SMAI (CSE 471)

Spring-2019

Assignment-2

Name: Karnati Venkata Kartheek

Roll Number 2018801010

**Question1**

**Part1(Q1.1)**

Results on validation data for Robot1 Dataset

Number of Neighbors = 12, distance measure = Euclidean

**-------------------------------------------------**

**Precision is 0.9230769230769231**

**Recall is 1.0**

**F1\_Score is 4.166666666666666**

**Accuracy is 0.96**

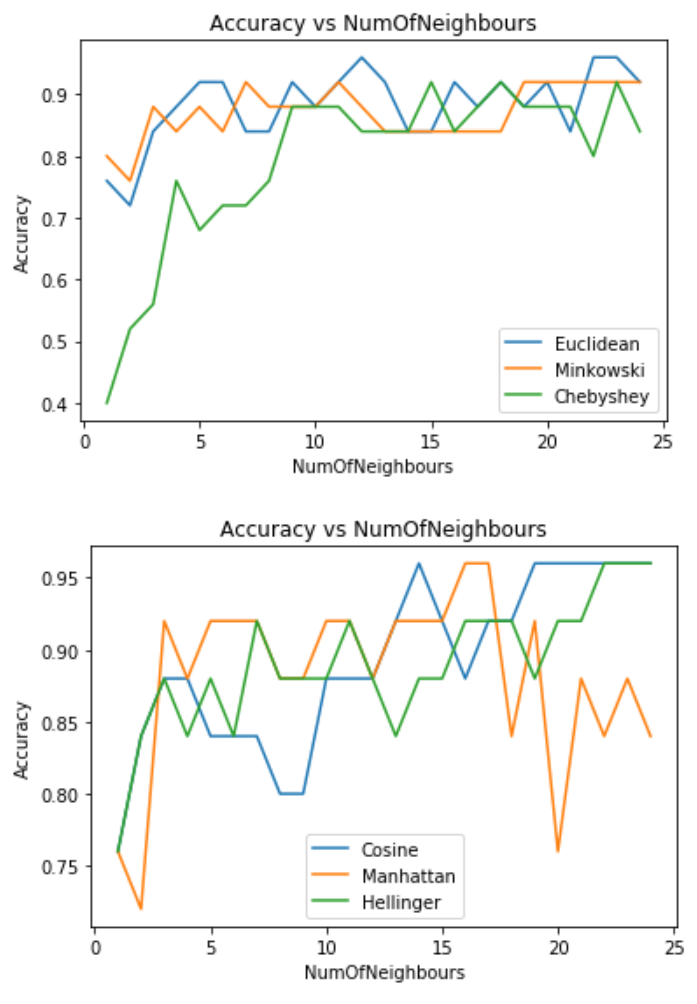
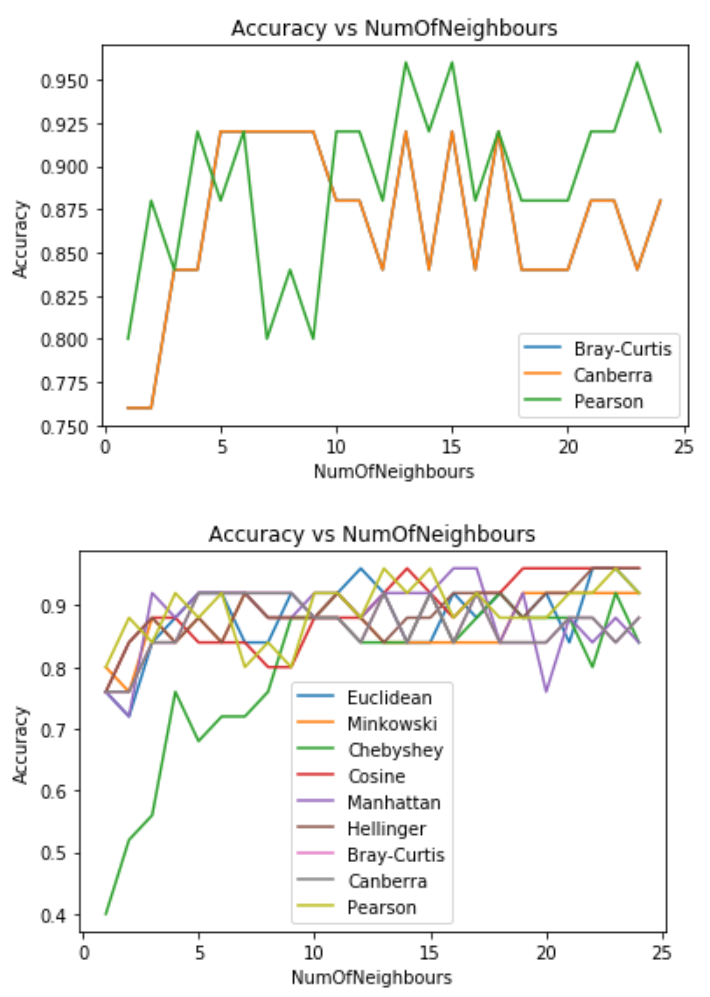
**-------------------------------------------------**

