Prashanth Kumaran
ID: 4549673

Homework 1 Report

1. Conceptual question: propose one new regression problem (one that we have not discussed in class) that you can solve with machine learning, and describe how you would go about solving it. Please give a short answer (text) to the following;
   Predicting the number of fantasy football points a given player will score in the upcoming season
   a. What features (x) would you use?
      Past Points, Age, Receptions, Touchdowns, Yards, Schedule Difficulty Rating
   b. What would the labels (y) be?
      Number of Points Scored in the current year
   c. How would you collect data?
      All of the data for games played is published online, including fantasy points scored
   d. Why might the problem turn out to be challenging?
      Variability with player position and team.

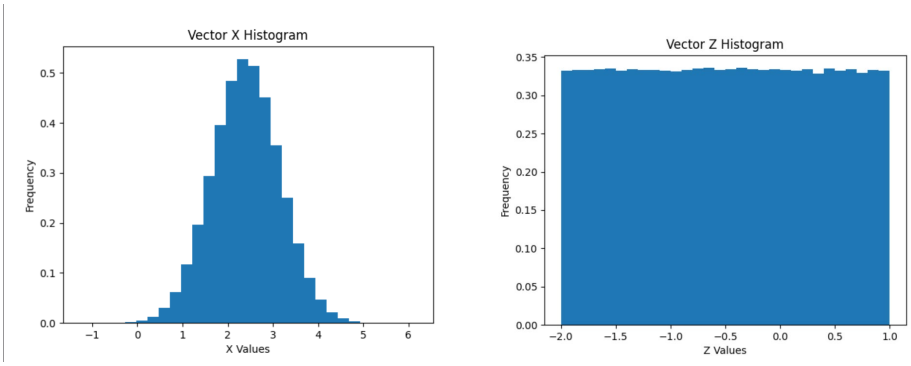
2. Repeat question 1 for a classification problem
   Predict which NHL playoff team will win the Stanley Cup based on regular season performance.
   a. What features (x) would you use?
      Wins, Points, Goals Scored, Goal Differential, Goalie Save Percentage
   b. What would the labels (y) be?
      Each of the 16 teams that could win the Stanley Cup
   c. How would you collect data?
      All of the data is published online
   d. Why might the problem turn out to be challenging?

      Rule changes from year to year, team performance changes from regular season to playoffs, only 98 previous data points to look at.

3. Basic Operations

   c. The histogram for x exhibits the bell curve expected in a Gaussian distribution. The histogram for z shows the flat top of a uniform distribution

Prashanth Kumaran
ID: 4549673



ps1-3-c-1.png *ps1-3-c-2.png*

d. Loop execution time: 0.1563 seconds

e. Optimized execution time: 0.0020 seconds. The built in function (np.add()) is more efficient to add a constant to a long vector than using a loop.

f. Trial 1: 267223 Elements, Trial 2: 266090 Elements. The difference between each trial is due to the finite sample size and the randomness of the uniform distribution.

4. Linear Algebra

b. Resultant = [5.72727273, -0.27272727, -0.95454545]

c. x1 = [-4, 0, 1] L1 Norm = |-4| + |0| + |1| = 5 L2 Norm = $\sqrt{(-4)^2 + 0^2 + 1^2} = \sqrt{17} = 4.1231$

x2 = [-2, -2, 0] L1 Norm = |-2| + |-2| + |0| = 4 L2 Norm = $\sqrt{(-2)^2 + (-2)^2 + 0^2} = \sqrt{8} = 2.828$

numpy outputs:

x1 L1 norm: 5.0

x1 L2 norm: 4.123105625617661

x2 L1 norm: 4.0

x2 L2 norm: 2.8284271247461903

Prashanth Kumaran
ID: 4549673

5. Splitting Data
    a.

```
116  # ////////////////////// Question 5 //////////////////////
117  print("---------------Question 5 Text Outputs----------------")
118
119  X = np.zeros((10, 3))
120
121  for i in range(10):
122      X[i][:] = i
123
124  y = np.arange(1, 11)
125
126  print("Matrix X:\n", X)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  ESP-IDF

x2 L2 norm: 2.8284271247461903
---------------Question 5 Text Outputs----------------
Matrix X:
 [[0. 0. 0.]
 [1. 1. 1.]
 [2. 2. 2.]
 [3. 3. 3.]
 [4. 4. 4.]
 [5. 5. 5.]
 [6. 6. 6.]
 [7. 7. 7.]
 [8. 8. 8.]
 [9. 9. 9.]]
```

d. Trial 1:

```
X_train:
 [[7 7 7]
 [2 2 2]
 [4 4 4]
 [0 0 0]
 [3 3 3]
 [9 9 9]
 [8 8 8]
 [5 5 5]]
X_test:
 [[1 1 1]
 [6 6 6]]
y_train:
 [7 2 4 0 3 9 8 5]
y_test:
 [1 6]
```

Trial 2:

```
X_train:
 [[4 4 4]
 [6 6 6]
 [0 0 0]
 [1 1 1]
 [3 3 3]
 [5 5 5]
 [9 9 9]
 [2 2 2]]
X_test:
 [[7 7 7]
 [8 8 8]]
y_train:
 [4 6 0 1 3 5 9 2]
y_test:
 [7 8]
```

Trial 3:

```
X_train:
 [[8 8 8]
 [1 1 1]
 [2 2 2]
 [3 3 3]
 [7 7 7]
 [6 6 6]
 [5 5 5]
 [4 4 4]]
X_test:
 [[9 9 9]
 [0 0 0]]
y_train:
 [8 1 2 3 7 6 5 4]
y_test:
 [9 0]
```

Prashanth Kumaran
ID: 4549673

The same submatrices are not produced every time X is split because of the random shuffling of the train indices selected.