Aashsih Kumar pcr902

November 26, 2019

Contents

1 Make Your Own

1. sample space,X : grades in prerequisite courses, overall GPA, study habits
2. Y: grades A,B,C,D,E
3.
4.
5.

2 Illustration of Markov's and Chebychev's Inequalities

1. I perform an experiment of choosing 20 random variables (with 0 for failure and 1 for success) 1,000,000 times. Then, calculate $\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha$ for $\alpha \in (0.5, 0.55, ..., 0.1)$ for each experiment. Then, calculate the empirical frequency for each α. See Part2.py file in `code.zip`
2. Plot for Empirical Frequency, Markov bound, chebychev bound

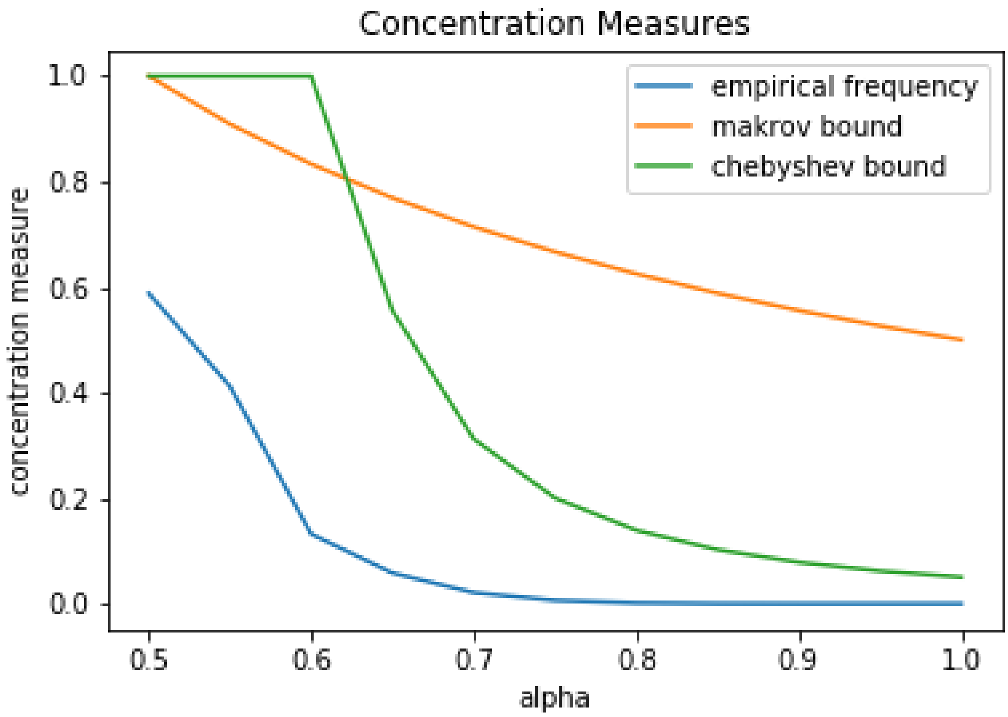

Figure 1: Empirical Frequency, Markov bound, chebychev bound

3. The above granularity of α is sufficient because X_i is a discreet random variable, it can be either 0 or 1. So, $\frac{1}{20}\sum_{i=1}^{20} X_i$ will belong to $(0, 0.05, ..., 0.1)$. so, for instance, values greater than 0.51 and 0.52 will be same and not provide any extra information.

4. X_i can be 0 or 1 and probability of success and failure is equal i.e. 0.5
   Expectation E[X] = 0.5
   According to the markrovs inequality

$$P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha) \leq \frac{1}{20} * \frac{\sum_{i=1}^{20} E[X_i]}{\alpha} \tag{1}$$

$$\leq \frac{0.5}{\alpha} \tag{2}$$

   Hence, Markrov Bound is $\frac{0.5}{\alpha}$
   Refer to above fig 1 for plot

5. let $Y = \frac{1}{20}\sum_{i=1}^{20} X_i$ be the random variable

$$E[Y] = E[\frac{1}{20}\sum_{i=1}^{20} X_i] \tag{3}$$

$$= 0.5 \tag{4}$$

here, $E[X^2] = E[X]$ as $X = X^2 \in 0, 1$

$$Var[\frac{1}{20}\sum_{i=1}^{20} X_i] = \frac{1}{20^2}\sum_{i=1}^{20} Var[X_i] \tag{5}$$
$$= \frac{1}{20^2}\sum_{i=1}^{20}(E[X_i^2] - (E[X_i])^2) \tag{6}$$
$$= 0.0125 \tag{7}$$