Results on validation data for Robot2 Dataset

Results on validation data for Iris Dataset

**Part1(Q1.2)**

Number of Neighbors vs Accuracy for Robot1 Dataset

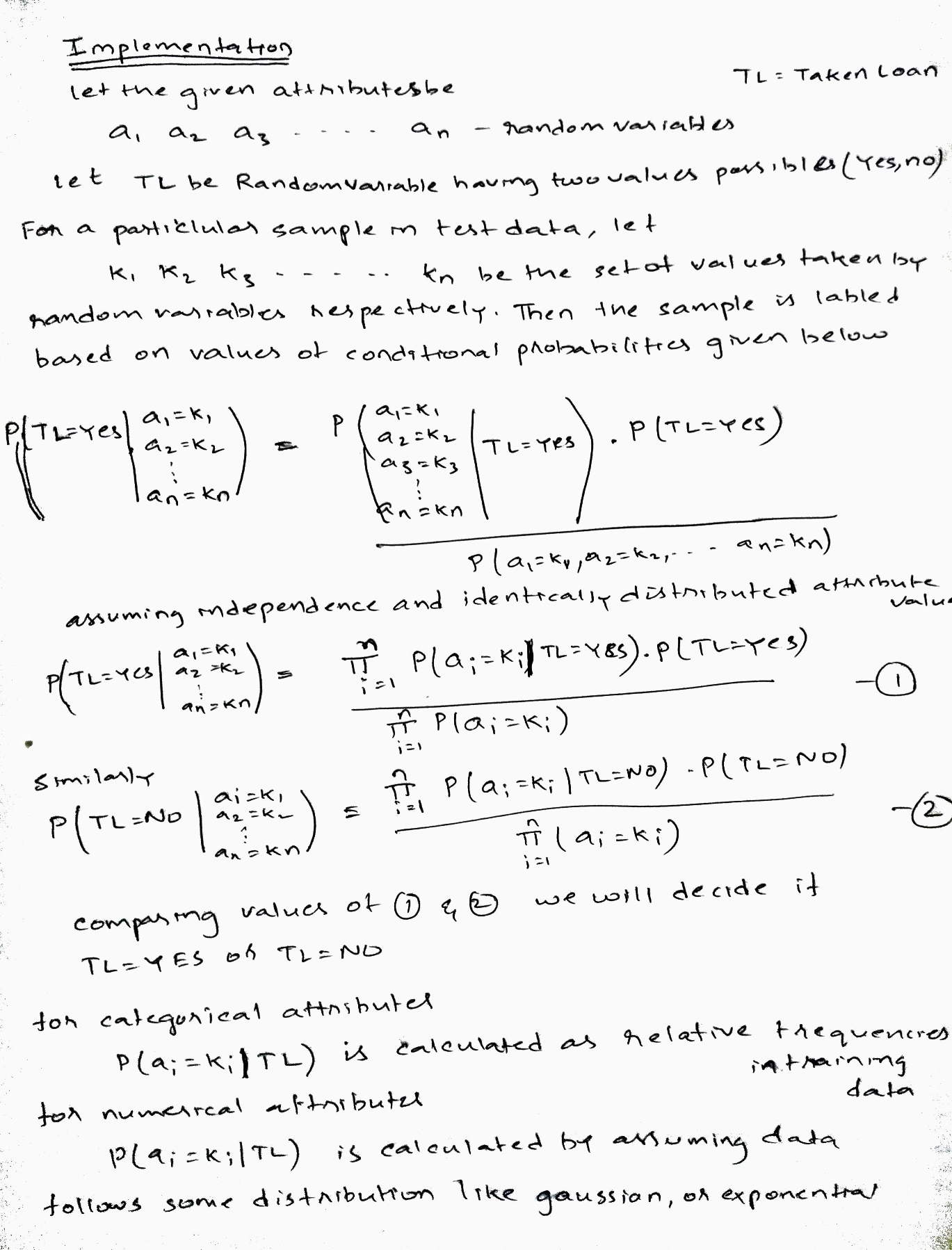
 

Number of Neighbors vs Accuracy for Robot2 Dataset

