Homework 6

Instructions
This homework contains **4** concepts and **7** programming questions. In MS word or a similar text editor, write down the problem number and your answer for each problem. Combine all answers for concept questions in a single PDF file. Export/print the Jupyter notebook as a PDF file including the code you implemented and the outputs of the program. Make sure all plots and outputs are visible in the PDF.

Combine all answers into a single PDF named andrewID_hw6.pdf and submit it to Gradescope before the due date. Refer to the syllabus for late homework policy. Please assign each question a page by using the "Assign Questions and Pages" feature in Gradescope.

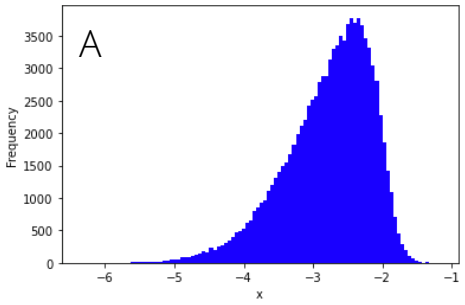
Here is a breakdown of the points for programming questions:

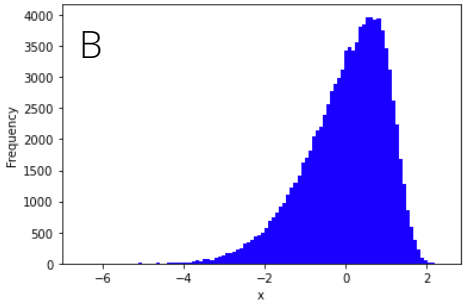
Name	Points
M6_HW1	30
M6_HW2	30
M6_L1_P1	6
M6_L1_P2	6
M6_L1_P3	6
M6_L2_P1	6
M6_L2_P2	6

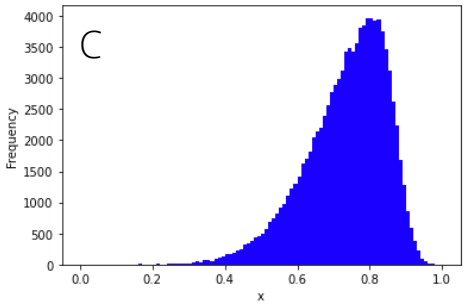
Problem 1 (2.5 points)

Multiple choice (select one)
Given the original data in A, the data in B and C appears to be:

1. B) Normalized and C) Standardized
2. B) Standardized and C) Normalized
3. B) and C) both unchanged from the original data







Problem 2 (2.5 Points)

Multiple Choice (select one)
Which scaling technique would be best to use on the following data:
X = [0.002, 0.01, 100000, 4000, 500, 0.00008, 7]

1. Normalization
2. Standardization
3. Log Transformation

Problem 3 (2.5 Points)

Compute the Pearson's correlation coefficient for the following two features by hand:
$x_1 = [8,4,0,-4]$, $x_2 = [-16,-12,-10,2]$

Problem 4 (2.5 Points)

Multiple choice (select one)
Consider the dataset with features x_1, x_2, x_3, and label y. We have generated the following correlation matrix, and would like to select a feature to remove. We have set the the following threshold $|r| > 0.9$ to drop features. Which of the features should be dropped?

1. x_1
2. x_2
3. x_3