Q1 1

5 Points

A classifier that gets 100% accuracy on the training set and 65% accuracy on test set is better than a classifier that attains 85% accuracy on the training set and 75% accuracy on the test set. (True/False) Why?

False
The second classifier has better test accuracy.
The second classifier reflects the true accuracy.
The first classifier is overfitting.

Q2 2

5 Points

A and B are two events. If P(A,B) decreases while P(A) increases, how P(B|A) acts (increase or decrease) Why?

decrease
P(A,B) = P(A|B) P(B) = P(B|A) P(A)
P(B|A) = (P(A|B) P(B)) / P(A)

When P (A, B) decreases, P (B | A) decreases
When P (A) increases, P (B | A) decreases
therefore P (B | A) is reduced

Q3 3

10 Points

Suppose we want to compute 10-Fold Cross-Validation error on 100 training examples. We need to compute error N_1 times, and the Cross-Validation error is the average of the errors. To compute each error, we need to build a model with data of size N_2, and test the model on the data of size $$N_3$

- (●) N_1 = 10, N_2 = 90, N_3 = 10
- ( ) N_1 = 1, N_2 = 90, N_3 = 10
- ( ) N_1 = 10, N_2 = 100, N_3 = 10
- ( ) N_1 = 10, N_2 = 100, N_3 = 100

