat) = 0.2
P(small | puma) = 0.15
P(large | puma) = 0.85

P(puma | large) = 0.85 * 0.1 / [P(puma)*P(large | Puma) + P(cat)*P(large | cat)] = 0.085 / [0.085 + 0.9*0.2] = 0.3208

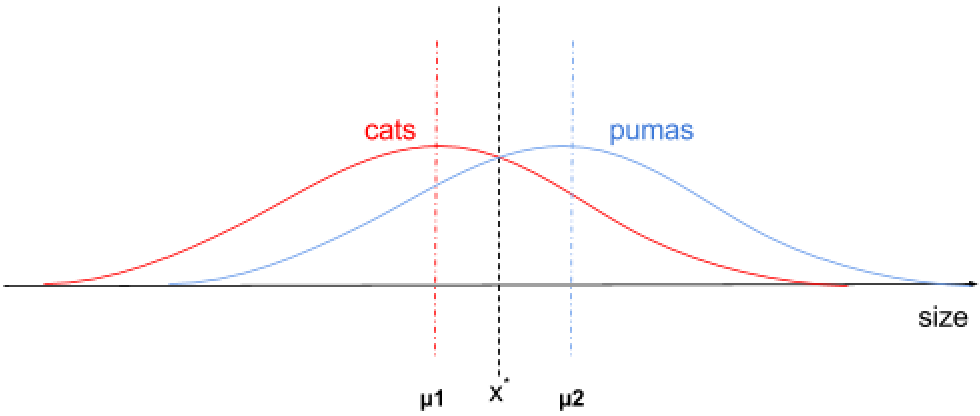
Ex3: Following the last example, assume that the variable size is continuous:

a) **Draw a figure (approximated) of the distribution functions p(size | cat) and p(size | puma). To do so, assume that the size of cats and pumas follow a normal distribution where cat's size tend to be smaller than pumas. Moreover, you know that the size of different cat species use to be more variable than in the case of pumas.**
b) **Over the same figure, and following the Bayesian Decision Theory, draw the 2 decision regions which classify cats and pumas with respect to their size. Draw also the region which corresponds to the decision error.**

Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966



Ex4: Assume that probabilities p(size|cat) and p(size|puma) follow two normal distributions N(μ1, σ1) and N(μ2, σ2) respectively. Given that p(cat) = p(puma) = 0.5 and σ1 = σ2 , find the size value x* which corresponds to the decision boundary classifying cats or pumas. Define x* as a function of μ1 and μ2.



$x^* = (\mu 1 + \mu 2) / 2$

<u>Multivariate Gaussian Mixture Model</u>

Ex5: Imagine that you have a hen and a goose in your home. One day, you find an egg but you do not know who have laid it. You measure the weight and the height of the egg, which are 60g and 5cm respectively.
Assuming that the laying frequency of the chickens is the double than the gooses one, and the following mean and covariance matrices;

$\mu_1 = (54,5)$ $\qquad \sigma 1 = \begin{pmatrix} 5 & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{2} \end{pmatrix}$

$\mu 2 = (65,6)$ $\qquad \sigma_2 = \begin{pmatrix} 8 & \frac{1}{5} \\ \frac{1}{5} & 1 \end{pmatrix}$

Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966

Which kind of egg do you think that you have?

Not studied yet

Ex6: Download the data from the dataset that you will find in https://archive.ics.uci.edu/ml/datasets/Iris and build a Multivariate Gaussian Mixture Model. Use it to classify the following unlabeled samples:

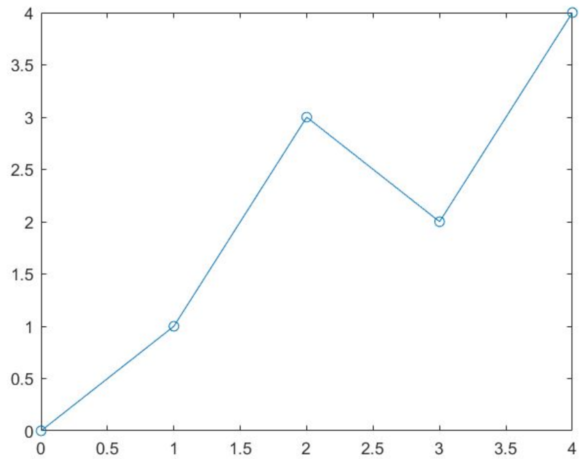
Sepal length	Sepal width	Petal length	Petal width
4,9	3,2	1,7	0,2
5	3,2	1,6	0,5
5,5	2,8	3,6	1,3
7,1	3,1	6,1	1,7

Not studied yet

PCA

Ex7: Suppose that we have a set of 2D samples: [0, 0], [1, 1], [2, 3], [3, 2] and [4, 4].

