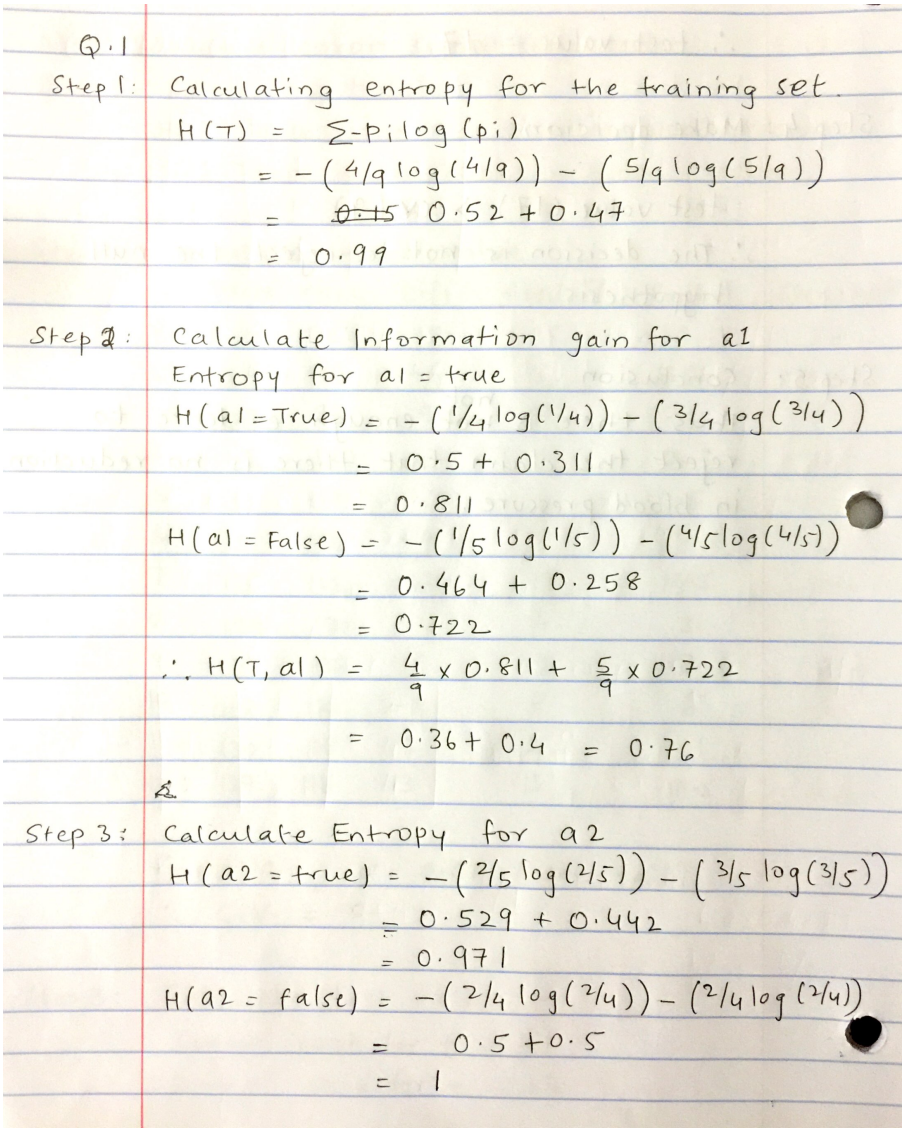
CSC591: Foundations of Data Science
HW4: Feature Selection, Resampling and Bootstrap, Bayesian Inference

Released: 11/13/15
Due: **11/23/15 (23:55pm);** (One day late: -25%; -100% after that).

Student Name: Parth Satra
Student ID: pasatra(200062999)

Q1. Information theory (10 points)
Using the data provided in the following table, rank the attributes based on information gain.

Answer
The solution is as follows:



$\therefore H(T, a_2) = \frac{5}{9} \times 0.971 + \frac{4}{9} \times 1$

$= 0.984$

Step 4: Calculate Entropy for a3.

$H(a3 = 1) = -(1/1 \log(1/1)) - 0 = 0$

$H(a3 = 3) = -(1/1 \log(1/1)) - 0 = 0$

$H(a3 = 4) = -(1/1 \log(1/1)) - 0 = 0$

$H(a3 = 5) = -(1/2 \log(1/2)) - (1/2 \log(1/2)) = 0.5 + 0.5 = 1$

$H(a3 = 6) = -(1/1 \log(1/1)) - 0 = 0$

$H(a3 = 7) = -(1/2 \log(1/2) - (1/2 \log(1/2)) = 0.5 + 0.5 = 1$

$H(a3 = 8) = -((1/1) \log(1/1)) - 0 = 0$

$\therefore H(T, a_3) = \frac{1}{9} \times 0 + \frac{1}{9} \times 0 + \frac{1}{9} \times 0 + \frac{2}{9} \times 1 + \frac{1}{9} \times 0 + \frac{2}{9} \times 1$

$+ 1/9 \times 0$

$= 2 \times \frac{2}{9} = \frac{4}{9} = 0.444$

Step 5: Calculate Informatn gain for each attribute

$\therefore I(T, a1) = H(T) - H(T, a_1) = 0.991 - 0.76 = 0.231$

$\therefore I(T, a2) = H(T) - H(T, a_2) = 0.991 - 0.984 = 0.007$

$\therefore I(T, a3) = H(T) - H(T, a_3) = 0.991 - 0.444 = 0.547$

Thus, attribute rank based on information gain is

$$\boxed{a_3 > a_1 > a_2}$$

In the above problem, it is assumed that the attribute 3 in the given data set to be categorical since there is no additional information available about the attribute. Having said that, if we know that that the attribute is distributed over a continuous wider range then the attribute can be transformed into a categorical form by techniques like binning. Binning basically assigns a categorical value to attribute between a certain range. Depending on what actually the attribute signifies the binning categories are defined.

Finally, from the output above we can see that since we have considered the attribute 3 as the categorical attribute directly, it has the highest rank. This is the side effect of information gain technique that it prefers attributes with wider spread as the one with higher gain. Attributes like primary key always will have the highest gain but they actually don't signify anything.