Q4 4

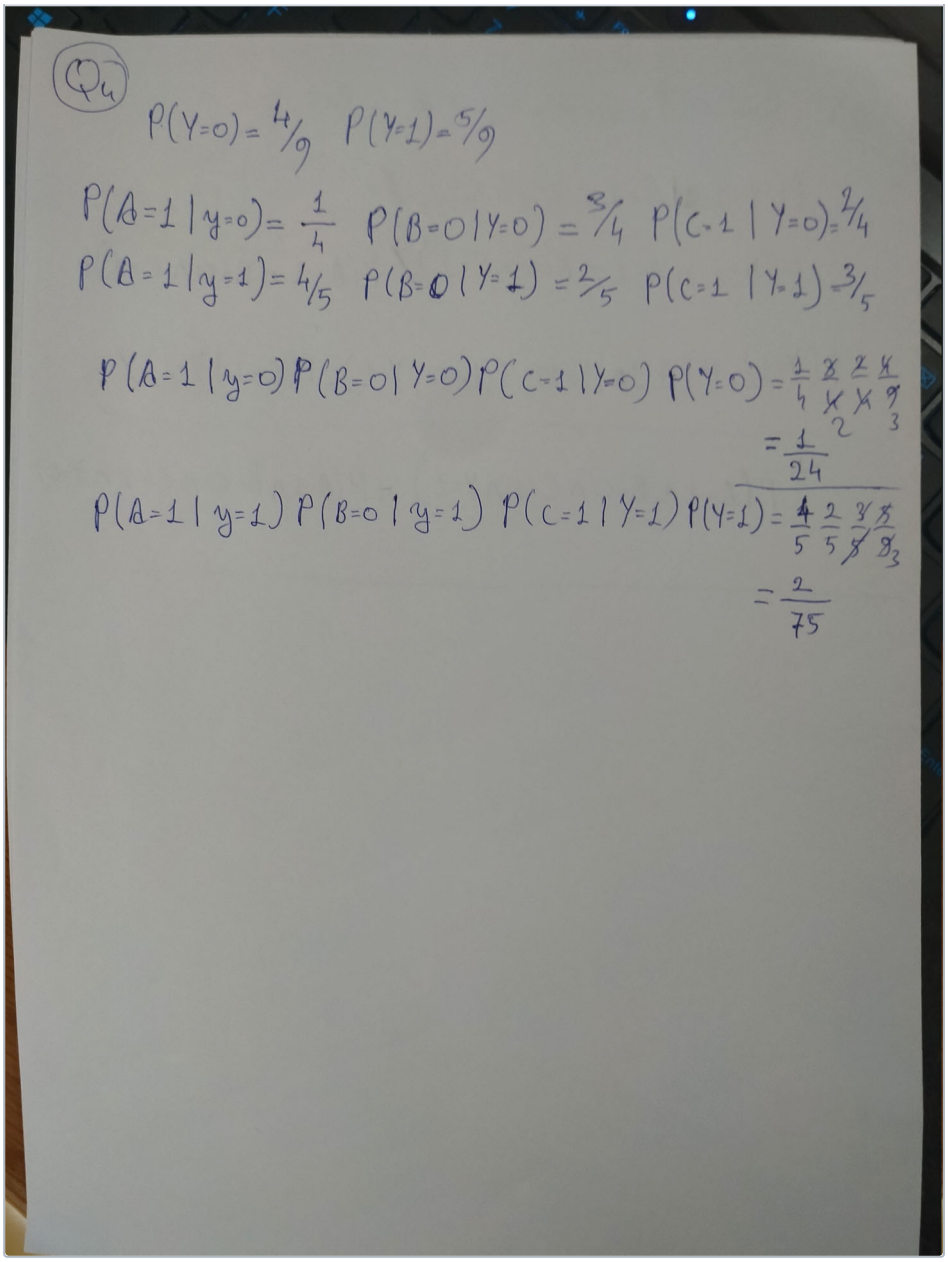
20 Points

Suppose we are given the following dataset, where A, B, C are input binary random variables, and y is binary output whose value we want to predict.

How would a naive Bayes classifier predict y given this input: A = 1, B=0, C=1.

A	B	C	y
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	1	0	1
1	1	1	1
1	0	1	0
1	0	1	1

▾ IMG_20210326_175447_N.jpg Download

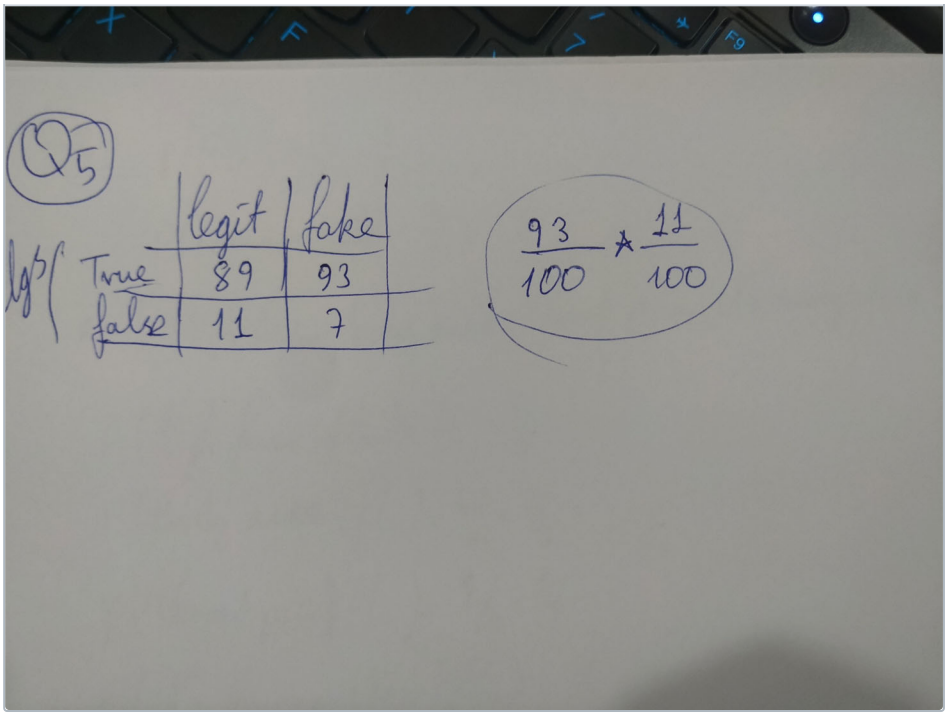


Q5 5

30 Points

Imagine that we develop an algorithm to predict fake reviews. Based on the previous experience, we know that 95% of the reviews are legit and 5% are fake. If a review is fake, there is a 93% chance that the algorithm predicts it as fake. If a review is legit, the algorithm classifies it as legit with 89% chance. What is the chance that the review is actually fake if the algorithm suggests it is fake.

