Q1.1 Proportion of variance explained (PVE) for the first 10 principal components of Red, Green and Blue is reported in the Code file.

A trend was noticed in the, there was a considerable drop in PVE between the first and second principal components. It was noticed that the trend continues but the drop in the next values was not much. Overall it can be said that variance was the maximum for the first principal component , and then there was an overall drop in values one after another.

Q1.2

Principal Component Analysis (PCA) is used for two purposes: First to reduce noise and secondly reduce dimensionality. Basically PCA refers to the projection of images, on the K dimensions, where these dimensions are of the highest variances.

For the reconstruction we are going to use Eigen Vector matrix which we obtained while doing PCA. Initially, we will start by taking mean (μ) of each of the matrix we have. The mean (μ) will be subtracted from the corresponding row of the image. The subtracted matrix will then be multiplied with the Eigen vector matrix and the transpose of Eigen vector matrix. The reconstructed image will be the multiplied image added with μ. The higher the number of PCA component used the better quality of image reconstructed will be obtained.

To decrease the presence of artificially added noise, Standard Deviation (σ) is obtained for the image data, Data divided by σ is subtracted from the original image and therefore noise is reduced.

Q2.1

Ordinary Lest Square function:

$$J_n = ||y - X\beta||^2 = (y - Xc)^T(y - X\beta)$$

Now taking the derivative with respect to vector β and set that gradient to zero.

$$\nabla_\beta J_n(\beta) = 2X^T(X\beta - y) = 0$$

$$X^T X\beta = X^T y$$

By solving for β we found:

$$\beta = (X^T X)^{-1} X^T y$$

The Matrix X can be called the design matrix. This matrix allow us to show the relationship of all the data points, through $y = X\beta$

Q2.4

- If there are large outliers then MSE doesn't tell you much as it is very much affected and in this scenario MSE is robust to outliers. The Mena **absolute error** is the absolute value of difference MAE is the average magnitude of errors in a set of predictions, and therefore greater error in predictions effects the MAE.
- RMSE, because it follows a normal distribution and tell how concentrated the data is.
- The cross-validation method is preferred on larger dataset because it tells you the parameter which will give greater required outcomes in less time. Also in K fold cross validation the dataset is afterwards trained on both the validation and training set. With larger data-sets, the hyper parameters can be found quickly and therefore training would be done in less time.
- With a data-set of 50000 I will prefer cross validation as data is trained on both validation and train set further it is presumably efficient.

Q3.1

It was see after the training, that the class label prediction was mostly higher for Positive labels. For the stochastic and mini-batch gradient the precision and, Recall were high.

For Stochastic it was seen that the class label prediction (N) was relatively higher. This suggests that this training model is much efficient for predicting positive labels. One reason for this could be the unevenness in the dataset provided.

Q3.2

Similar results were obtained for this part; the full-batch gradient produced almost the same results, Accuracy for Positive labels were relatively higher. It is also noticed, that

Q3.3

The weight corresponding to a feature shows us how important a feature is; the weightage tells us the usefulness of the particular feature in predicting the labels. The feature normalization is used to bring the values closer to each other, keeping the ratio, because of this we are able to deal with features without disrupting the flow of code.

Comparison of both continuous features and categorical features are possible, and as per my understanding the most important features are Age, Gender.

Q4.1

Hard Margin SVM is relatively profound to outliers where as Soft Margin SVM avoids outliers. In this setting as per me, Hard Margin performs slightly better than soft margin.

Q4.3

The Non-linear SVM model performed better in terms of accuracy. In non-linear 'rbf' the parameters taken were 'gamma' and 'C'. In the boxplot it can be seen that NON-linear SVM performed better as it had greater mean accuracy. If it had been trained directly on image pixels, then the accuracy might have been a little better as it is plausible that some features had been lost while extracting inception features.

In this scenario, the gamma = 'scale' performed better in every instance. The large C allows us to avoid the misclassification of data. Gamma shows the influence of single sample on future predictions. The higher the value the greater the influence.