Logistic and Softmax Regression for Facial Emotion Classification

Devanshu Desai
Department of Data Science
University of California San Diego
San Diego, CA 92093
dbdesai@ucsd.edu

Nishant Mysore
Department of Bioengineering
University of California San Diego
San Diego, CA 92093
nmysore@ucsd.edu

Ranak Roy Chowdhury
Department of Computer Science and Engineering
University of California San Diego
rrchowdh@eng.ucsd.edu

Abstract

Emotion classification is the means by which one may distinguish one emotion from another. Emotion computing, or sentiment analysis, is one of the most active research areas. In this work, we try to distinguish between human emotions from the images of their faces. We developed a logistic and a softmax regression based model to develop our framework. Results show that simple regression models perform reasonably well in learning facial features that correspond to each emotion.

1 Introduction and Methods

1.1 Background

The dataset that was used to train our logistic regression and softmax models is the Extended Cohn-Kanad Dataset [1]. We used six separate emotions (happiness, sadness, surprise, anger, disgust, and fear). The data was separated into two catagories: aligned, and unaligned, where the aligned set was preprocessed to align faces. Due to variance in the number of images of each emotion, a balanced sampler was used to draw the same number of images for each facial expression.

1.2 Cross Validation Procedure

We used a k-fold cross Validation procedure where the data was divided into k mutually exclusive sets. We repeated the following procedure k times: Two sets were designated as holdout and test, while the other (k-2) sets were trained on. The holdout error and testing error were computed over 50 epochs. The model with the lowest validation error after running k-times would be selected as the "best" model, and its accuracy on the test set would be computed. This procedure was completed for all the following experiments unless otherwise noted.

1.3 Principal Component Analysis

We performed Principal Component Analysis (PCA) on the images for dimensionality reduction. We used Turk and Penland's Trick [2] to reduce computational expense, and projected over k principal components. Each experiment performed will note the number of principal components used.

2 Logistic Regression

Logistic Regression is used to model the probability of a certain class or event existing such as pass/fail, win/lose, or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1 and the sum adding to one.

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled ”0” and “1”. In the logistic model, the log-odds (the logarithm of the odds) for the value labeled “1” is a linear combination of one or more independent variables (“predictors”); the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labeled “1” can vary between 0 (certainly the value “0”) and 1 (certainly the value “1”), hence the labeling; the function that converts log-odds to probability is the logistic function, hence the name. The defining characteristic of the logistic model is that increasing one of the independent variables multiplicatively scales the odds of the given outcome at a constant rate, with each independent variable having its own parameter; for a binary dependent variable this generalizes the odds ratio.

2.1 Happiness Vs Anger on Resized

We trained a logistic regression model on the happiness vs Anger dataset, on the resized images. Our best Test loss was 0.684, and our Best Test Accuracy was 55.6%. This model did not perform that well because the images were not aligned, and such the model could not properly distinguish faces.

