

CS400 Machine Learning - Exam 1 (Take home)

Instructions

1. All data sets are available in Blackboard -> Data files folder.
2. Create a separate python file for each section.
3. Comment the files appropriately.
4. Zip and Submit all python files, other files (if needed based on the question) in Blackboard.

Section 1

1. Using the Numpy package, complete the following questions.
 - a. Create an array with values ranging from 5 to 25. Print the array created.
 - b. Reverse a vector (first element becomes last). Print the reversed array.
 - c. Create a random array of size 30 with float values and find the mean value. Print the array and the mean value.
2. Using the Pandas package, write a program to create a Pivot table and find number of adult males, adult females and children using the following data – titanic.csv. You will have to research online for this question on how to create a pivot table.

This section will only require submission of a python script.

Section 2

1. Create a pie chart for the following data – iris.csv. The pie chart should show the distribution of the different species in the data. Hint, the number of species is uniform across the data, meaning all species occur the same number of times in the table. Your code should count how many times each species occurs and then use those values to create a pie chart. The graph should be saved in a PNG file. Submit the .py and .png for this question.
2. Draw a line chart and a bar chart for the following data – flights.csv. X axis is the year, Y axis the number of flights. Discuss which graph shows you better results. Include all elements of a complete graph. Save the line chart and the bar chart in separate PNG files. Include your discussion in the Python file using comments. This question requires submission of a .py and 2 .png files.

Section 3

Apply the regression models to the following data - Employee_Compensation_SF.csv.

Questions to answer:

1. Are “Health/Dental” benefits dependent upon “Salaries”, “Overtime” and “Retirement”?
2. If the answer is yes, then which type of regression best fits this analysis – Linear, Multiple or Polynomial?
3. This question requires submission of a python script that contains the implementation of the model that best fits the data.
4. Include one graph using linear regression with a discussion on what that graph represents.
5. Include one graph using polynomial regression with a discussion on what that graph represents.

Section 4

Apply the regression models to the following data – Social_Network_Ads.csv.

Questions to answer:

1. Are “EstimatedSalary” benefits dependent upon “Gender”, “Age”?
2. If the answer is yes, then which type of regression best fits this analysis – Linear, Multiple or Polynomial?
3. This question requires submission of a python script that contains the implementation of the model that best fits the data. Include one graph with a discussion on what that graph represents.
4. Include one graph using linear regression with a discussion on what that graph represents.
5. Include one graph using polynomial regression with a discussion on what that graph represents.