

Homework # 1

September 8, 2020

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 and variance of all the elements in M ;
- (3) Use "for loop" to calculate the mean and variance of each row of M .
- (4) Use matrix operation instead of "for loop" to calculate the mean of each row of M , hint: create a vector of ones using `np.ones(100, 1)`?
- (5) Calculate the inverse matrix M^{-1}
- (6) Verify that $M^{-1}M = MM^{-1} = I$. Are the off-diagonal elements exactly 0, why?

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 and Rolling-Dice distribution). Your teammate provides you with a simulated data sample "**sample_trials.csv**". In the file, each column contains 5000 numbers drawn from one of the 5 distributions. However, the columns are not labeled properly and you have to figure out the labels yourself as your teammate is on vacation.

- (1) Do the columns have discrete value or continuous value? How many unique values does each column have?
- (2) What are the min, max, mean, variance of the columns?
- (3) Investigate the distribution of each column by plotting the histograms. Make sure you choose the bin size properly.
- (4) Based on the analysis above, label each column with the distribution name and explain why.
- (5) Knowing the mean, variance, write down the formulas of the distributions.
- (6) Simulate another 5000 samples of Bernoulli distribution with the same set of parameters. Write it into a text file.