▾ IMG_20210326_175456_N.jpg Download



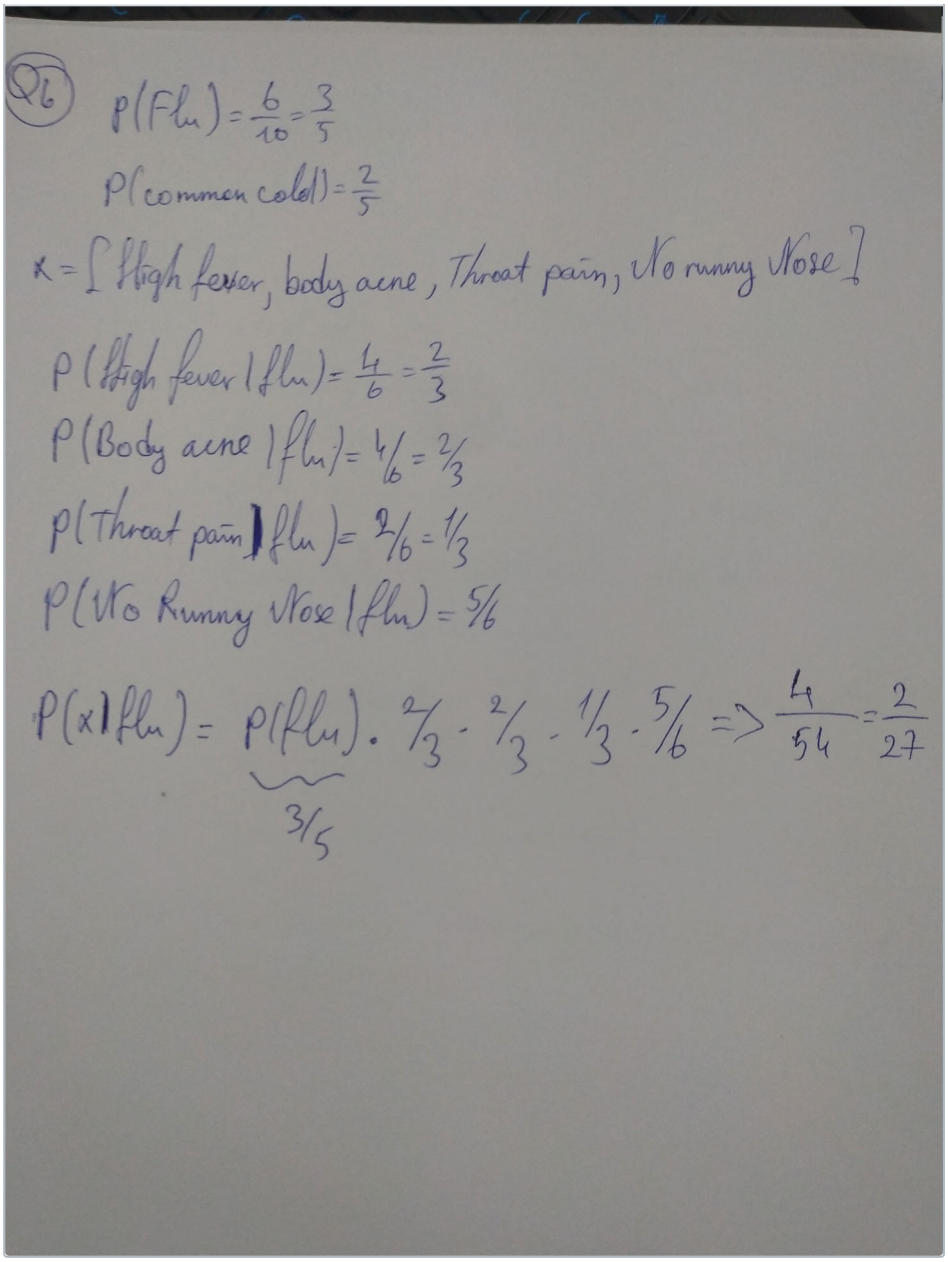
Q6 6

15 Points

Consider a naive Bayes classifier trained on the dataset given below. A new patient comes who has x=[High Fever, Body Ache and Throat Pain, but NO runny nose]. Calculate the posterior probability of having Flu given these symptoms

	Fever	Body Ache	Runny Nose	Throat Pain	Disease
1	High	Yes	No	Yes	Flu
2	High	Yes	No	No	Flu
3	High	No	Yes	No	Flu
4	Medium	Yes	No	No	Flu
5	Medium	No	No	No	Flu
6	High	Yes	No	Yes	Flu
7	Low	No	Yes	Yes	Common cold
8	Low	No	Yes	Yes	Common cold
9	Low	Yes	No	No	Common cold
10	Medium	No	Yes	Yes	Common cold