The time taken to solve this question is 35mins.

Q2. Dimensionality Reduction (20 points)
Find the principal components of the following two-dimentional data

x1 = 1,0,-1
x2 = -1, 1, 0

Answer
Step 1: Calculate the mean for x1 and x2
Mean(x1) = 0
Mean(x2) = 0

Step 2: Calculate the covariance Matrix
Cov(x, y) = $\frac{\sum(x\ i)(y\ i)}{n}$ (Considering the mean for both x, y is 0.
Cov(x1, x1) = {(1)(1) + (0)(0) + (-1)(-1)}/3 = 2/3
Cov(x1, x2) = {(1)(-1) + (1)(0) + (-1)(0)}/3 = -1/3
Cov(x2, x2) = {(-1)(-1) + (1)(1) + (0)(0)}/3 = 2/3

Thus covariance matrix is

	X1	X2
X1	2/3	-1/3
X2	-1/3	2/3

Step 3: Calculate Eigen Values

$$cov_{matrix} - \lambda I = 0$$

Step 3 : Calculate Eigen Values

$$\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$

$$\therefore \begin{bmatrix} (2/3 - \lambda) & -1/3 \\ -1/3 & (2/3 - \lambda) \end{bmatrix} = 0$$

$$\therefore (2/3 - \lambda)^2 - (-1/3)^2 = 0$$

$$\therefore (2/3 - \lambda - 1/3)(2/3 - \lambda + 1/3) = 0$$

$$\therefore \lambda = 1/3 \quad \text{or} \quad \lambda = 1$$

Step 4 : Calculate Eigen Vectors

$$\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = 1 \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix}$$

$$\therefore 2/3\, e_{11} - 1/3\, e_{12} = e_{11}$$

$$\therefore -1/3\, e_{12} = 1/3\, e_{11}$$

$$\therefore e_{12} = -e_{11} \qquad \text{---------- (1)}$$

$$-1/3\, e_{11} + 2/3\, e_{12} = e_{12}$$

$$\therefore -1/3\, e_{11} = 1/3\, e_{12}$$

$$\therefore e_{12} = -e_{11}$$

Now

$$e_{11}^2 + e_{12}^2 = 1$$

From (1) we have $e_{12} = -e_{11}$

$$\therefore e_{11}^2 + (-e_{11})^2 = 1$$

$$\therefore 2e_{11}^2 = 1$$

$$\therefore \boxed{e_{11} = 1/\sqrt{2}}$$

$$\therefore \boxed{e_{12} = -1/\sqrt{2}}$$

The second eigen vectors can be

$$\begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} e_{21} \\ e_{22} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} e_{21} \\ e_{22} \end{bmatrix}$$

$$\therefore 2/3\, e_{21} - 1/3\, e_{22} = \frac{1}{3} e_{21}$$

$$\therefore 1/3\, e_{21} = 1/3\, e_{22}$$

$$\therefore \boxed{e_{21} = e_{22}} \quad \text{---------- (2)}$$

Now $e_{21}^2 + e_{22}^2 = 1$

From 2 we get

$$e_{21}^2 + e_{21}^2 = 1$$

$$\therefore 2e_{21}^2 = 1$$

$$\therefore \boxed{e_{21} = 1/\sqrt{2}}$$

$$\therefore \boxed{e_{22} = 1/\sqrt{2}}$$

Thus two principal components are

$$e_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad \& \quad e_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

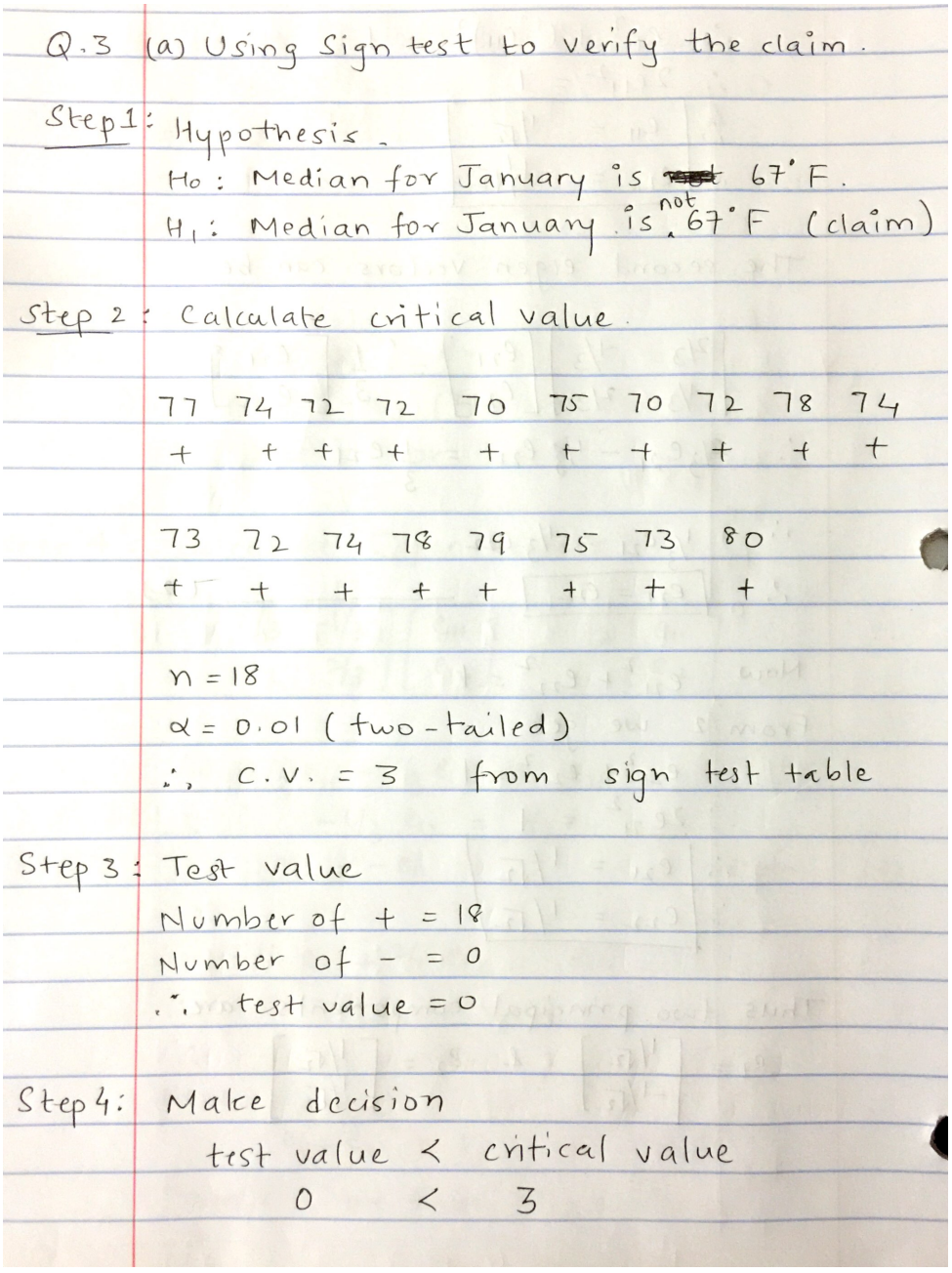
Total time taken to solve the question is 35mins.

