

## 1. A very brief description of your model and its implementation.

The model is a logistic regression model that uses the cross entropy loss function as its loss function to update the weights with the following formula.

$$\theta_{t+1} = \theta_t - \eta \nabla L(f(x; \theta), y)$$

The model runs over 10 independent 70:30 train:test data splits. Each split is run on two different variations of the model, a variation that implements stochastic gradient descent and another that implements gradient descent over the whole dataset. It outputs the average accuracy, loss, precision, recall, and f-score over the 10 splits and also outputs the loss and accuracy for each split. The final weights array is the coefficients to be used to calculate the probability a prediction.

```
def sigmoid(row, weights):  
    yhat = np.dot(row, weights)  
    return 1/(1+np.exp(-yhat))
```

## 2. The most important feature in the dataset.

The most important dataset in the feature was found by scaling each attribute in the dataset and then running a model over the dataset for a few iterations and the average of the weights were obtained. The weight/coefficient with the highest absolute value was taken as the most important one, in this case, the **first** attribute was found to be the most important attribute.

**3. The final train and test metrics (loss, accuracy, recall, precision and f-score) achieved by your model with GD and SGD**

The final average test metrics were obtained as follows :

Average accuracy for SGD is 99.83009708737863  
Average accuracy for GD is 91.52912621359224

Average loss for SGD is 0.005786569970404106  
Average loss for GD is 0.5191721946546213

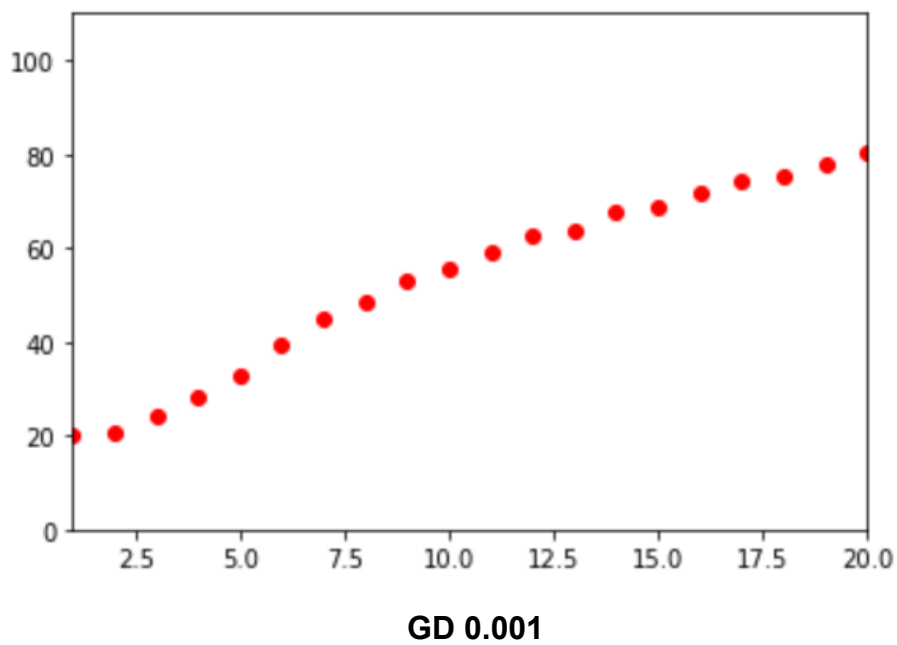
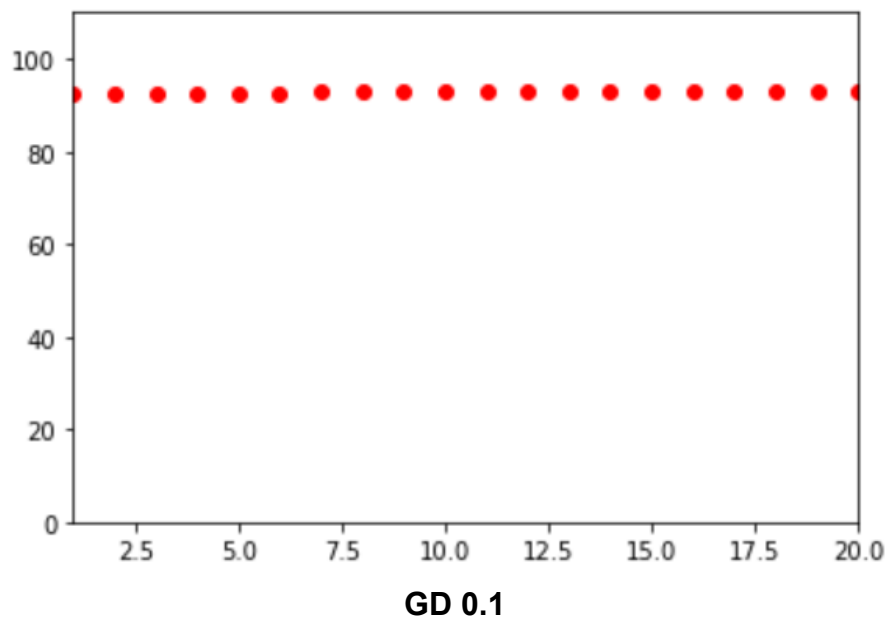
Average recall for SGD is 99.8360655737705  
Average recall for GD is 92.04027185799794

