Assignment 4

Danny Kearns

2023-11-29

Part 1 can be completed on paper or using LaTeX on an RMarkdown document. The rest of the parts should be turned in as a knitted RMarkdown (.html, .pdf, .docx)

1. Consider the table below with the samples as the columns and the variables as the rows:

Variable	Sample1	Sample2	Sample3	Sample4	Sample5
X_1	2	3	5	6	10
X_2	3	4	6	7	11

1a. First start by calculating the covariance matrix (denoted as Σ). Set up the equation, but use `R` to calculate the matrix.

1b. Find the eigenvalues by calculating $\det(\Sigma - \lambda I)$ and solving for λ

1c. Compute the eigenvectors corresponding to each eigenvalue using the properties of eigenvalues and eigenvectors:

$$\Sigma e_i = \lambda e_i$$

Where e_i is the ith eigenvector. Then convert e_i to a unit eigenvector using the formula:

$$\frac{e_i}{||e_i||}$$

1d. Finally compute the percentage of contribution from each eigenvector by summing the eigenvalues together and dividing each individual eigenvalue by the total. Draw a Scree plot to show the differences in contribution from the principal components

2. Copy the starter code below. This is your dataset (Make sure you have `dplyr` installed)

```
library(dplyr)
iris_test = iris[order(iris$Species),]
rownames(iris_test) = paste(substr(iris_test[,5],1,2),
                            sample(1:2000, 150, replace = FALSE),sep = "")

set.seed(641)
iris_test = sample_n(iris_test, 20)
```

With `R`, create a PCA plot with PC1 on the X-axis and PC2 on the Y-axis. Construct a Scree plot.

3. Using the dataset attached to the assignment, `honey.csv`, first split the data as 60% training data and 40% validation data, using the `TVHSplit` function created in the notes (keep `iseed` as the default value so that the sampling is consistent). Train 3 models using `average_price` as your y variable:

1. A decision tree using the `rpart` package
2. A model based recursive partitioning tree model from the `partykit` package
3. A boosted tree from the `gbm` package

Compare the validation R^2 for each model