Q3. Nonparametric Tests (30 points)

(a) (10 points) A CNN meteorologist states that the median daily high temperature for the month of January in Raleigh is 67° Fahrenheit. The high temperatures (in degrees Fahrenheit) for 18 randomly selected January days in Raleigh are listed below. At α = 0.01, can you reject the meteorologist's claim?

77 74 72 72 70 75 70 72 78 74 73 72 74 78 79 75 73 80

Answer



Q.3 (a) Using Sign test to verify the claim.

Step 1: Hypothesis.

H_0: Median for January is 67° F.

H_1: Median for January is not 67° F (claim)

Step 2: Calculate critical value.

77 74 72 72 70 75 70 72 78 74

\+ + + + + + + + + +

73 72 74 78 79 75 73 80

\+ + + + + + + +

$n = 18$

$\alpha = 0.01$ (two-tailed)

∴, C.V. = 3 from sign test table

Step 3: Test value

Number of + = 18

Number of − = 0

∴ test value = 0

Step 4: Make decision

test value < critical value

0 < 3

Thus we reject the null hypothesis.

Step 5: Thus, there is enough evidence to reject the claim that median daily high temperature for Raleigh in January is 67° F.

(b) (10 points) A drug company postulates that their vaccine will decrease the number of colds. The following table shows the results (number of colds before and after vaccination) of an experiment involving 14 subjects. At α= 0.05, verify if the company's claim is true.

Answer

Q.3(b) Using sign test to test the hypothesis.

Step 1: State hypothesis

H_0: The number of colds will not reduce

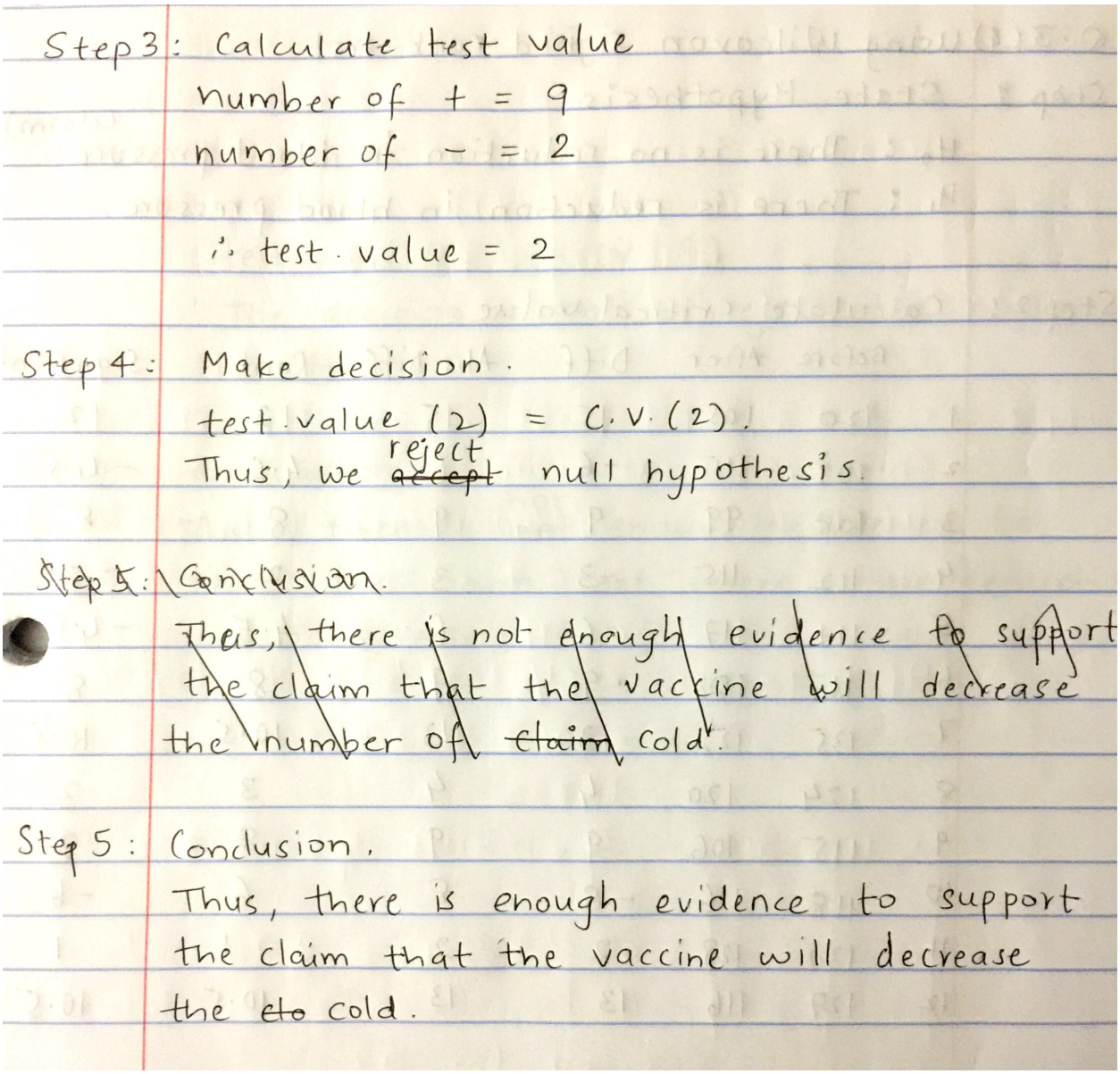
H_1: The number of colds will reduce. (claim)

Step 2: Critical Value.