▾ IMG_20210326_175529.jpg Download



Q7 7

15 Points

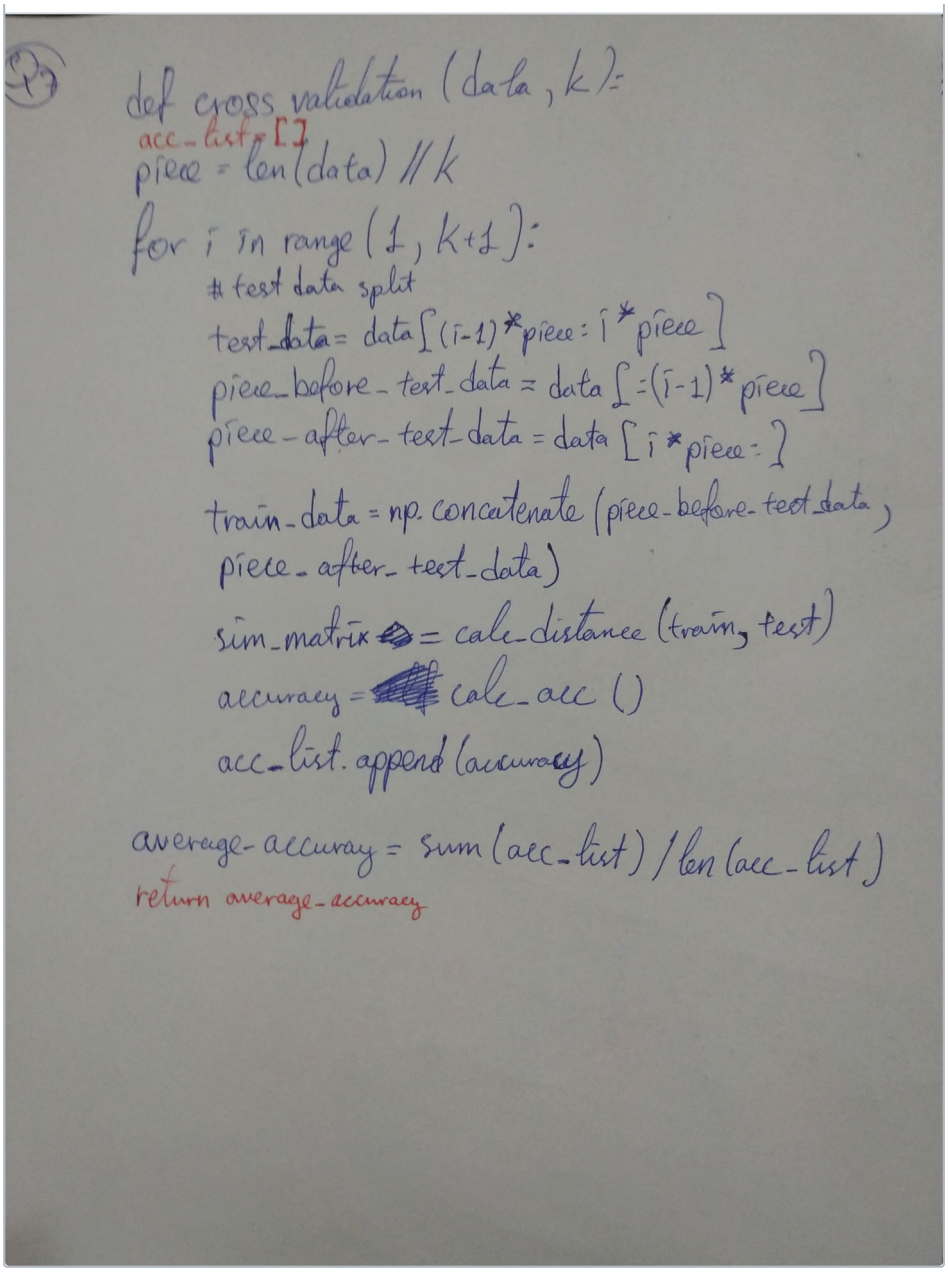
Fill the given functions for the k-fold cross validation. (You can use train_knn, and predict_knn functions in your code)

```
def cross_validation(data,k)

...

return average_accuracy
```

▾ IMG_20210326_175555.jpg Download



```
def cross_validation(data, k):
    acc_list = []
    piece = len(data) // k
    for i in range(1, k+1):
        # test data split
        test_data = data[(i-1)*piece : i*piece]
        piece_before_test_data = data[:(i-1)*piece]
        piece_after_test_data = data[i*piece:]

        train_data = np.concatenate(piece_before_test_data,
        piece_after_test_data)
        sim_matrix = cale_distance(train, test)
        accuracy = cale_acc()
        acc_list.append(accuracy)

    average_accuray = sum(acc_list) / len(acc_list)
    return average_accuracy
```

Quiz-2

GRADED

STUDENT

MEHMET TAHA USTA

TOTAL POINTS

75 / 100 pts

QUESTION 1

1 **5** / 5 pts

QUESTION 2

2 **5** / 5 pts

QUESTION 3

3 **10** / 10 pts

QUESTION 4

4 **20** / 20 pts

QUESTION 5

5 **5** / 30 pts

QUESTION 6

6 **15** / 15 pts

QUESTION 7

7 **15** / 15 pts