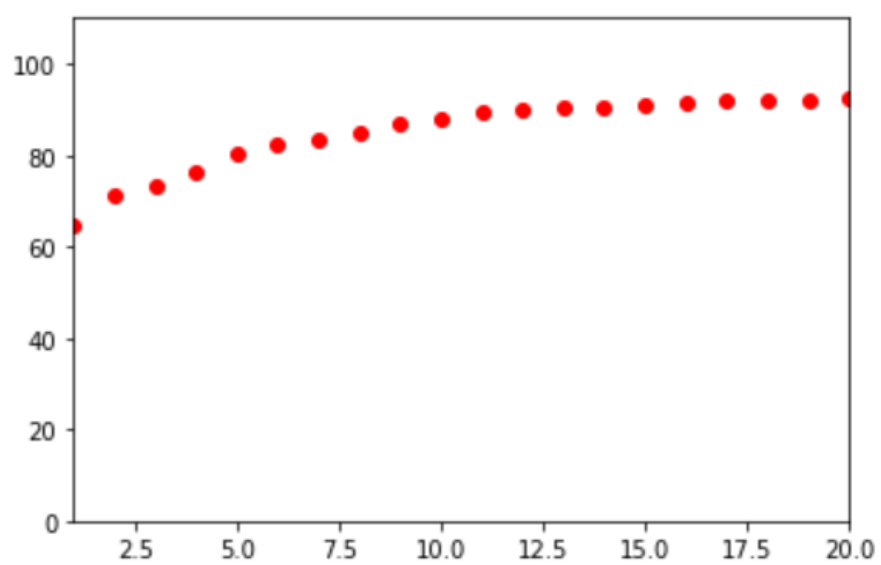
Average precision for SGD is 99.78260869565217  
Average precision for GD is 99.72972972972973

Average f-score for SGD is 0.998092643051771  
Average f-score for SGD is 0.9570329980865871

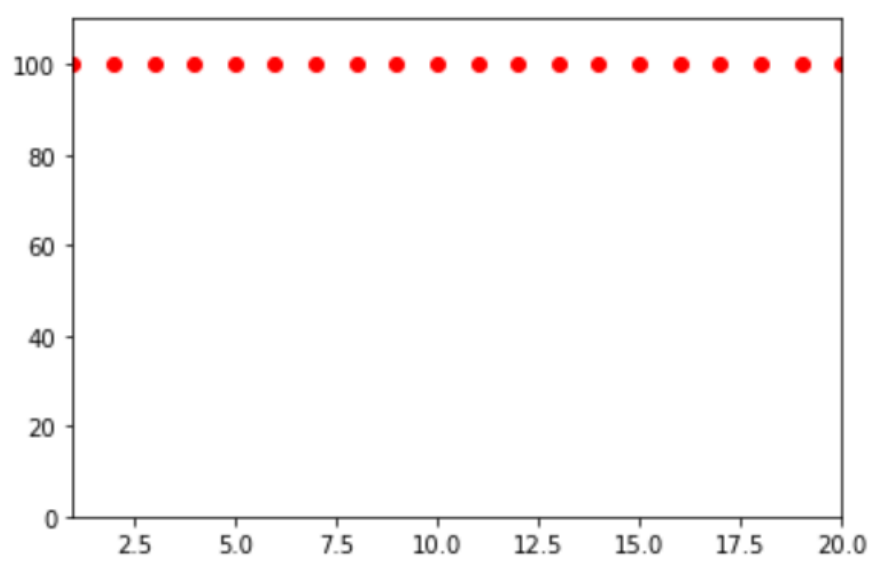
**4. Plots of accuracy for three different learning rates using GD and SGD, i.e. three plots for GD with varying learning rate (say 1, 2, 3) and three plots for SGD with the same set of learning rate on any one data split.**