1	0	2	−
2	1	1	0
3	2	0	+
4	2	2	0
5	2	3	−
6	3	2	+
7	3	1	+
8	3	3	0
9	3	2	+
10	4	1	+
11	4	3	+
12	5	2	+
13	5	4	+
14	6	3	+

$\therefore n = 11$, $\alpha = 0.05$ (single-tail)

$\therefore$ C.V. = 2 from sign test table.



(c) (10 points) Readings from a study conducted to measure impact of a food supplement on blood pressure (before and after the supplement) is given in the following table. At α = 0.01, can you reject the claim that there is no reduction in blood pressure after taking supplement?

Answer

Q.3(c) Using Wilcoxon signed rank test

Step 1: State Hypothesis

H_0 : There is no reduction in blood pressure (claim)

H_1 : There is reduction in blood pressure.

Step 2: Calculate critical value.

	Before	After	Diff	Abs. Diff	Rank	Sign Rank
1	120	105	15	15	12	12
2	109	115	-6	6	4.5	-4.5
3	108	99	9	9	8	8
4	112	115	-3	3	2	-2
5	111	117	-6	6	4.5	-4.5
6	117	108	9	9	8	8
7	135	122	13	13	10.5	10.5
8	124	120	4	4	3	3
9	115	106	9	9	8	8
10	118	126	-8	8	6	-6
11	130	128	2	2	1	1
12	129	116	13	13	10.5	10.5

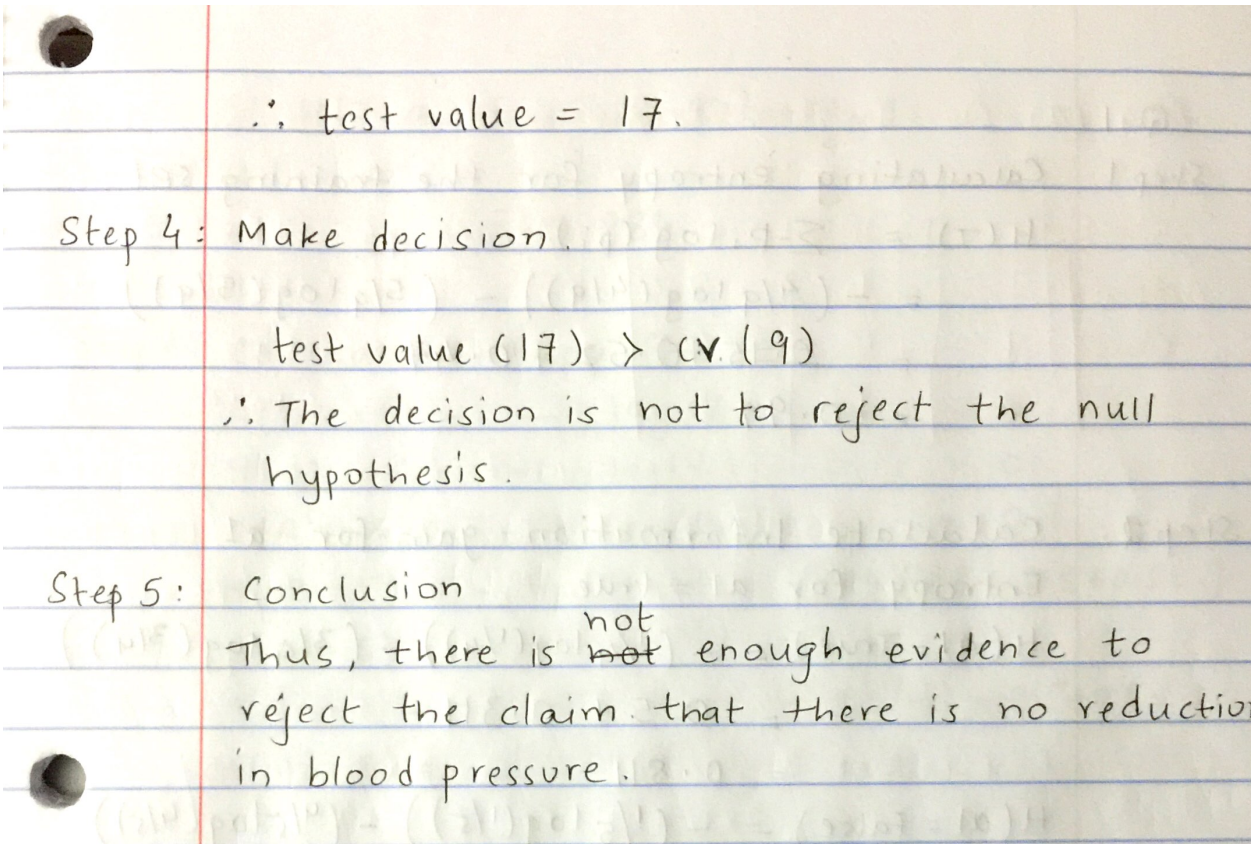
$\therefore n = 12$, $\alpha = 0.01$ (one-tail)

$\therefore$ C.V. $= 9$

Step 3: Calculate test value.

Sum of rank for + $= 61$

Sum of rank for − $= 17$



Total time taken in solving all the three parts of question 3 was 90mins.

