PARTH RATHOD

CSP - 571 Fall 2021

A20458817

Recitation Answers

Chapter 4

Question 4)

a) Feature X is uniformly distributed on [0,1] and range of observation is within 10%. Hence available observations we can use for predictions is:

```
For X [0.55,0.65] = ((0.65 - 0.55) / (1-0)) * 100
= 10%
```

b) We are given with the set of observations with P= 2 features X1 and X2 that are also uniformly distributed over the same range such that (X1, X2) € [0,1] X [0, 1]

```
So, the fraction of available observations that we will use to make
prediction is given by:
= (10%) * (10%)
= 1%
```

c) We are given with the set of observations with P= 100 features and all of which are uniformly distributed over the range between 0 and 1.

```
So, the fraction of available observations that we will use to make
prediction is given by:
= (0.1)^100 * 100
= (10)^-98 %
```

d) From the above observations we can conclude that as the number of features increases, the percentage of observations that are used to predict KNN becomes very small. Hence, more features leads to fewer neighbors.

e)

```
For P= 1 => length = (0.10)
For P= 2 => length = (0.10)^1/2 = 0.316
For P= 100 => length = (0.10)^1/100 = 0.977
```

From the above, we can conclude that when we wish to make predictions for the test observations that contains on average 10% of the training observations, then if we have large number of features, it will be better to include all the features.

Question 6)

a)

Logistic Regression with multiple variables is given by:-

$p(X)$ = e ^ (β0 + X1β1 + x2β2) / 1 + e ^ (β0 + X1β1 + X2β2)

β0 = -6, β1 = 0.05, β2 = 1, e = 2.71828

X1= 40 , X2 = 3.5
Substituting in Equation gives
p(X) = 37.755

b)

Logistic Regression with multiple variables is given by:-

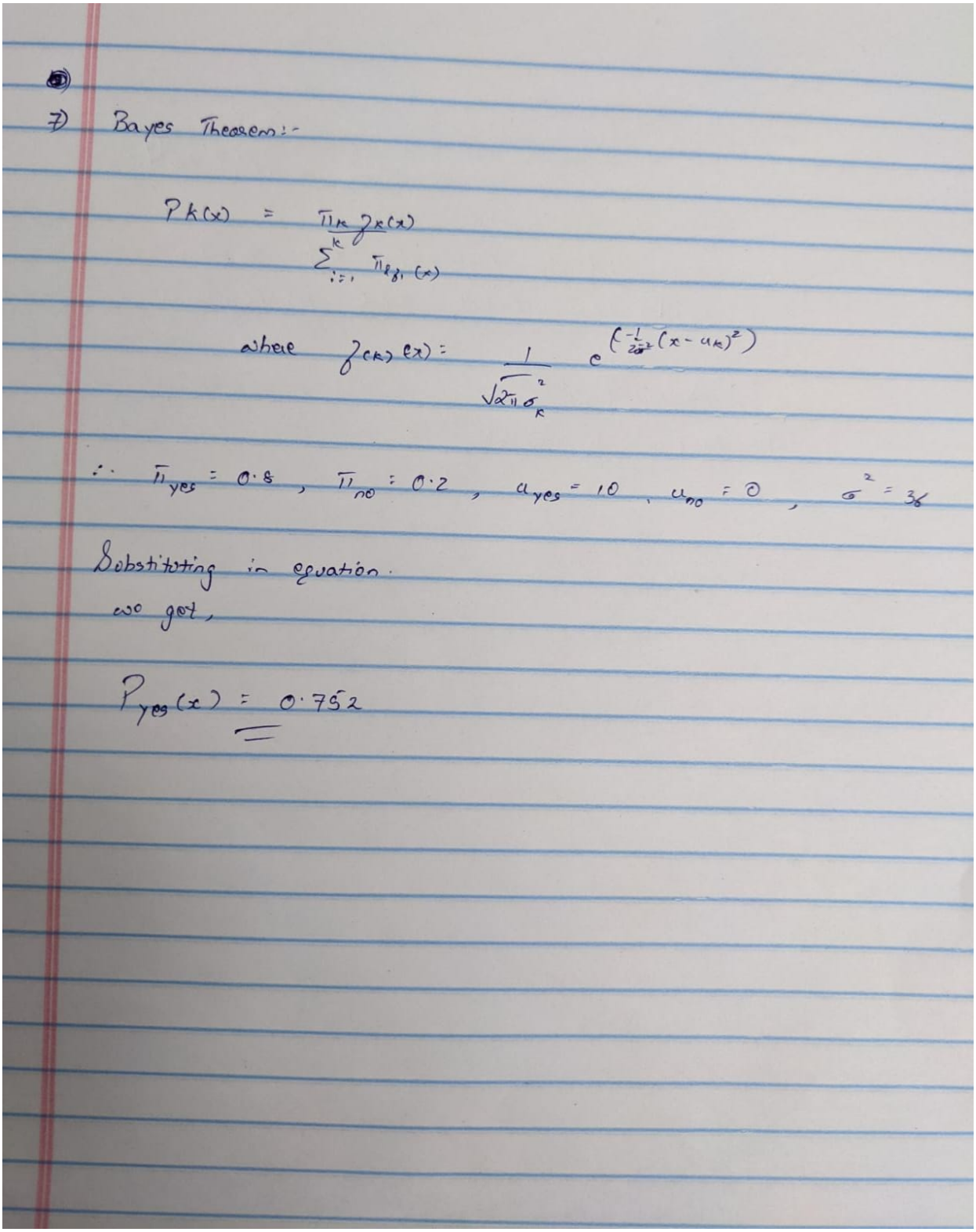
$p(X)$ = e ^ (β0 + X1β1 + x2β2) / 1 + e ^ (β0 + X1β1 + X2β2)

