Homework # 1

August 10, 2019

1 Linear Algebra in Numpy

(1) Create a random 100-by-100 matrix M, using numpy method "np.random.randn(100, 100)", where each element is drawn from a random normal distribution.

(2) Calculate the mean of all the elements in M;

(3) Use "for loop" to calculate the mean of each row of M.

(4) Use matrix operation instead of "for loop" to calculate the mean of each row of M, simultaneously, hint (create another matrix np.ones(100, 1))?

(5) Calculate the inverse matrix of M: M^{-1}

(6) Verify that $M^{-1}M = MM^{-1} = I$, are the off-diagnoal element exactly 0?

2 Probability Distribution

You have recently joined a data science team and working on a project that needs to simulate 5 types of distributions (Bernoulli, Poisson, Gaussian, uniform distribution and Rolling-Dice distribution). Your teammates provides you with a simulated data sample "**sample_trials.csv**". In the file, each column contains 5000 data sample draw from one of the 5 distributions. However, your teammate did not label them properly. Since your teammate is on vacation, you decide to figure them out by yourself.

(1) Do the columns have discrete value or continuous value? How many unique values does each column have? What are the min, max values of the columns

(2) As a first step, you investigate the distribution probability of each column by plot their histograms.

(3) Use your educated guess and label the column properly using the distribution name.

(4) Calculate the mean and variance of each column. Describe the distributions and their key parameters. Write down the distribution formula if possible.