Q4. The following table summarizes the results of logistic regression classification results on 3 datasets construed using LOOCV method. (**15 points**)

(a) (5 points) Compute the (i) construct contingency table, (ii) compute overall accuracy, (ii) precision, (iv) recall, and (v) F-measure for each classification prediction.

(b) (10 points) Construct a fourth classifier by taking majority vote on 3 classifiers, and then compute all the measures asked in (a) for the fourth classifier.

Answer

Q.4 (a) Classification Prediction 1

Contingency table

Actual Class \ Predicted Class	1	2
1	5	4
2	1	5

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$= \frac{5+5}{5+5+4+1} = \frac{2}{3}$$

Precision (p): $TP/(TP+FP)$

$= 5/(5+1) = 5/6$

Recall (r) $= TP/(TP+TN)$

$= 5/(5+4)$

$= 5/9$

F-measure : $\frac{2 * TP}{(2*TP+TN+FP)} = \frac{2\times5}{(2\times5+4+1)} = 2/3$

Classification Prediction 2

Contingency table

Actual Class \ Predicted Class	1	2
1	5	4
2	2	4

$$\text{Accuracy} = \frac{5+4}{15} = 9/15$$

Precision (p) $= \frac{5}{5+2} = \frac{5}{7}$

Recall (r) $= \frac{5}{5+4} = 5/9$

$$\text{F-measure} = \frac{2\times5}{(2\times5+4+2)}$$

$$= \frac{10}{16} = 5/8$$

Classification Prediction 3

Contingency table

Actual Class \ Prediction Class	1	2
1	6	3
2	2	4

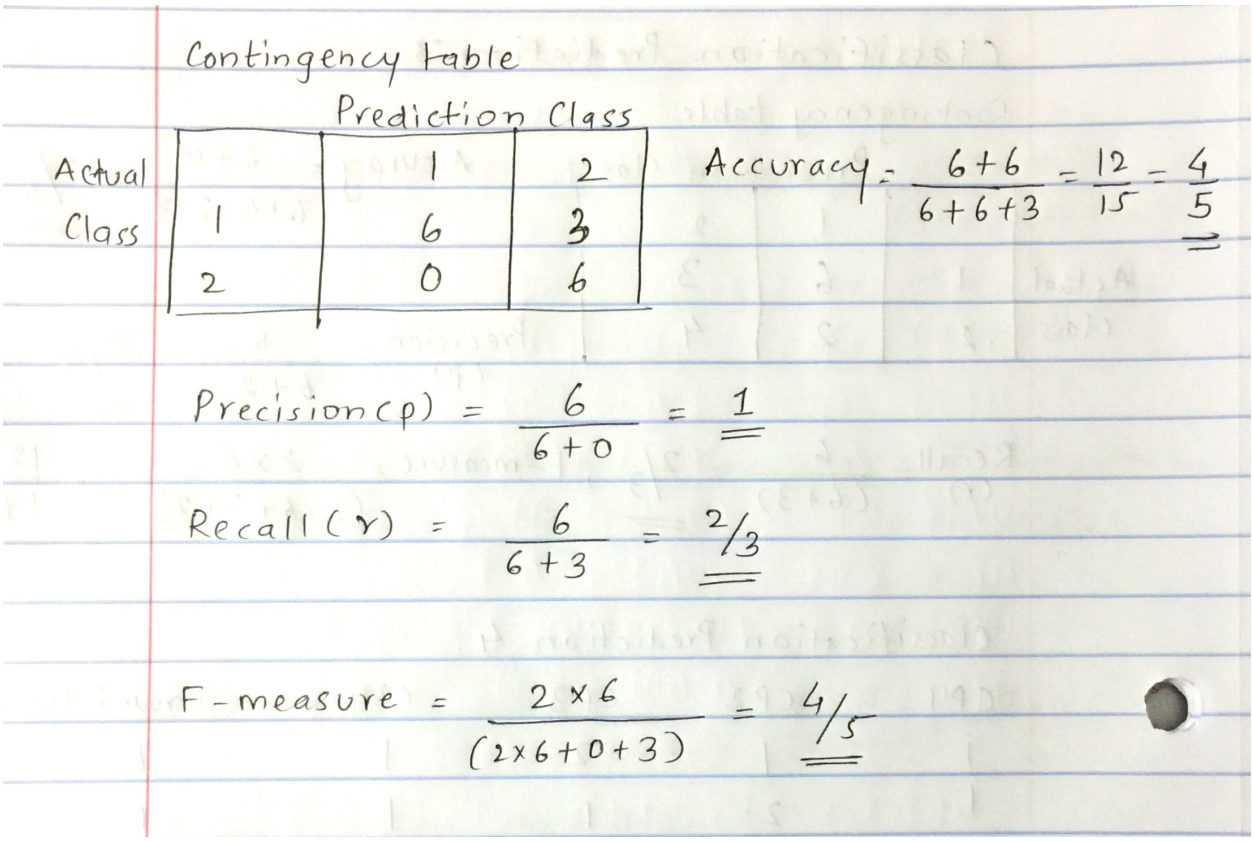
$$\text{Accuracy} = \frac{6+4}{6+4+2+3} = \frac{2}{3}$$

$$\text{Precision (P)} = \frac{6}{6+2} = \frac{3}{4}$$

$$\text{Recall (r)} = \frac{6}{(6+3)} = \frac{2}{3} \; ; \; \text{F-measure} = \frac{2 \times 6}{(2 \times 6 + 3 + 2)} = \frac{12}{17}$$

Classification Prediction 4.

CP1	CP2	CP3	CP4	Ground truth
1	1	2	1	1
1	2	1	1	1
2	1	1	1	1
1	1	2	1	1
1	1	1	1	1
2	2	2	2	2
2	2	1	2	1
2	1	2	2	2
2	2	1	2	2
2	2	1	2	2
2	2	1	2	1
2	1	2	2	2
2	2	1	2	1
1	1	2	1	1
1	2	2	2	2



The total time taken to complete both the parts is 45mins.