then, according to the chebychev inequality

$$P(|Y - E[Y]| \geq \epsilon) \leq \frac{Var[Y]}{\epsilon^2} \tag{8}$$
$$P(|Y - 0.5| \geq \epsilon) \leq \frac{0.0125}{\epsilon^2} \tag{9}$$
$$P(Y \geq \epsilon + 0.5) \leq \frac{0.0125}{\epsilon^2} \tag{10}$$

By comparing $P(Y \geq \alpha)$ with equation 10, $\alpha = \epsilon + 0.5$
$\epsilon = \alpha - 0.5$
So, $P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha) \leq \frac{0.0125}{(\alpha-0.5)^2}$
Hence, chebychev Bound is $\frac{0.0125}{(\alpha-0.5)^2}$
Refer to above fig 1 for plot

6. comparison between the plots In the above plot, the chebychev bound is tight bound whereas markrov bound is comparatively looose bound. Empirical frequency, markrov bound and chebychev bound all of them are highest at $\alpha = 0.5$ and decreases rapidly when α increases.

7. Probability for $\alpha = 0.95, 1.0$

$$P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq 0.95) = \binom{20}{19} * (\frac{1}{2})^{19} * (\frac{1}{2}) + \binom{20}{20} * (\frac{1}{2})^{20} \tag{11}$$
$$= 20 * (\frac{1}{20})^{20} + (\frac{1}{20})^{20} \tag{12}$$
$$= 21 * (\frac{1}{20})^{20} \tag{13}$$

$$P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq 0.1) = \binom{20}{20} * (\frac{1}{2})^{20} \tag{14}$$
$$= (\frac{1}{20})^{20} \tag{15}$$
$$\tag{16}$$

3 Tightness of Markov's Inequality

4 Digits Classification with Nearest Neighbours

5 Nearest Neighbours for Multiclass Classification

Algorithm K Nearest Neighbors (K-NN) for Multiclass Classification with Y = [0,1,2,.....,9]

- Input: A set of labeled points $(x1, y1), ..., (xn, yn)$ and a target point x that has to be classified.
- Calculate the distances $d_i = d(xi, x)$.
- sort the distances d_i in ascending order
- calculate the frequency of all the labels y corresponding to the first K d_i that are closest to x. The label with highest frequency will be the output of KNN. In case of tie when more than one label has the highest frequency, then compute the average of d_i corresponding to those labels in first K points. and the label with least average distance will be the output.

6 Linear Regression

1. $$Y = [y_1, ..., y_n]^T \tag{17}$$
   $$X = \begin{bmatrix} x_1 & 1 \\ . & . \\ . & . \\ x_n & 1 \end{bmatrix} \tag{18}$$
   $$(w, b)^T = (X^T X) X^T Y \tag{19}$$

   See Part6.py file in `code.zip`

2. $h(x) = w^T x + b$
   Two parameters of the model
   Weight, w = 9.489
   Bias, b = -10.426
   Mean Square error = 0.012