a) **Draw the data and compute the covariance matrix.**

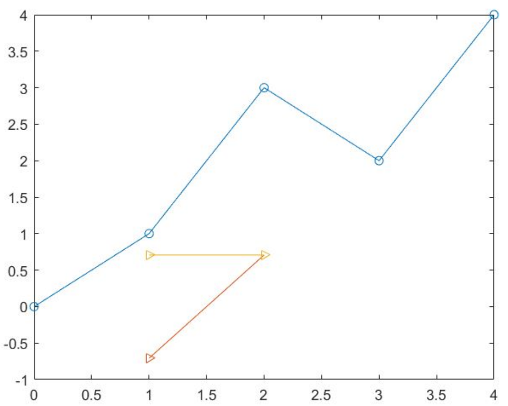


covMtx =

2.5000 2.2500
2.2500 2.5000

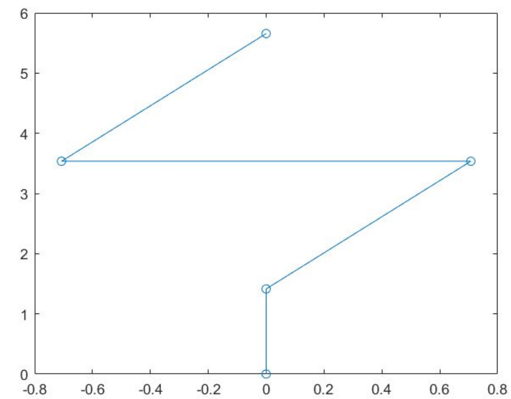
b) **Apply PCA over these samples and find the basis where the data have the maximum variance. To do so, find the eigenvalues and eigenvectors of the data covariance matrix. Draw these basis in the same figure than the previous exercise.**

Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966



Eigenvectors and data

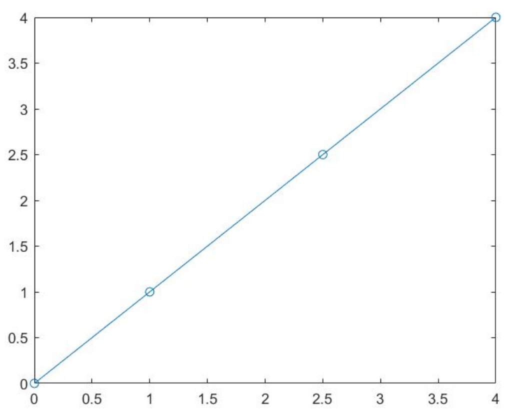
c) **Project the data over the obtained basis with PCA. Discuss the relation existing between the covariance of the projected data and the eigenvalues computed in b).**



New based values

The data rotates with the eigenvectors to maximize the separation.

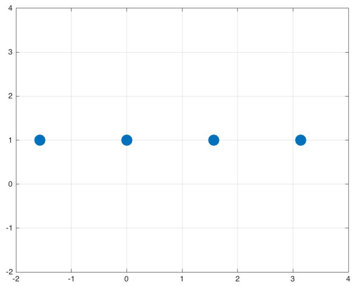
d) **Project and re-project the data using only the basis of maximum variance. Draw the results over the original data.**



Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966

Ex8: Repeat the previous exercise with the Cartesian samples [0, 1], [0, -1], [1, 0], [-1, 0] but using cylindrical coordinates (you first have to transform the data to cylindrical coordinates)

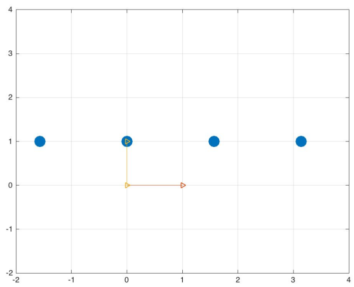
a) Draw the data and compute the covariance matrix.



covariance =

4.1123	0
0	0

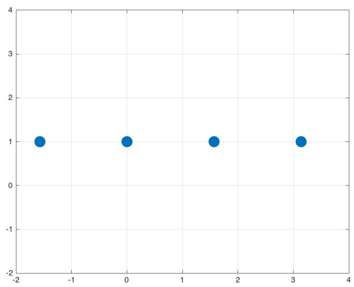
b) Apply PCA over these samples and find the basis where the data have the maximum variance. To do so, find the eigenvalues and eigenvectors of the data covariance matrix. Draw these basis in the same figure than the previous exercise.



pca_coeff =

1	0
0	1

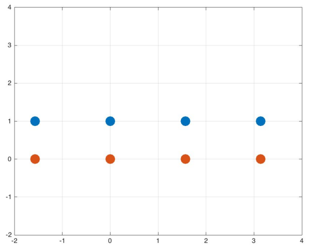
c) Project the data over the obtained basis with PCA. Discuss the relation existing between the covariance of the projected data and the eigenvalues computed in b).



Because pca returns an identity matrix, the data is not changed in any way.

Jorge Aleman Gonzalez - 174128
Miquel Floriach Clark - 172966

d) Project and re-project the data using only the basis of maximum variance. Draw the results over the original data.



With the basis of maximum variance (1,0) the values of the Y axis are lost.

Ex9: Apply PCA to the data from exercise 5 and classify again the unlabeled samples using The Multivariate Gaussian Mixture Model with the two more significant "new attributes" found with the PCA.