Given below are three plots of SGD and GD accuracies vs epochs for the learning rates **0.1**, **0.01** and **0.001** :

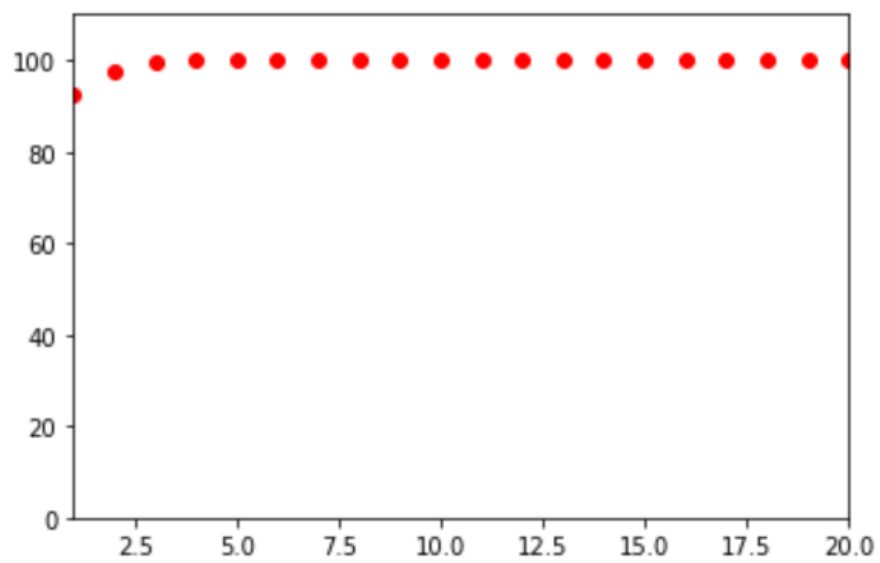




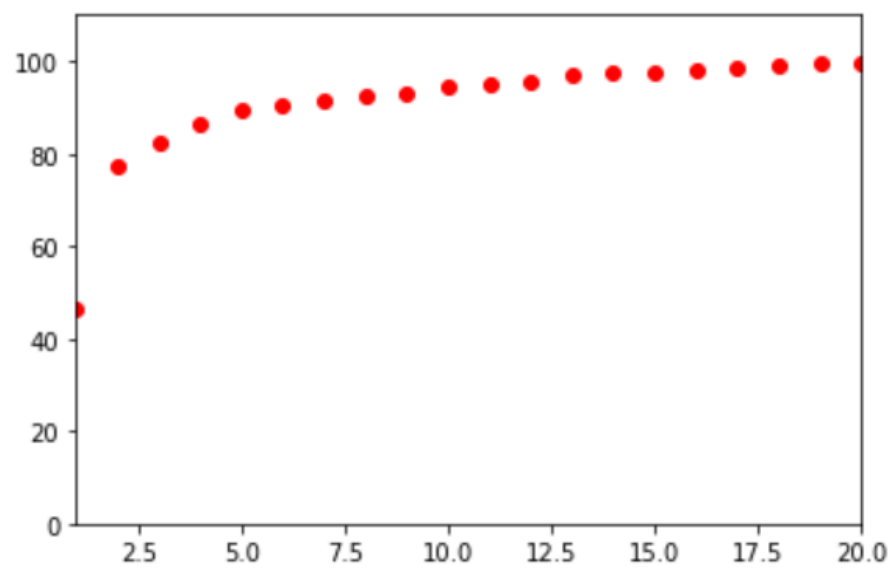
**GD 0.01**



**SGD 0.1**



**SGD 0.01**



**SGD 0.001**