vs Angry.png

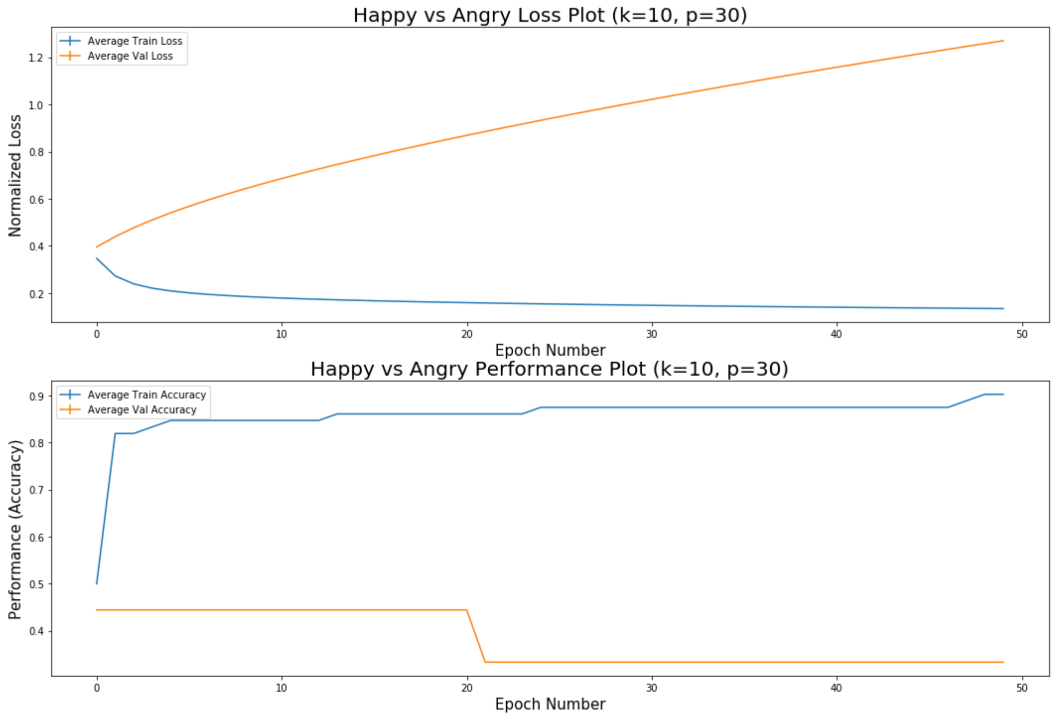

Figure 1: Comparison of the variation between training and validation normalized loss and accuracy with epoch for the Logistic regression classifier for happy vs Angry classes. The training loss goes down with epoch as the weights get progressively better and more representative of the facial characteristics. However, the validation loss goes higher with increasing number of epochs. Moreover, the validation loss is higher than the training loss because the model has been trained on the training set. Therefore, it doesn't perform as well on the validation set as it does on the training set.

To understand this, we need to look at the first four principal components.

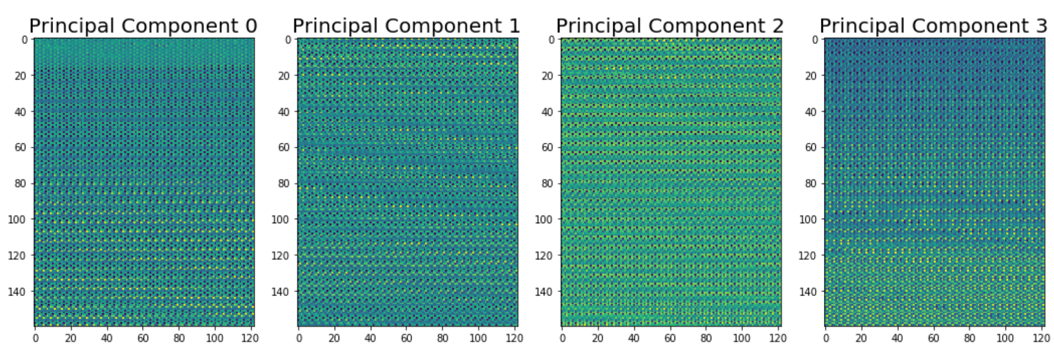

Figure 2: These components look very noisy, which is a result of using images that are not aligned properly. Thus, the principal component analysis fails to have any significant effect on improving performance.

2.2 Happiness vs Anger on Aligned Dataset

Let's first look at our first four principal components:

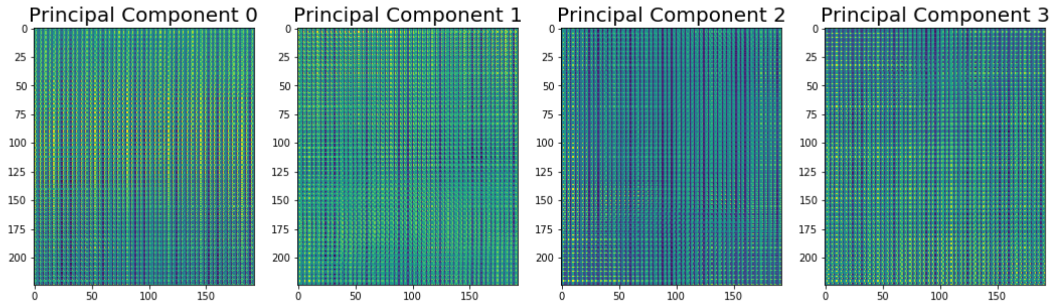

Figure 3: These components are much more dense than the aligned principal components. Also, these are way less noisy than the resized ones as these do not pack information about the subject's background. All faces are aligned on the same scale.

We trained another logistic regression model on the Happiness vs Anger categories on the aligned dataset. We used a learning rate of 0.01, and generated a test loss of 0.37, with a best test accuracy of 98.9% (0.03)

With a learning rate of 0.5, our best test loss was 2.05, with a best test accuracy of 94.4% (0.13). This model performed worse than our optimal learning rate model.

With a learning rate of 1e-20, , our best test loss was 0.34, and our best test accuracy was 94.4% (0.13). This learning rate was much too low, and resulted in a severe decrease in performance.

vs Angry lr=1e-20.png

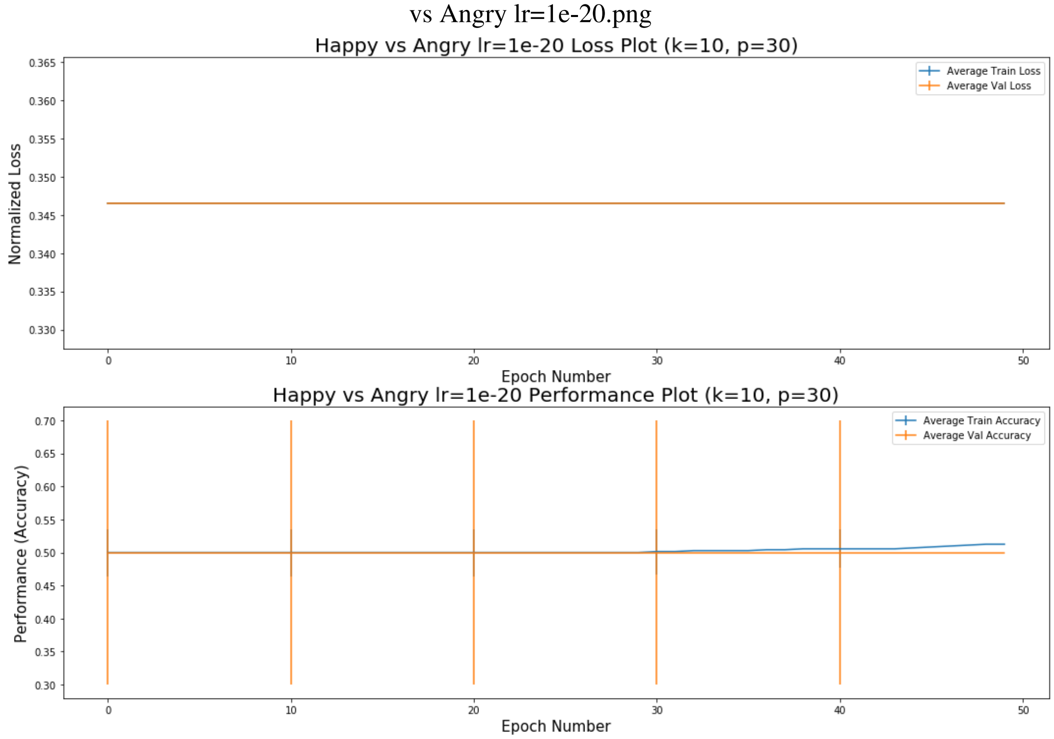

Figure 4: This is a graph of the happiness vs Angry model trained at a much too low learning rate. As you can see, model performance is extremely poor, and the network is unable to learn due to the extremely low learning rate.