R5. PCA and Validation (25 points)

(a) (10 points) Implement Q5 from the book (under 5.4 exercises)

Answer

Code:

```
rm(list = ls())
library(ISLR)
attach(Default)

# a. Logistic regression model using "income" and "balance".
glm_fit = glm(default ~ income + balance, data = Default, family = "binomial")
print("Logistic Regression Model Summary using income and balance")
print(summary(glm_fit))

# b. Use Validation set approach and estimate the test error of this model.
validate_lrm = function(split_ratio) {
```

```
  print(paste0("Model validation for Default data set using split ratio = ", split_ratio))
  data_size = dim(Default)[1]
  train = sample(data_size, data_size * split_ratio)
  glm_fit = glm(default ~ income + balance, data = Default,
              family = "binomial", subset = train)
  prob = predict(glm_fit, newdata = Default[-train,], type = "response")
  predictions = rep("No", length(prob))
  predictions[prob > 0.5] = "Yes"
  conf_mat = table(predictions, Default[-train,]$default)
  print(conf_mat)
  res_err = (conf_mat[1, 2] + conf_mat[2,1]) / length(predictions)
  print(res_err)
}

# Assuming 50/50 to be the split for validation set approach.
validate_lrm(0.5)

# c. Repeating validations in b with different split ratios.
#    The ratios taken here are 90/10, 75/25, 60/40.
validate_lrm(0.9)
validate_lrm(0.75)
validate_lrm(0.6)

# d. Logistic Regression model for Default data set using income, balance and student.
#    Considering split ratio as 75/25
print("Model validation for Default data set with student, income and balance using split ratio =
0.75")
data_size = dim(Default)[1]
train = sample(data_size, data_size * 0.75)
glm_fit = glm(default ~ income + balance + student, data = Default,
            family = "binomial", subset = train)
prob = predict(glm_fit, newdata = Default[-train,], type = "response")
predictions = rep("No", length(prob))
predictions[prob > 0.5] = "Yes"
conf_mat = table(predictions, Default[-train,]$default)
print(conf_mat)
res_err = (conf_mat[1, 2] + conf_mat[2,1]) / length(predictions)
print(res_err)
```

Output:

(a) [1] "Logistic Regression Model Summary using income and balance"
glm(formula = default ~ income + balance, family = "binomial",
data = Default)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4725  -0.1444  -0.0574  -0.0211   3.7245

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.154e+01  4.348e-01 -26.545  < 2e-16 ***
income       2.081e-05  4.985e-06   4.174 2.99e-05 ***
balance      5.647e-03  2.274e-04  24.836  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1579.0  on 9997  degrees of freedom
AIC: 1585

Number of Fisher Scoring iterations: 8
```

(b) Use Validation set approach and estimate the test error of this model. Assuming split ratio = 50/50

```
[1] "Model validation for Default data set using split ratio = 0.5"
predictions   No   Yes
          No  4818  103
          Yes   23   56
[1] 0.0252  - error
```

(c) Repeating validations in b with different split ratios.

```
[1] "Model validation for Default data set using split ratio = 0.9"

predictions  No  Yes
          No   952  28
          Yes   4   16
[1] 0.032

[1] "Model validation for Default data set using split ratio = 0.75"

predictions   No   Yes
          No  2400   64
          Yes   8     28
[1] 0.0288
```

[1] "Model validation for Default data set using split ratio = 0.6"

```
predictions   No   Yes
        No  2403   58
        Yes    9   30
[1] 0.0245
```

As we see the error variation is random as the split ratio between the training/testing samples decreases. In general, the selection of split ratio depends on the sample data. The training sample should have all variations of all classes to be predicted and sufficient amount of test data to check for accuracy of the model. If more and more training data are selected and less test data then the model gets over-fitted and if less training data is selected then model gets under-fitted. Hence a balance is necessary and it depends on the samples available.

(d) Logistic Regression model for Default data set using income, balance and student. Considering split ratio as 75/25

```
predictions   No   Yes
        No  2420   50
        Yes    8   22
[1] 0.0232
```

Though in this case the value of error is reduced. This is just due to random selection of samples. If the code is run again with different random selection for training and testing data, we get different results. Thus adding student to the model doesn't seem to affect the model.

Total time taken by this question is approx. 200min

(b) (15 points) Implement PCA based Eigenface generation (see class slides). Use the data given in the faces-corrected.zip)
(Briefly describe your steps and include top 10 eigenface images; submit code separately).

Answer

Code:

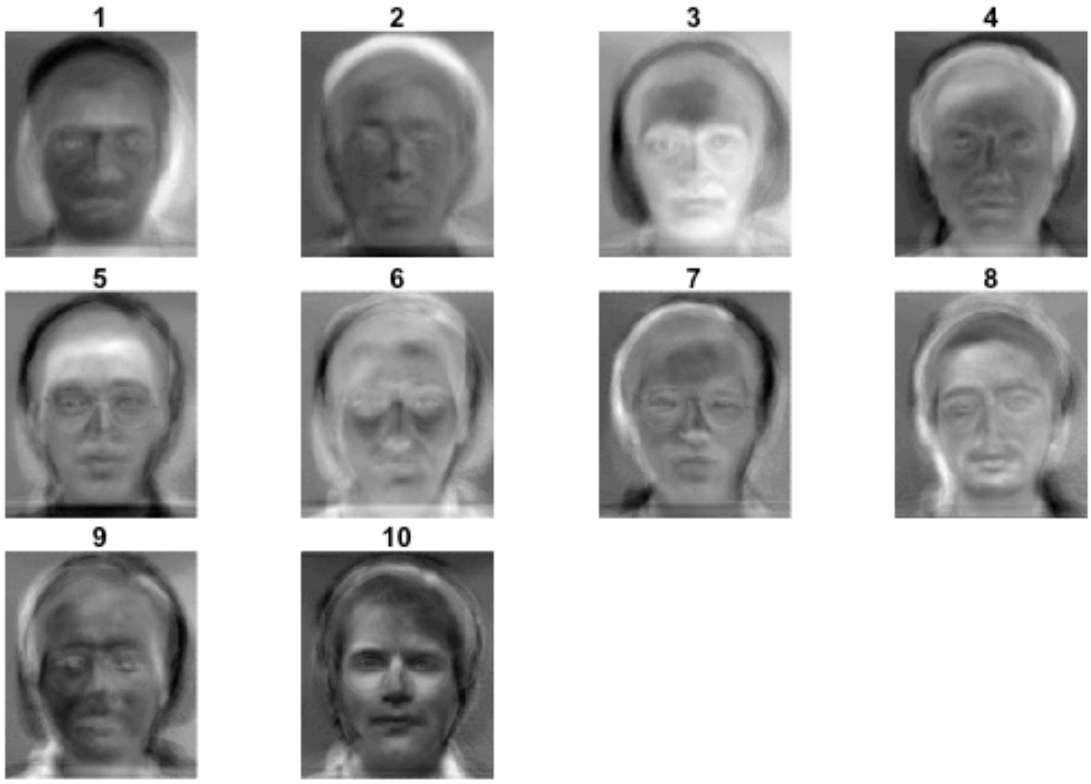
```
rm(list = ls())
library(pixmap)
library(ppls)
```