Number of Neighbors vs Accuracy for Iris Dataset

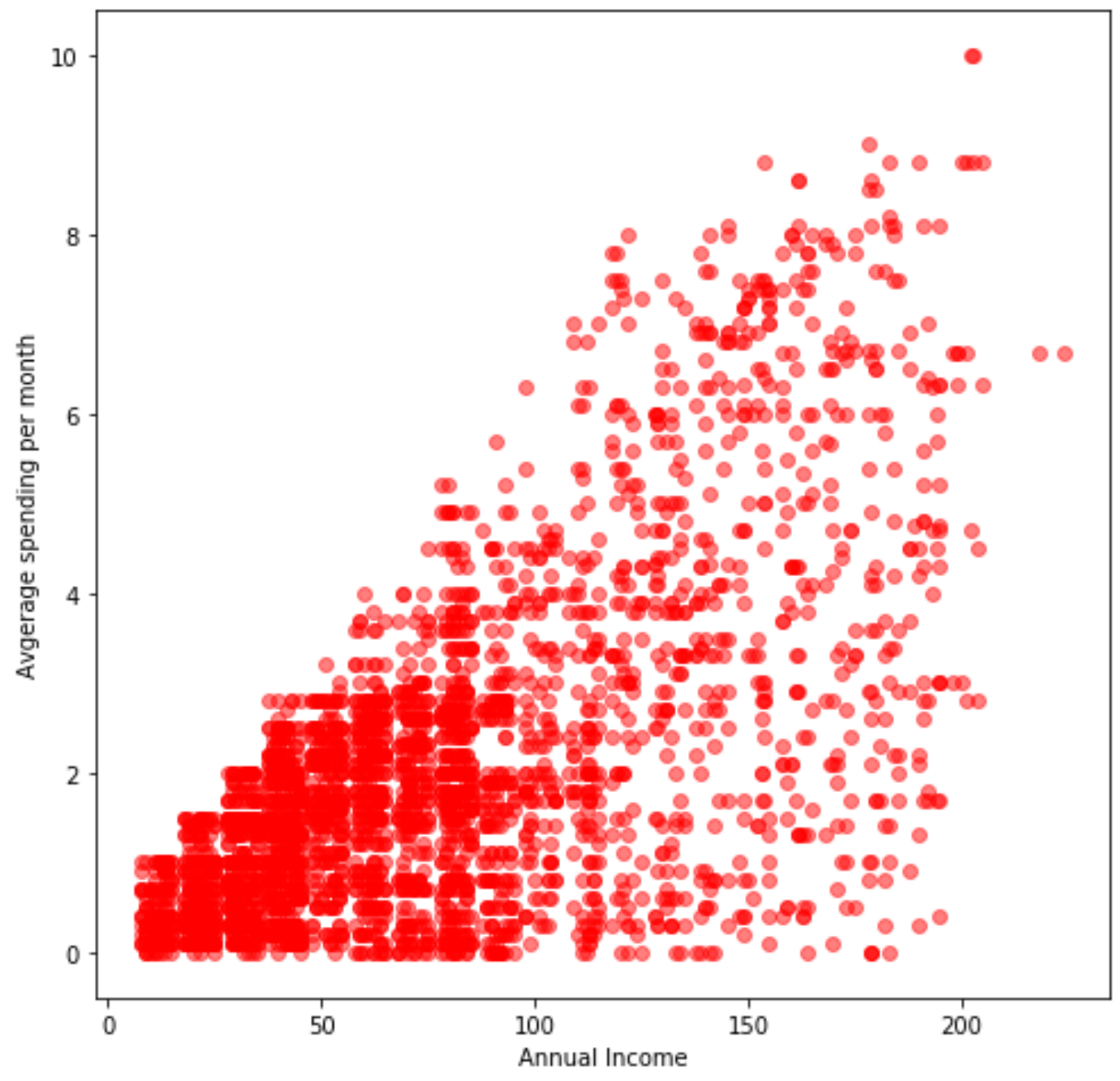
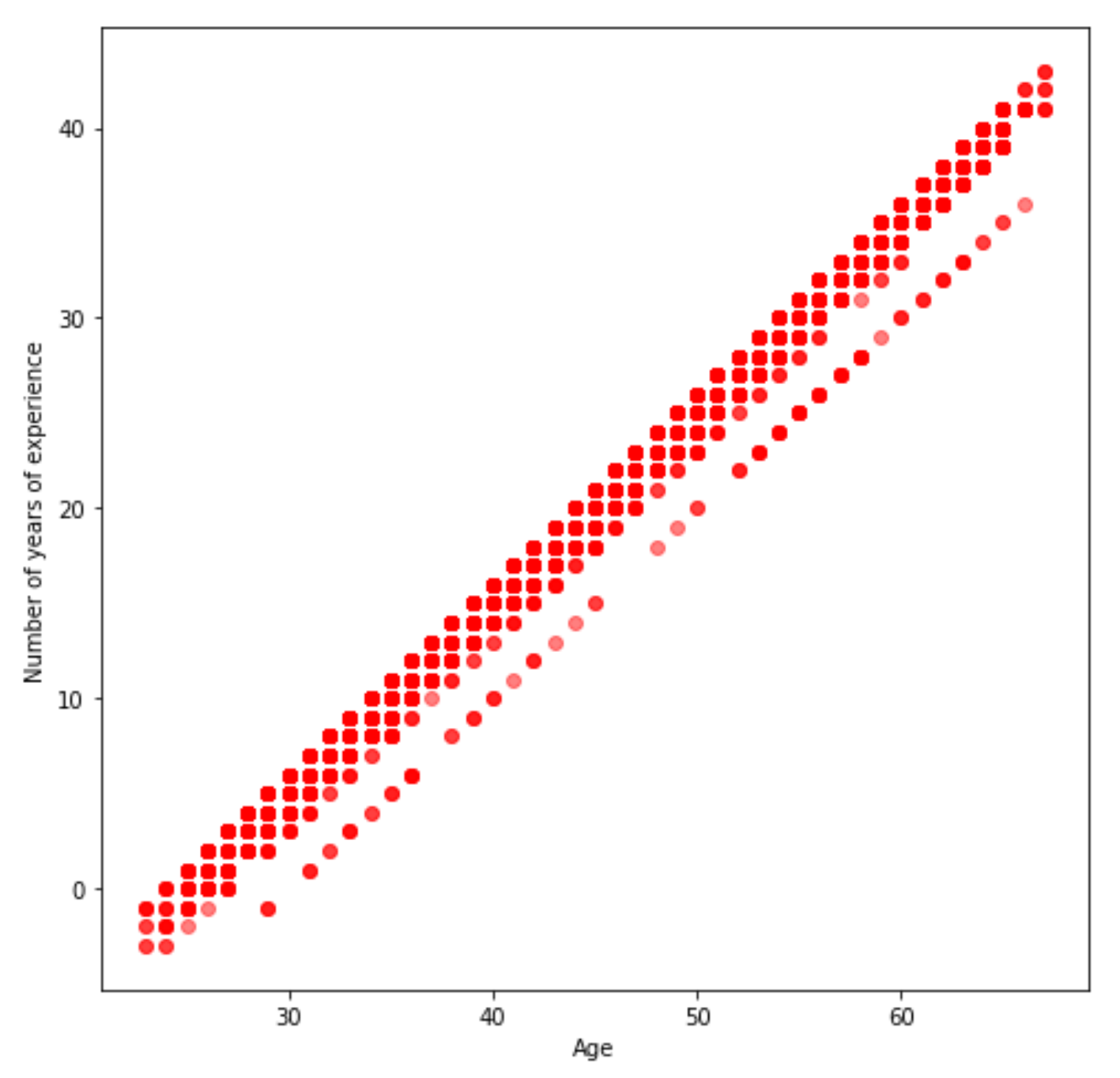
**Question2**

