

Exercise 6 : Conditional Probability and Bayes' Theorem, k -Nearest Neighbor Classification**Exercise 6-1 : Conditional Probability**

Suppose that of all individuals buying a certain digital camera, 60% include an optional memory card in their purchase, 40% include an extra battery, and 30% include both a card and battery. Consider randomly selecting a buyer and let $A = \{\text{memory card purchased}\}$ and $B = \{\text{battery purchased}\}$.

Then $\Pr(A) = 0.6$, $\Pr(B) = 0.4$, and $\Pr(\text{both purchased}) = \Pr(A \cap B) = 0.3$.

- (a) Given that the selected individual purchased an extra battery, what is the probability that an optional card was also purchased?
- (b) Given that the selected individual purchased a memory card, what is the probability that an optional extra battery was also purchased?

Exercise 6-2 : Bayes' Theorem

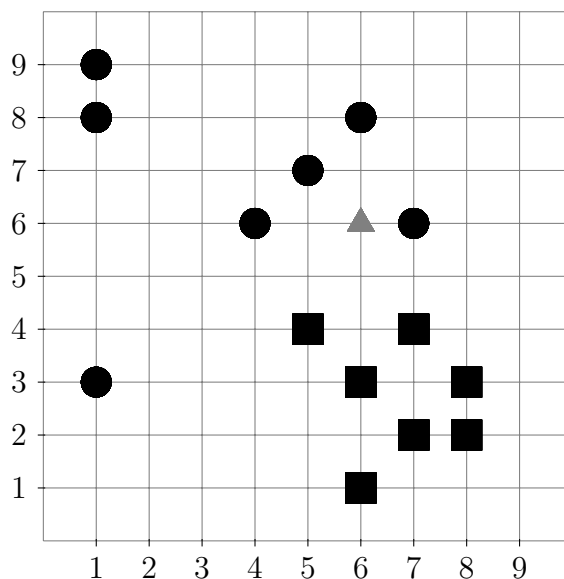
Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time.

If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

Exercise 6-3 : Nearest neighbor classification

The 2D feature vectors in the figure below belong to two different classes (circles and rectangles). Classify the object at $(6, 6)$ — in the image represented using a triangle — using k nearest neighbor classification. Use Manhattan distance (L_1 norm) as distance function, and use the non-weighted class counts in the k -nearest-neighbor set, i.e. the object is assigned to the majority class within the k nearest neighbors. Perform k NN classification for the following values of k and compare the results with your own “intuitive” result.

- (a) $k = 4$
- (b) $k = 7$
- (c) $k = 10$



Exercise 6-4 : Nearest Neighbor classification

Find a scenario where we have a set of at least four points in 2 dimensions, such that the Nearest Neighbor classification ($k = 1$) only gives incorrect classification results when using any of these points as query points and the rest as training examples. Use Euclidean distance as distance function.