```
# Extracts grey pixels from each face
get_pixels <- function(x_i) {
  return(c(x_i@grey))
}

# Read all the face files
face_files = list.files(".")
faces = lapply(face_files, read.pnm)
faces = as.array(faces)

# Extract all the grey pixels from the faces
face_mat = lapply(faces, get_pixels)
face_mat = as.data.frame(face_mat)

# Calculate mean and normalize the faces
mean_mat = rowMeans(face_mat)
face_mat = sweep(face_mat, 1, mean_mat, "-")

# Compute covariance matrix and eigen vectors
cov_mat = cov(as.matrix(face_mat))
mat_eigen = eigen(cov_mat)
face_vectors = as.matrix(face_mat) %*% mat_eigen$vectors

# Print the top 10 images
op <- par(mfrow = c(3,4),
        oma = c(1,1,1,1) + 0.1,
        mar = c(0,0,1,1) + 0.1)
for(i in 1:10) {
  top_face = face_vectors[,i]
  top_face = normalize.vector(top_face)
  dim(top_face) = c(231,195)
  plot(pixmapGrey(top_face), main = i)
}
```

Following steps where followed to get the top 10 eigen faces:

1. Extract all the pixel values of each of the face in the data set.
2. Subtract mean pixel value from each pixel in all the faces.
3. Calculate the covariance matrix of for all the faces.
4. Calculate the eigen values and vectors for the generated covariance matrix.
5. The first 10 eigen vectors have the largest eigen values and hence are the top 10 faces eigen faces

Plot of the top 10 eigen faces.



Total time taken in solving this question was approx. 300 mins.

Bonus Questions (2% of grade)

B6. (**a**) (Chapter 5) Answer Q2 from the book under 5.4 exercises (1%). We will now derive the probability that a given observation is part of a bootstrap sample. Suppose that we obtain a bootstrap sample from a set of n observations.

1. What is the probability that the first bootstrap observation is not the jth observation from the original sample? Justify your answer.

Answer

The probability of selecting the jth element from the set of n observations as the first sample is 1/n, since each element is equally likely. Thus the probability of not selecting the jth observation from the original sample as the first bootstrap sample is thus **(1 – 1/n)**.

2. What is the probability that the second bootstrap observation is not the jth observation from the original sample? Justify your answer.

Answer

While bootstrapping each selection is independent and each selected element is replaced back into the original sample pool. Thus same element is equally likely to be selected the next time. Thus the probability of selecting jth element the second time is still 1/n irrespective of whether it was previously selected or not. Thus the probability of not selecting the jth element second time is **(1 – 1/n).**

3. Argue that the probability that the jth observation is not in the bootstrap sample is $(1 - 1/n)^n$.

Answer

As discussed in the previous questions, the probability of selecting the jth element at any time is 1/n. Thus not selecting jth element for any of the individual selections is (1 – 1/n). Now, these selections are done n times and each time jth element is not selected. Thus the overall probability of not selecting the jth element in (1 – 1/n) * (1 – 1/n) * … * (1 – 1/n), n times. Which is $(1 - 1/n)^n$.

4. When n = 5, what is the probability that the jth observation is in the bootstrap sample?

Answer

As discussed in the previous question, probability of not selecting the jth element at all is $(1 - 1/5)^5$. So the probability of jth observation being selected in the bootstrap sample is $(1 - (1 - 1/5)^5)$, which is **0.67232**.

5. When n = 100, what is the probability that the jth observation is in the bootstrap sample?

Answer

Similarly the probability that jth observation is in the bootstrap sample is $(1 - (1 - 1/100)^{100}) =$ **0.63397**

6. When n = 10,000, what is the probability that the jth observation is in the bootstrap sample?

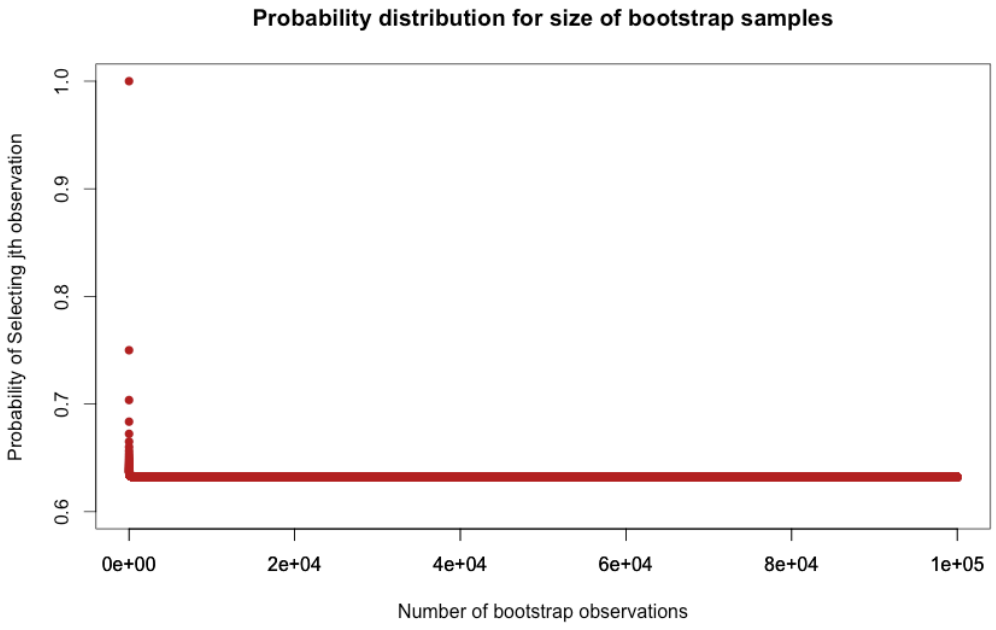
Answer

Similarly the probability that jth observation is in the bootstrap sample is $(1 - (1 – 1/10000)^{10000})$ = **0.63213**

7. Create a plot that displays, for each integer value of n from 1 to 100,000, the probability that the jth observation is in the bootstrap sample. Comment on what you observe.

Answer

The plot for the probability is as follows:



From the plot above we can observe that as the value of n increases the probability value asymptotically reaches a value close to 0.632. Suggesting that approx. 63% of the original sample in a part of the bootstrap.

8. We will now investigate numerically the probability that a bootstrap sample of size n = 100 contains the jth observation. Here j = 4. We repeatedly create bootstrap samples, and each time we record whether or not the fourth observation is contained in the bootstrap sample. Comment on the results obtained.

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

Answer

The output for the above code is **0.6367.** This is similar to the previous conclusion. Here the probability for the jth observation in this case 4^{th} observation to be present in bootstrap sample on an average is **0.6367**. This results shows that as you repeat the bootstrapping process more and more times the probability of an element being present in bootstrap sample tends to be close to 0.632 on an average.

The total time taken to solve this question was approx. 110mins.

B6. (**b**) (Chapter 6) Answer Q8 (only parts (a) to (c) from the book under section 6.8 exercises (1%). In this exercise we will generate simulated data, and will then use this data to perform best subset selection.

1. Use the rnorm() function to generate a predictor X of length n = 100, as well as a noise vector of length n = 100.
2. Generate a response vector Y of length n = 100 according to the model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$$

   where β_0, β_1, β_2 and β_3 are constants of your choice.

3. Use the regsubsets() function to perform best subset selection in order to choose the best model containing the predictors X, X^2, …, X^{10}. What is the best model obtained according to C_p, BIC, and adjusted R^2? Show some plots to provide evidence for your answer and report the coefficients of the best model obtained. Note you will need to use data.frame() function to create a single data set containing both X and Y.

Answer

The complete code for this question is as follows:

```
#Question B6 part (B)
# Question 1: Generating values of x and eps using rnorm
# setting seed to get the same random values for rnorm
set.seed(2)
x = rnorm(100)
eps = rnorm(100)

# Question 2: Generating response vector Y
# Coefficient values are taken by choice
```

b0 = 10
b1 = 3
b2 = -2.25
b3 = 0.05
*y = b0 + b1 * x + b2 * x^2 + b3 * x^3 + eps*

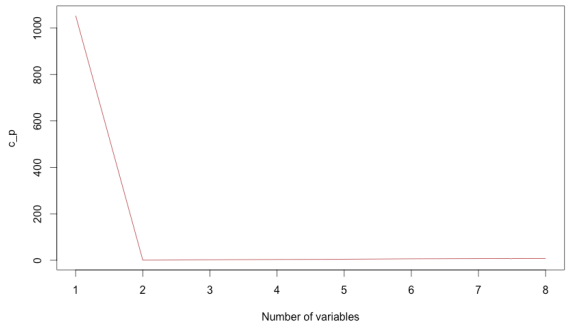
```
# Question 3: Determining the best model along with some plots as evidence.
library(leaps)
sample_data = data.frame(y, x)
model = regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6) + I(x^7) + I(x^8) + I(x^9) +
I(x^10), data = sample_data)
model_summary = summary(model)
plot(model_summary$cp, type = "l", col = "firebrick", xlab = "Number of variables", ylab = "c_p")
readline()
plot(model_summary$bic, type = "l", col = "steelblue", xlab = "Number of variables", ylab =
"bic")
readline()
plot(model_summary$adjr2, type = "l", xlab = "Number of variables", ylab = "Adjusted R^2")

# Finding the coefficents of the best fitted model.
print(coef(model, which.max(model_summary$adjr2)))
```

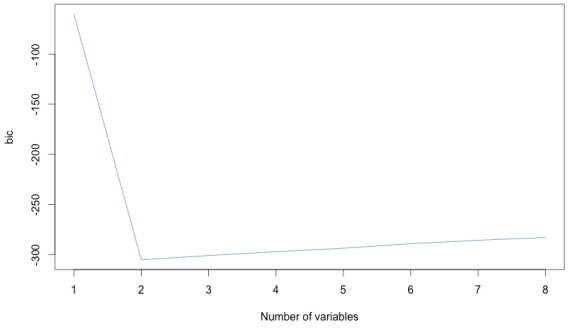
Output:

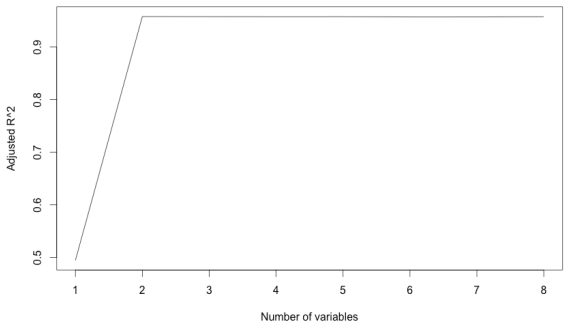
Coefficient for the best model fit is as follows:

(Intercept)	x	I(x^2)
9.979393	3.099172	-2.208522

The following plots show some evidence for the best model.







In the above plots we can see that we must pick 2 variable model for each of C_p, BIC and Adjusted R^2 and the output summary above shows the coefficients of the best model fit.

Total time taken to solve this question was approx. 135mins