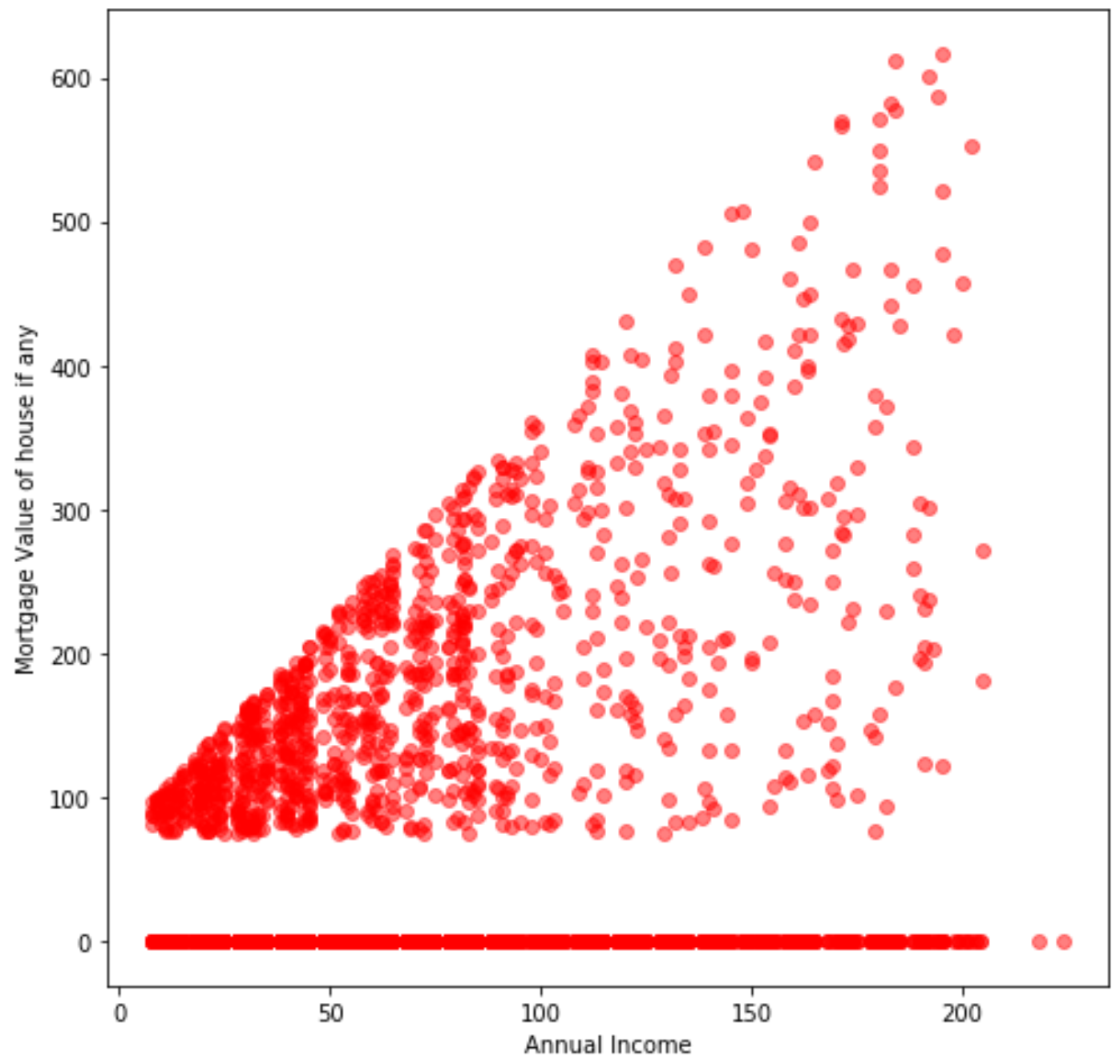
**Implementation**