β0 = -6, β1 = 0.05, β2 = 1, e = 2.71828

P(x) = 0.5 , X2 = 3.5

Substituting in equation gives
X1 = 50hrs

Question 7)



7) Bayes Theorem:-

$$P_k(x) = \frac{\pi_k f_k(x)}{\sum_{i=1}^{k} \pi_{g_i}(x)}$$

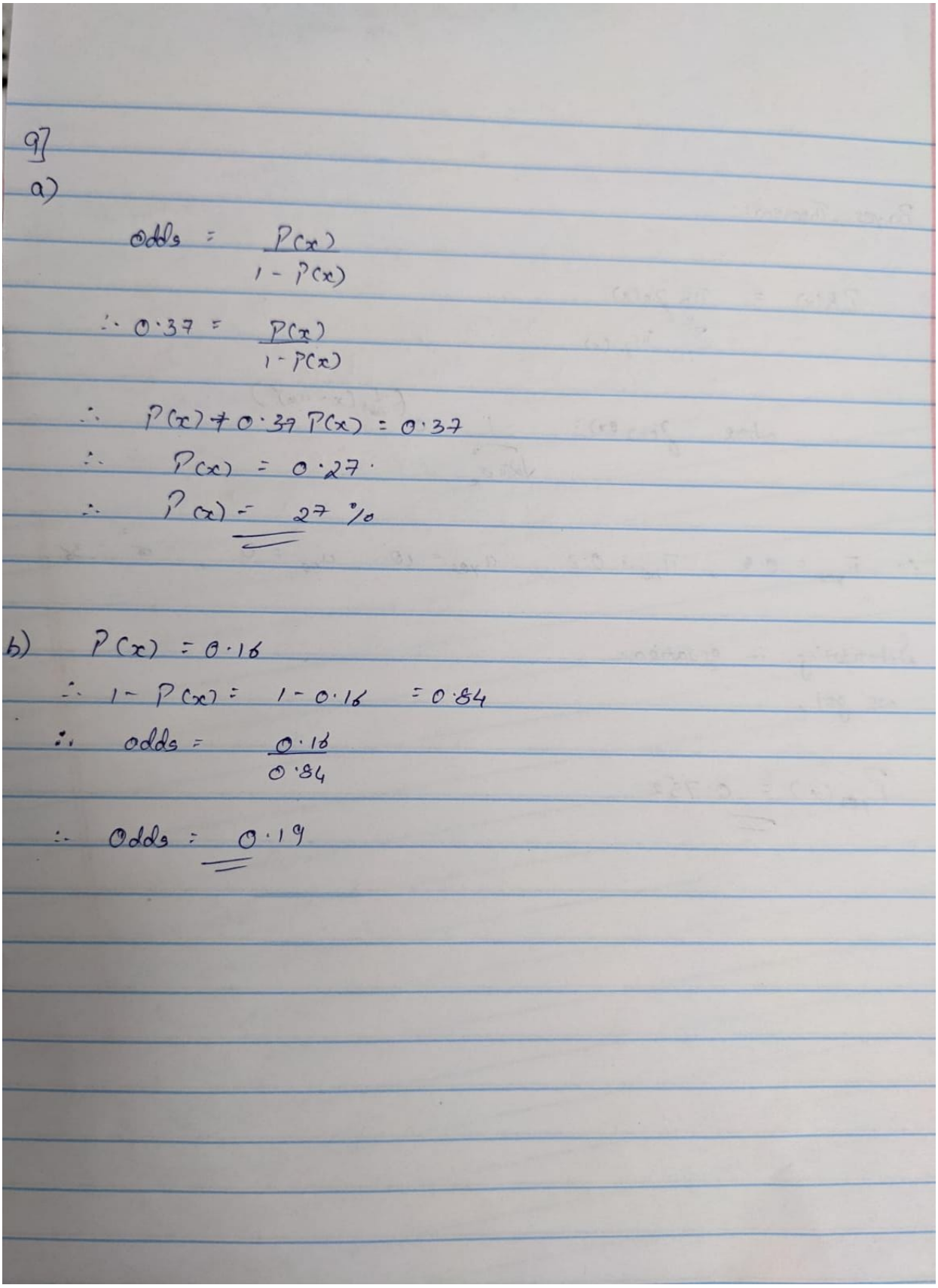
where $f_{(k)}(x) = \frac{1}{\sqrt{2\pi}\sigma_k^2} e^{\left(\frac{-1}{2\sigma^2}(x-u_k)^2\right)}$

$\therefore \pi_{yes} = 0.8$, $\pi_{no} = 0.2$, $u_{yes} = 10$, $u_{no} = 0$, $\sigma^2 = 36$

Substituting in equation.
we get,

$$P_{yes}(x) = 0.752$$

Question 9)



Chapter 5

Question 2)

a)

Given: total number of observations = n
Since bootstrap allows sampling with replacement
Every observation in original sample is independent and has equal probability to appear in each bootstrap observation
Probability that first bootstrap observation is the jth observation from original sample is = 1/n
Probability that first bootstrap observation is not the jth observation from original sample is = 1 – (1/n)

b)

Given: total number of observations = n
Since bootstrap allows sampling with replacement
Every observation in original sample has equal probability to appear in each bootstrap observation
Probability that Second bootstrap observation is the jth observation from original sample is = 1/n
Probability that Second bootstrap observation is not the jth observation from original sample is = 1 – (1/n)

c)

p(jth observation is not in the first bootstrap sample) = 1 – ($1/n$)
p(jth observation is not in the second bootstrap sample) = 1 – (1/n)
:
p(jth observation is not in the nth bootstrap sample) = 1 –(1/n)

P(jth observation is not in the bootstrap sample) = (1 – (1/n)) * (1 - (1/n)) * upto n times
= (1 - (1/n)) ^ n

d)

From above answer we can replace n by 5 and get the answer
P(jth observation is in the bootstrap sample) = (1 - (1/5)) ^ 5
= 0.67