3. Linear Regression

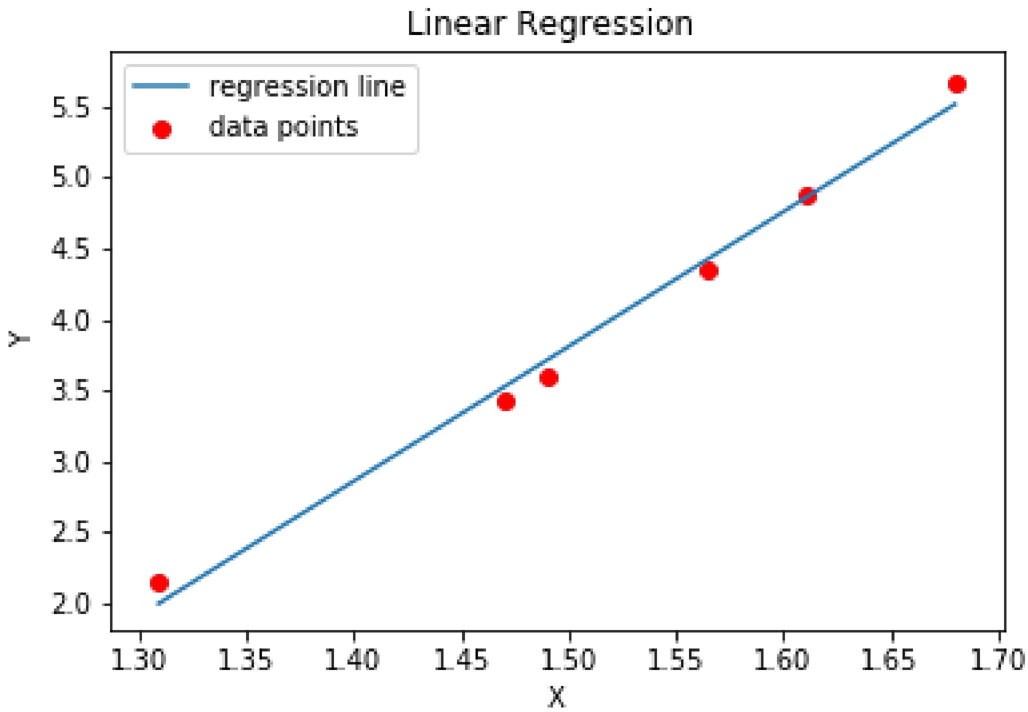

Figure 2: Plot of data and regression line

4. Variance = 1.267
   Mean Square Error = 0.012
   Quotient of Variance and Mean Square Error = 0.0097

5. Transformed Linear Regression
   Mean square Error = 0.00049

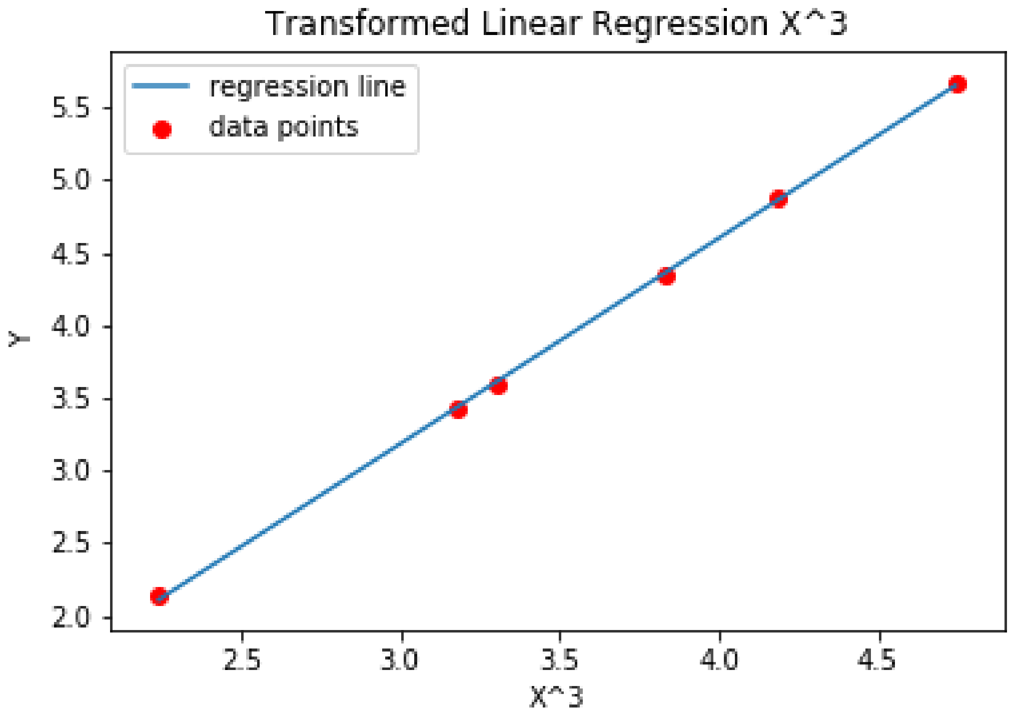

Figure 3: Plot of transformed data and regression line