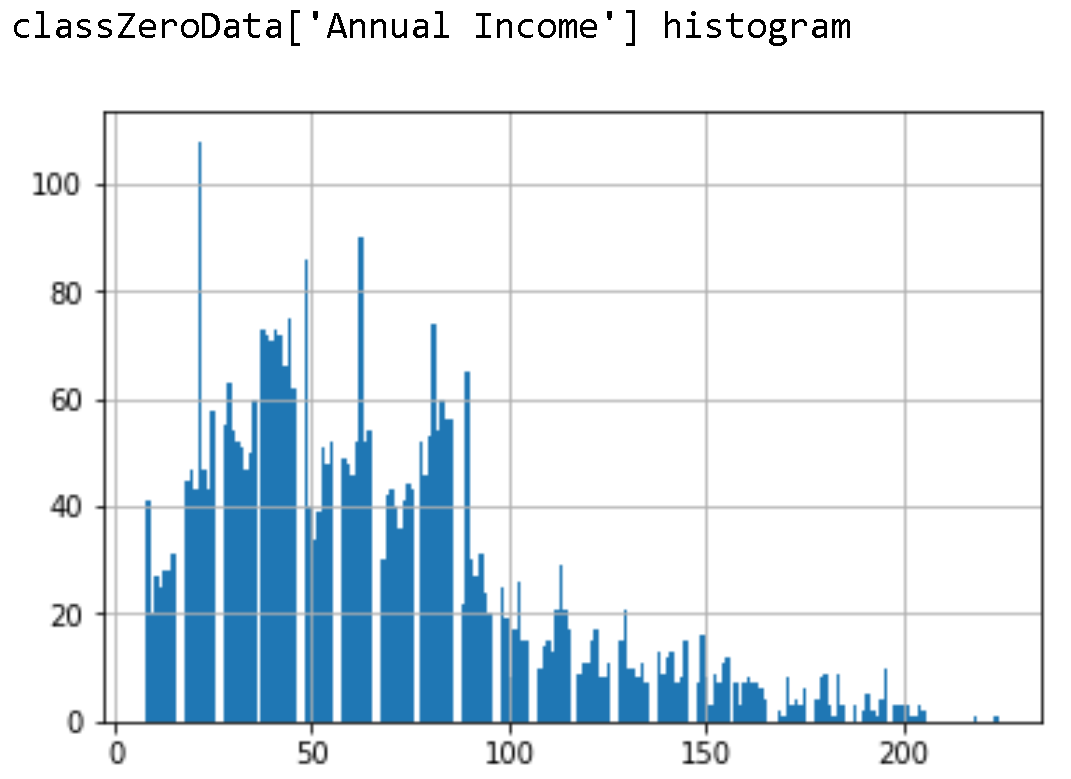
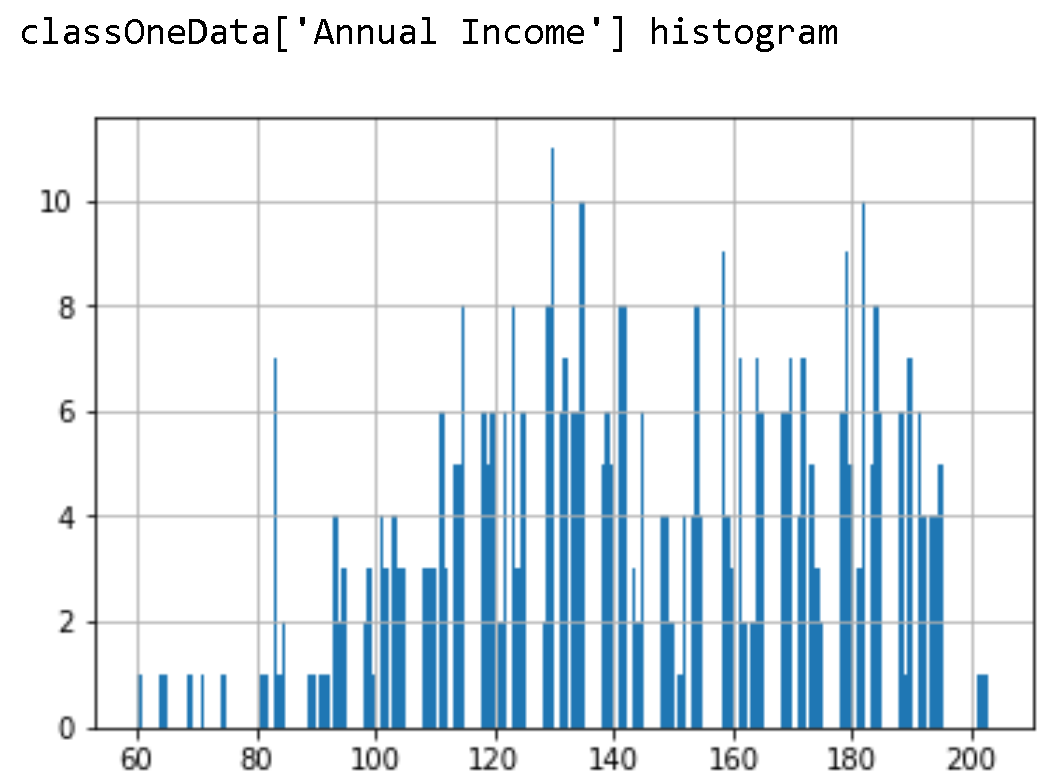


