Machine Learning Algorithms: exercise 3 30.03.2023

1. Let us consider the observation table below. It contains information whether the train is on time or not under four variables that are assumed to be independent. Use the information and create Naïve Bayes classifier. What is the most probable class for case (weekday, winter,high, heavy,?)

Day	Season	Wind	Rain	Class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

2. Load in the file data3.xlsx and estimate 2D Gaussian models for class C_1 and class C_2. Model can be estimated by calculating mean and covariance for classes C_1 and C_2. First 100 samples are from class C_1 and following 100 from class C_2. Use your model and classify the remaining 20 points. What are the accuracy, specificity and sensivity of this classifier? (First ten points belong to class C_1 and the remaining ten points belong to class C_2)

3. Classify the remaining points using only Mahalanobis distance between the points and means from 2D Gaussians.

4. Classify now the last 20 points using k=1 and k=3 nearest neighbor rules and Euclidean distance measure. Which of the values of k seem to give more accurate results? Repeat the classification using Manhattan, Cosine and Chebyshev distances as proximity measure. (doc knnsearch) Is there a measure that is the most suitable for classifying the points into classes C_1 and C_2.

5. Repeat the classification using linear discriminant classifier. (doc fitcdiscr)

6. Classify the data once more using naïve Bayes classifier. (doc fitcnb).