

**Exercise 13 : Probability, k -means algorithm, Probability and statistics,
Support Vector Machine****Exercise 13-1 : Probability**

Consider these probabilities : $Pr(B) = 0.02$, $Pr(T|B) = 0.99$, $Pr(T|B^c) = 0.05$, Calculate following probabilities :

- (a) $Pr(T)$
- (b) $Pr(B|T)$
- (c) $Pr(B|T^c)$

Exercise 13-2 : k -means algorithm

Consider the case that the distance measure :

$$\text{dist}(x, y) = \sqrt{(|x_1 - y_1|, |x_2 - y_2|) \begin{pmatrix} 4 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}}$$

for two-dimensional points is chosen to be used for the Lloyd-Forgy implementation of the k -means algorithm with $k = 2$. The algorithm has calculated the cluster centroids $\mu_1 = (0, 2)$ and $\mu_2 = (4, 7)$. Which cluster are the following points will be assigned based on k-means algorithm in the subsequent step?

- (a) $(4, 0)$
- (b) $(2, 5)$

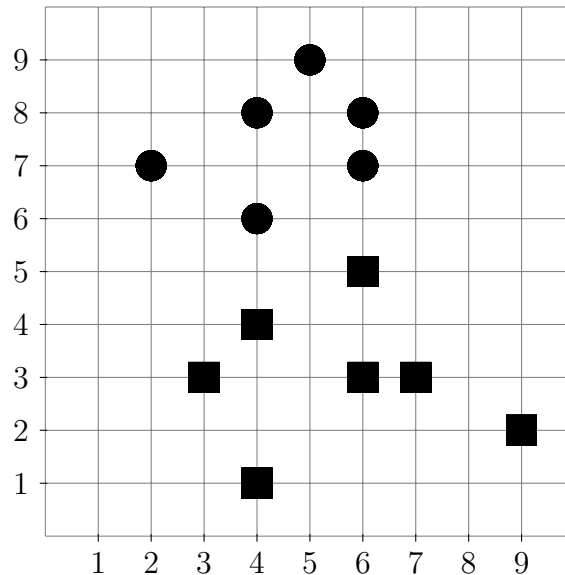
Exercise 13-3 : Probability and statistics

Clint throws darts at a circular target of radius r and is equally likely to hit any point in the target. Let X be the distance of Clint's hit from the center.

- (a) Find the CDF.
- (b) Find the PDF.
- (c) Find the mean.
- (d) Find the variance.

Exercise 13-4 : Support Vector Machine

The figure below plots a two-dimensional feature space in horizontal and vertical axes. The marks on the grid correspond to observations that are used by a binary hard-margin Support Vector Machine (SVM) classifier with linear kernel function $k(a, b) = a^T b$ for training. The labels of the positive class are denoted as circles and the negative class as rectangles. The data points are denoted as (X_1, X_2) , where X_1 is the horizontal and X_2 the vertical coordinate.



Evaluate the following statements for this SVM.

- (a) Which data points are support vectors.
- (b) Plot decision boundary. Which coordinates the decision boundary passes through.
- (c) Which data points assign to square class.