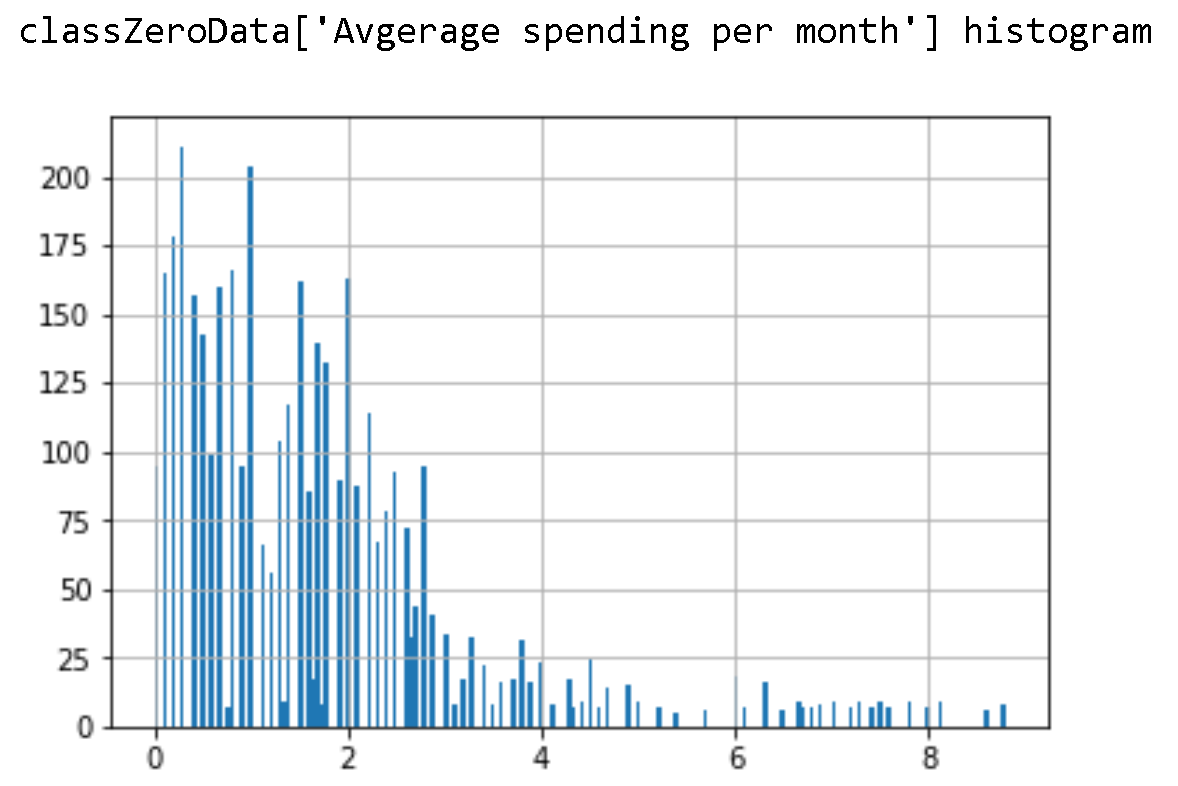
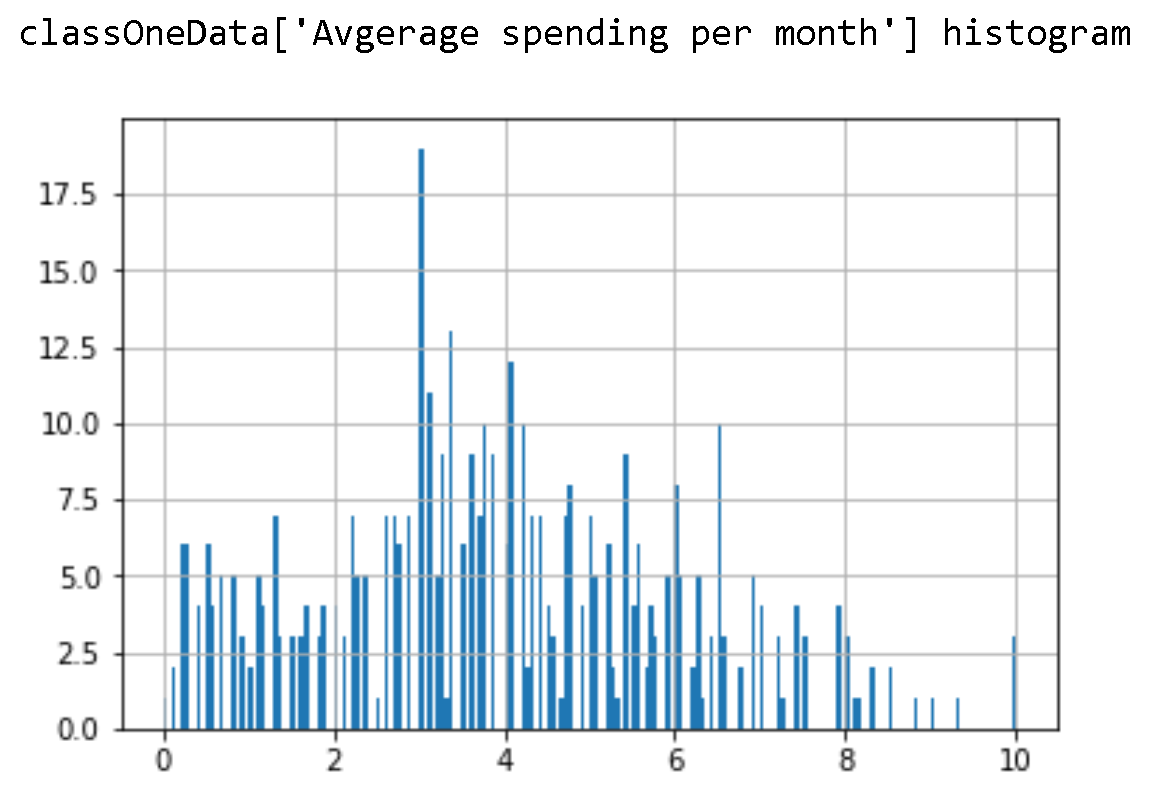
**Observations**

Here the assumption of independence used in Naïve Bayes is not followed by some attributes as they are correlated, here number of years of experience and Age have almost linear relationship, and Annual income and Average spending per month also show some dependency. Annual income and Mortgage value of the house also show some dependency.