vs Angry lr=1e-2.png

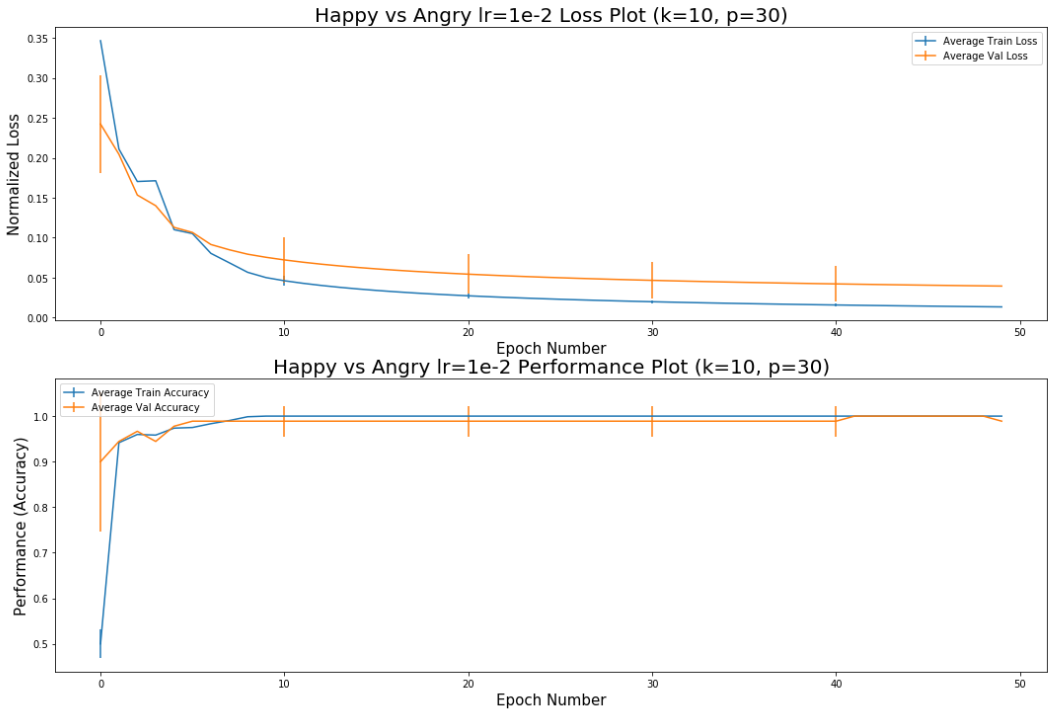

Figure 5: This is a graph of the happiness vs Angry model trained at a correct learning rate. As you can see, model loss steadily decreases, and the network is able to learn due to the optimal learning rate.

