CS 464 HOMEWORK 1 REPORT

Elif Gamze Güliter 21802870

Q1)

1.1

P(A, heads) in 8th trial= $(1 – P_1P_3)^7(P_1P_3)$

(getting heads in the first 7 trials) + (getting heads in 8th trial) = 1

Getting heads in 8th trial = 1 – (getting heads in the first 7 trials)

1.2

E[X + Y] = E[X] + E[Y] E[S] = $10\ P_3P_1\ +10P_2 - P_3P_2$

1.3

1.3.b

P(oliver predicting heads) = 0.99

P (oliver predicting heads 8 times in a row)) = 0.99^8

1.3.c

- P(oliver not head | result = heads) =

P(oliver not heads and result = heads) / P(result is heads)

- P(result = heads| oliver not head) =

P(oliver not heads and result = heads) / P(oliver not heads)

- P(oliver not heads and result = heads) = P(result = heads| oliver not heads) *

P(oliver not heads)

P(oliver not heads)= 0.01

P(result = heads| oliver not heads) = 0.01

P(oliver not heads and result = heads) = 0.0001

P(oliver not head | result = heads) = 0.0001 / P (result = heads)

= 0.0001 /0.4 = 0.00025

Q2)

2.1 I used Euclidian distance metric simply because it is easier to implement, and it is the measure of line distance between two points. However, since the range is very different for each parameter, I needed to implement normalization. Other distance metrics include Manhattan and Minkowski. Manhattan is simply the sum of differences between rows and columns. Minkowski is a method which generalize Euclidean and Manhattan metrics.

2.2 Because not every parameter in data has the same effect on classification. For example, if a feature is present in every label, we cannot use it as classification purposes since it does not have a remarkable effect on label. Furthermore, some features mislead the outcome. Therefore, we should implement feature selection technique in ML models.

2.4 By using backward selection I increased the accuracy from %71 to %75. I eliminated Skin Thickness, Insulin and BMI features. Below is the table of accuracies and running times I got during backward elimination.

Features Eliminated	Running Time (seconds)	Accuracy
None	**1,52**	71,00%
pregnancy	**2,62**	70,42%
Glucose	2,072	59%
Blood Pressure	2,337	67,84%
Skin Thickness	2,5219	69,12%
Insulin	2,3164	74,22%
BMI	2,5772	71,72%
DPF	2,34	71,07%
Age	2,43	73,67%
Insulin, BMI	1,62	71,07%
Insulin ,Age	2.32	74,32%
BMI, DPF	1,96	73,67%
Insulin , DPF , BMI	1,64	73,67%
Insulin , Age , BMI	1,77	70,42%
Skin Thickness, Insulin , BMI	1,93	75,62%

Q3)

3.2

Multinomial Model

Since we used Naïve Bayes approach, we assume all parameters are independent. Therefore, we used number of words in vocabulary parameters. If we add the prior probabilities, we used 3459 parameters in total. However, I did not implement Mutual information for feature extraction.

3.4

After estimations, the multinomial model gave better percentages than Bernoulli Model. I think, this is because the multinomial model looks for the number of occurrences of a word in a document, on the other hand Bernoulli only looks for whether it occurs or not. Therefore, for noisy data, Bernoulli model is too naïve.

For running time, it takes almost 18 seconds for Multinomial Model to run. On the other hand, it takes 20 seconds for Bernoulli Model to run. However, the difference is very small and can be due to computer.