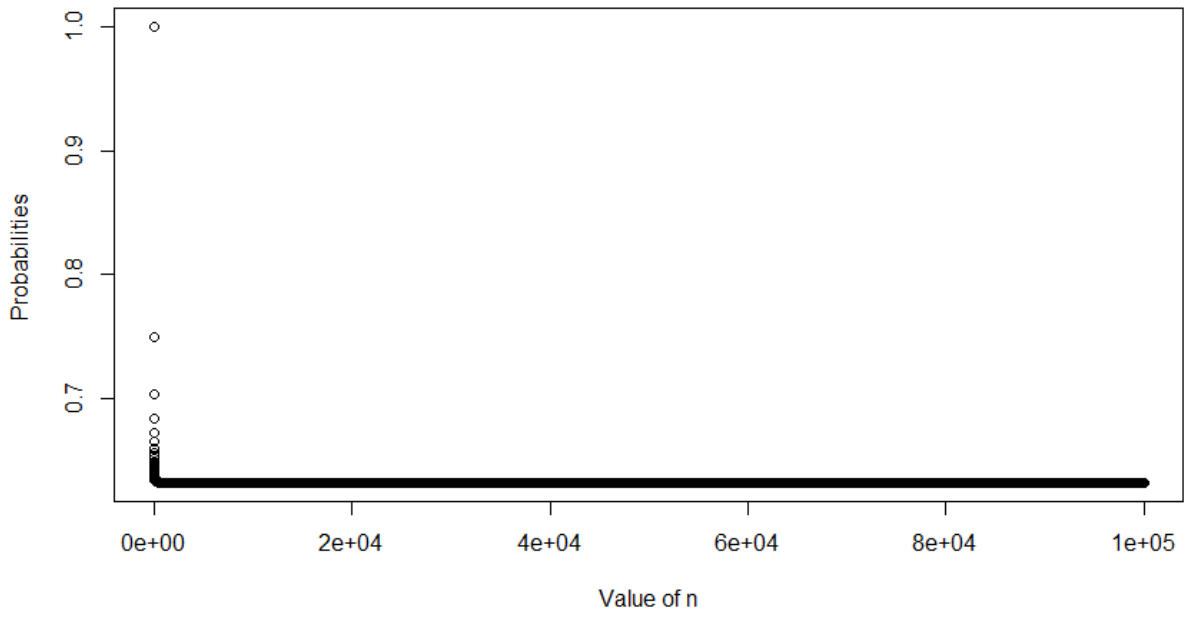
e)

P(jth observation is in the bootstrap sample) = 1 - (1 - (1/n))^n
= 1 - (1 - (1/100))^100
= 0.634

f)

```
P(jth observation is in the bootstrap sample) = 1 - (1 - (1/n))^n
                                              = 1 - (1 - (1/1000))^1000
                                              = 0.632
```

g) plot(1:100000, 1-(1-1/1:100000)^(1:100000), xlab= "Value of n",ylab ="Probabilities")



Probability quickly reaches to 0.62 and remains constant as n tends to infinity

h)

```
data <- rep(NA, 10000)
for (i in 1:10000)
{
  data[i] <- sum(sample(1:100, rep = TRUE) == 4) > 0
}
mean(data)

Answer:- 0.6289
```

Question 3)

a)

In k-fold cross validation,set of observations is randomly divided into k groups of approx equal size. The first fold is treated as validation set and remaining k-1 as where the fit method is used. The MSE is then computed. This process is repeated k times, each time a different group of observation is treated as validation set. The k-fold CV is computed by doing average of these values, CV (k) = 1/k * Summation(MSE (i) for i 1 to k)

b)

Validation set Approach:

Advantages:

This approach is conceptually simple.

This approach is easy to implement.

Disadvantages:

Validation estimate of test error rate can be highly variable, depends on which observations are included in training set and which observations are included in validation set.

In this approach, only observations which are included in training set are used to fit the model and model can perform worse if fewer.

LOOCV:

Advantages:

The LOOCV cross-validation method is a subset of k-fold cross-validation with k=n.

In this technique, each observation is only considered once in a validation set and (n-1) times in a training set.

This approach gives approximately unbiased estimates of the test error since each training set contain (n -1) observations.

Disadvantages:

It necessitates fitting the potentially computationally costly model n times, as opposed to k-fold cross-validation, which necessitates fitting the model just k times, where k < n.

This method has a greater variance than k-fold cross-validation (since we are averaging the outputs of n fitted models trained on a virtually similar set of observations, these outputs are highly correlated, and the mean of these outputs is higher).