vs Angry lr=5e-1.png

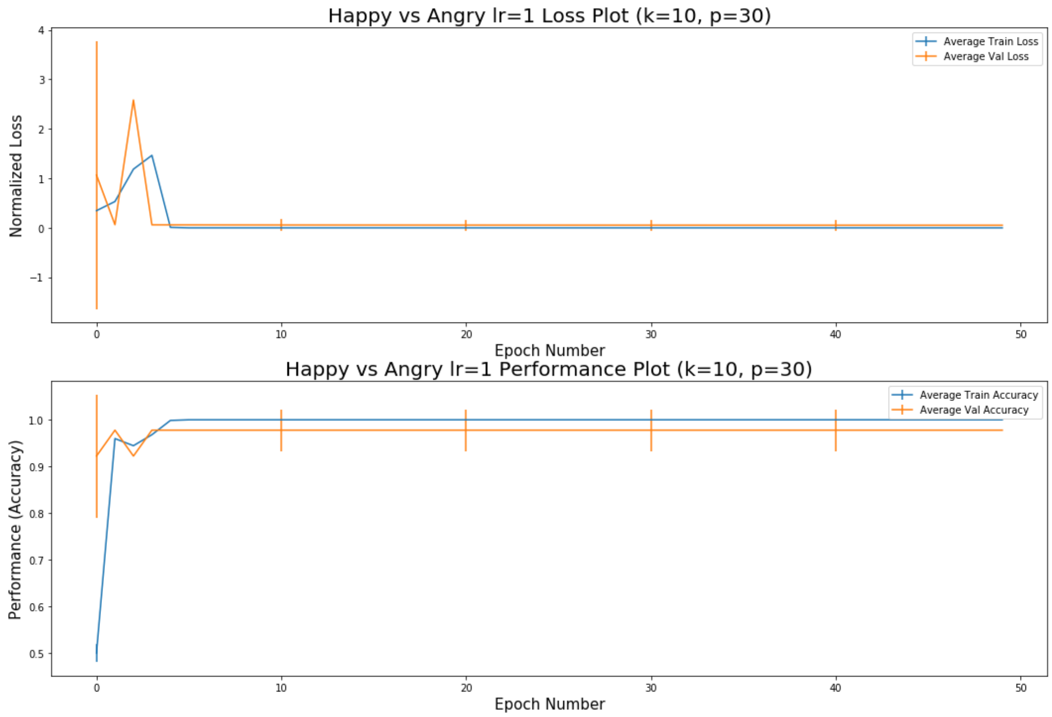

Figure 6: This is a graph of the happiness vs Angry model trained at a much too high learning rate. As you can see, model performance is not as good as that of our optimal learning rate, and the network is not able to learn effectively

2.3 Fear vs Surprise on aligned dataset

We trained another logistic regression model on Fear vs Surprise in the aligned dataset. This generated a Average Test loss of 0.153, and an Average Test Accuracy of 94.0%, when trained with a learning rate of 0.01

This model performs a bit worse than our Happiness vs Anger model, which we think is because of the fact that the emotions of fear and surprise are very similar, and thus, the model has a more difficult time distinguishing the two emotions.

