PRML Tutorial 1

13^{th} Aug 2021

Regression vs. Classification vs. Clustering

Identify which paradigm should be used in the following scenarios:

Problem 1: Given a set of spam mails and another set of non-spam mails, one needs to arrive at method to characterize a new mail as spam or non-spam.

Solution:

Problem 2: Given measurements of multiple flower, we need to group together the flowers such that samples belonging to the same group are similar and the flowers belongs to different groups are dissimilar.

Solution:

Problem 3: Given the attributes of house along with their prices, we need to predict the cost of a house given its attributes.

Solution:

Problem 4: An eCommerce website wants to recommend similar products to similar customers for which they are required to group similar customers together.

Solution:

Problem 5: Estimating the crop yield given the amount of fertilizer (per unit land area) and the amount of water (per unit land area).

Solution:

Problem 6: A bank needs to predict whether a customer is eligible for a loan or not.

Solution:

Problem 7: Search engines typically have to deal with large amount of documents, and hence computing similarity of the query with all the documents might leading to very slow performance of the Information Retrieval (IR) system. For this reason, documents can be grouped together so that the IR system can focus on selected groups of documents to fetch relevant documents to the query.

Solution:

Probability Refresher

Problem 1: A survey shows that out of 100 men 5 play football and out of 250 women 10 play football. What is the probability that a player chosen randomly among the a population containing equal number of men and women plays football.

Solution

Problem 2: The past data of a clinic shows that 10% of the people coming to the clinic have liver disease. The past data also suggests that 5% of the people coming to the clinic consume alcohol. Also, it is known that 7% of the people having liver disease consume alcohol. An alcoholic patient comes to the clinic to have the liver test. What is the probability that his test turns out to be positive i.e. he has a liver disease?

Solution

Problem 3: Consider two discrete Random variables X and Y with joint PMF as shown in the table below.

	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$
$X = 2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$
$X = 3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$

3.a Find $P(X > 2, Y <= 3)$

Solution

	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$
$X = 2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$
$X = 3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$

3.b Find the marginal PMFs of X and Y

Solution

	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$
$X = 2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$
$X = 3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$

3.c Find $P(Y = 2|X = 1)$

Solution

	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$
$X = 2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$
$X = 3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$

3.d Are X and Y independent?

Solution