As the histograms were more of exponential the exponential distribution is used to calculate the probability density for Annual Income, Average Spending per month attributes.

And also we can observe here that people who haven’t taken loan(class zero) are more concentrated in low annual income (0-100) where as people who have taken the loan are more concentrated in (100-200). That means from the data people who have high annual income are more probable to take the loan. Similarly because of the dependency between annual income and average spending per month we can see people with low annual income are spending less per month thus are concentrated more in range(0-4) from the histogram.

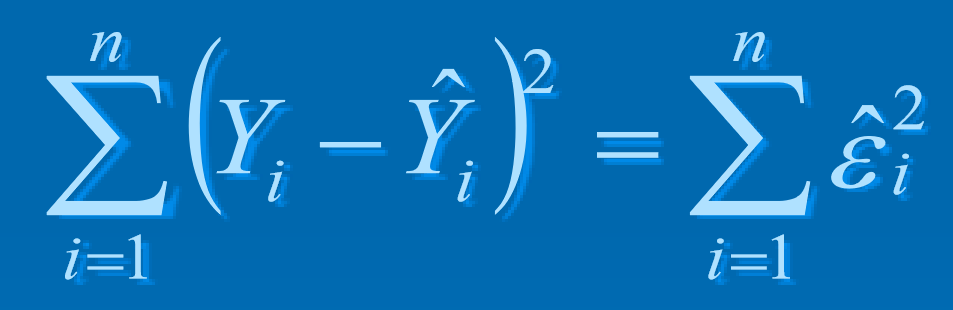
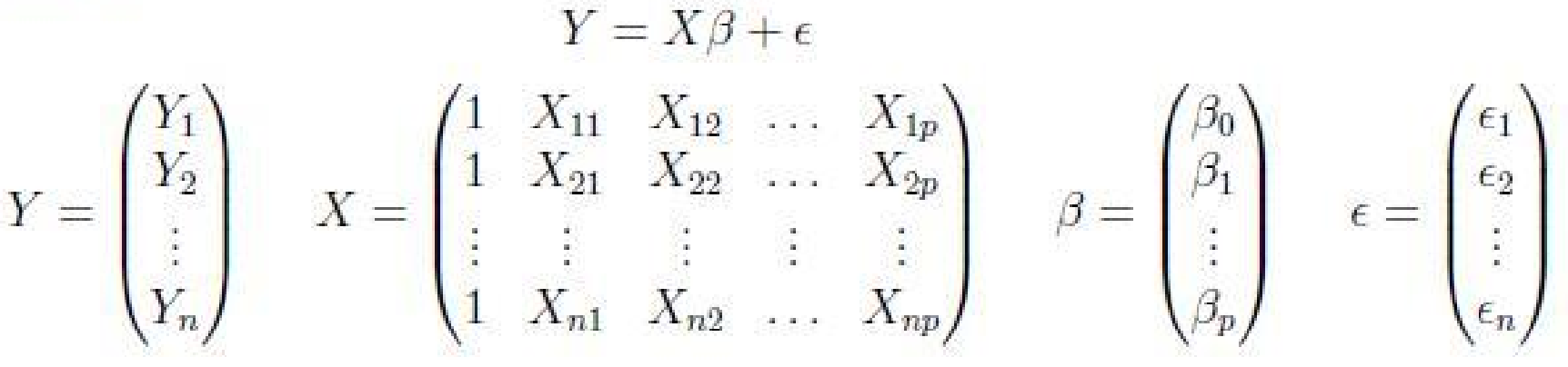
**Results On Validation Data for Loan Dataset:**

Accuracy on the validation data is: **0.9155555555555556**

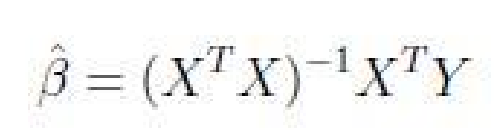
**Question3**

**Part1(Q3.1)**

Linear regression model is implemented using Matrix method. Here Minimization of Sum of the square errors is done.

The coefficients are obtained from



And thus predicted vector is given by



**Part2(Q3.2)**

The Mean Square Error (**MSE) is 0.0026201880591150122**

The Mean Absolute Error (**MAE) is 0.03950626790977792**

The Percentage Error (**MPE) is 0.017943715233214853**

Here as the error values are less that 1 the squares of those values are getting smaller than the absolute values thus MAE is more than MSE. From the positive value of mean percentage error we can see that there are more positive errors than the negative errors.

**Part3(Q3.3)**