vs Surprise.png

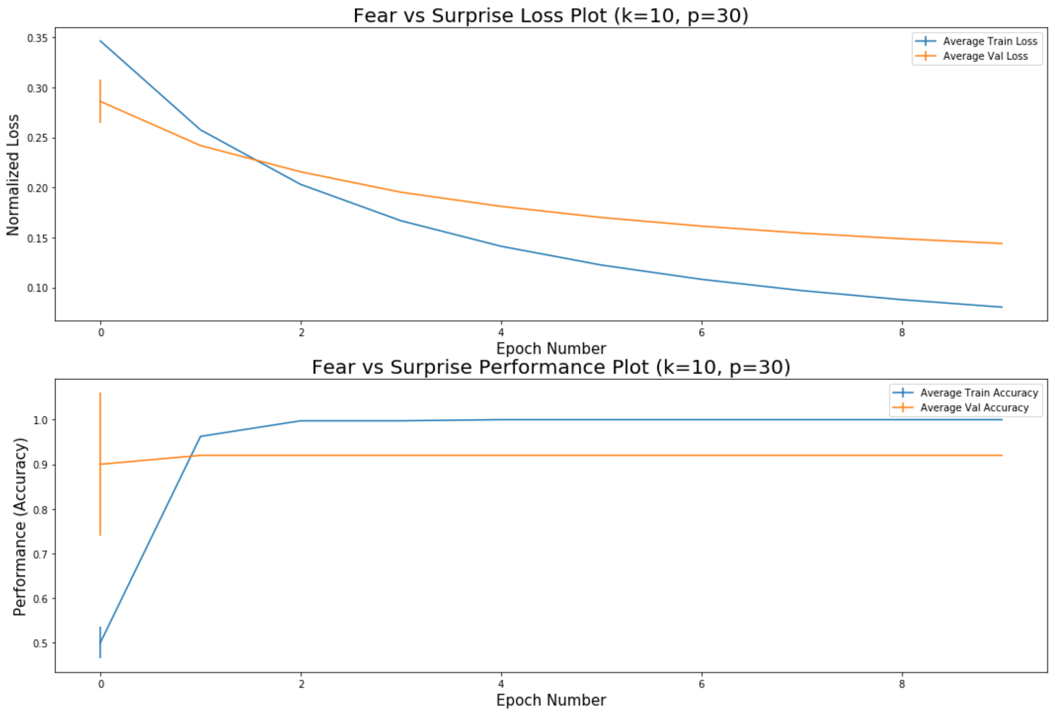

Figure 7: Comparison of the variation between training and validation normalized loss and accuracy with epoch for the logistic regression classifier for the Fear vs Surprise Classes. As expected both the training and validation loss goes down with epoch as the weights get progressively better and more representative of the facial characteristics. However, the validation loss is higher than the training loss because the model has been trained on the training set. Therefore, it doesn't perform as well on the validation set as it does on the training set. Both the training and validation accuracies are high compared to the baseline accuracy of random guessing for 2 classes, which is 50%. The training accuracy is still higher than the validation accuracy because the model has been trained on the training set.

3 Softmax Regression

Softmax regression is a generalized form of logistic regression which can be used in multi-class classification problems where the classes are mutually exclusive.

	AVERAGE TEST ACCURACY (STD)	AVERAGE TEST LOSS (STD)
1.	0.747 (0.03)	0.154 (0.02)

3.1 Evaluation on all six emotions

Six Emotions.png

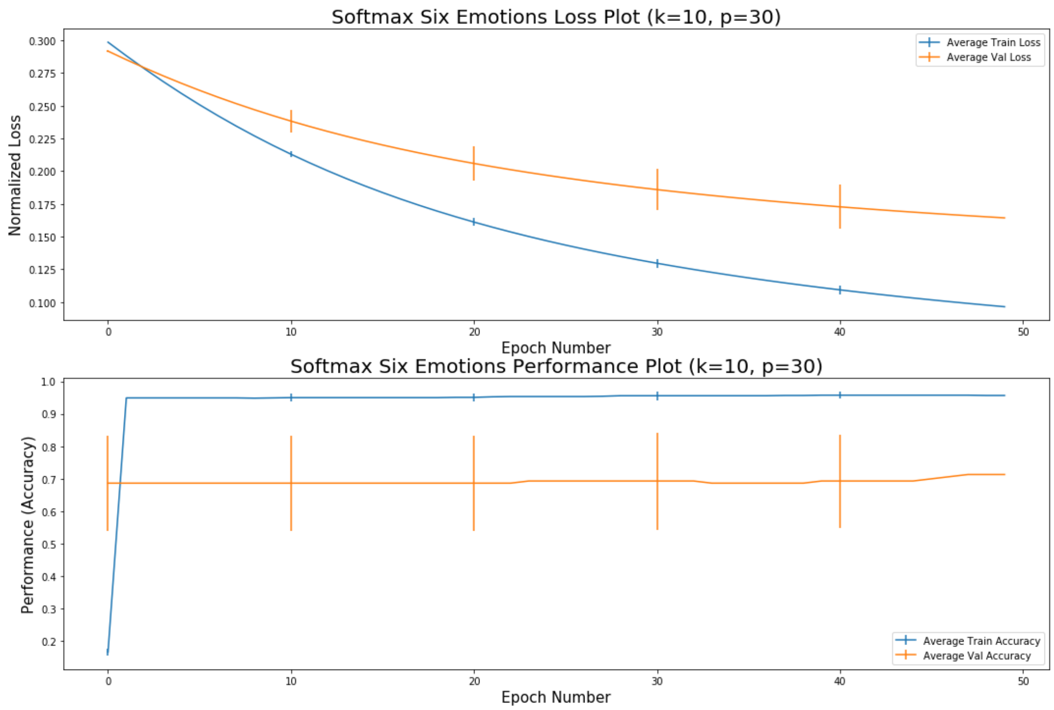

Figure 8: Comparison of the variation between training and validation normalized loss and accuracy with epoch for the softmax regression classifier. As expected both the training and validation loss goes down with epoch as the weights get progressively better and more representative of the facial characteristics. However, the validation loss is higher than the training loss because the model has been trained on the training set. Therefore, it doesn't perform as well on the validation set as it does on the training set. The validation loss has a higher standard deviation across folds than the training loss. Both the training and validation accuracies are high compared to the baseline accuracy of random guessing for 6 classes, which is 16.67%. The training accuracy is still higher than the validation accuracy because the model has been trained on the training set. The validation accuracy has a higher standard deviation across folds than the training accuracy.

	Fear	Surprise	Sadness	Happiness	Anger	Disgust
Fear	2	0	0	0	0	0
Surprise	0	2	0	0	0	0
Sadness	0	1	3	0	1	0
Happiness	0	0	0	2	0	0
Anger	0	0	0	0	1	0
Disgust	0	0	1	0	0	2

Table 1: Confusion Matrix of our results. The rows encode the real category, and the columns encode the predicted category. We misclassified 2/15 examples in the testing set. Misclassifications were in category 3 and 5

3.2 Batch versus Stochastic Gradient Descent

v Stochastic.png

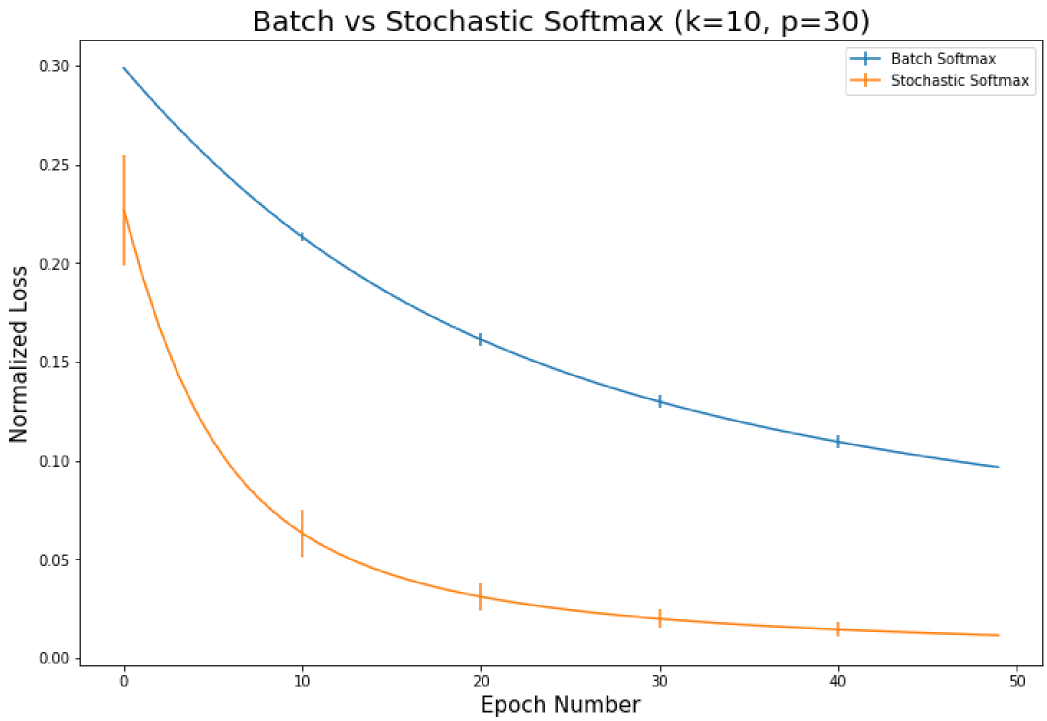

Figure 9: Figure showing our Softmax model being trained via Batch and Stochastic Stochastic Gradient Descent converges faster than Batch Gradient Descent. This is because in stochastic gradient descent, we update our weights M times per epoch, where M is the number of training examples. However, in Batch Gradient Descent, we only update weights once per epoch, which leads to slower convergence time.

1.

3.3 Visualize The Weights

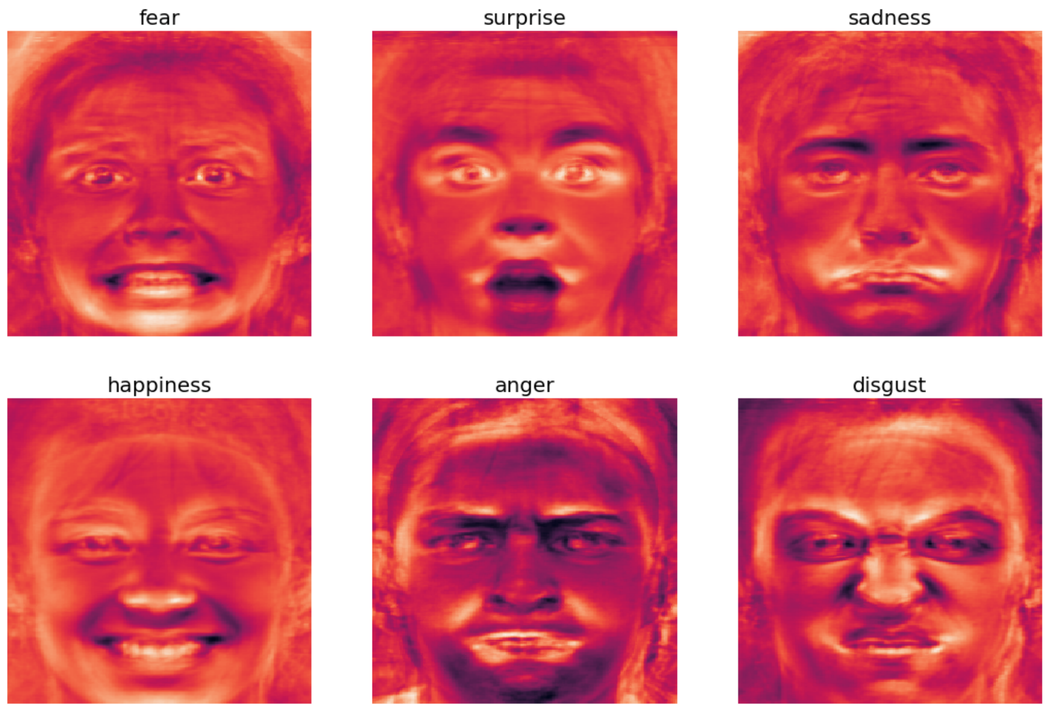

Figure 10: Visualization of the six weight vectors learnt from the training data. As evident from the diagrams, the weights reflect the salient feature of each emotion. For example, the weights capture that happiness and fear both have a strong correlation with our lips, teeth and nose. Because smile and fear generally widens our lips and nose, and exposes our teeth (first image of the second and first row). Similarly, sadness evokes a contraction of the lips as highlighted by the third image in the first row. Surprise tends to make our eyes go wider (second image in the first row) while anger and disgust causes people to frown (second and third image of the second row). As illustrated by the figures, the weight vectors are highlighting the appropriate section of the face for each emotion.

3.4 Extra Credit

We made some modifications to out softmax regression so that it automatically picks the lowest common denominator whenever there is a class imbalance in our training set.

4 Individual Contributions to the Project

Devanshu worked on implementing the K-fold Cross Validation algorithm, as well as the Softmax regression functions and code. He generated figures and accuracy measurements for the report.

Nishant worked on the PCA algorithm and debugging the logistic and softmax regression functions. He also worked on some graphic utilties, and helped to write the background and logistic sections of the report.

Ranak developed the functions for logistic regression, batch gradient descent, stochastic gradient descent, cross-entropy and accuracy evaluation. He also wrote the softmax regression sections of the report.