Report Assignment 3

By:

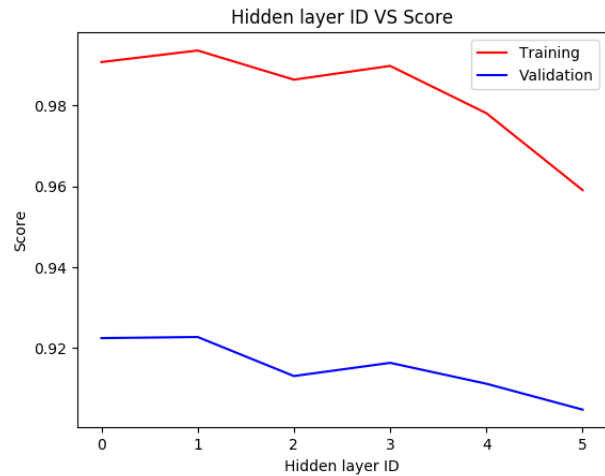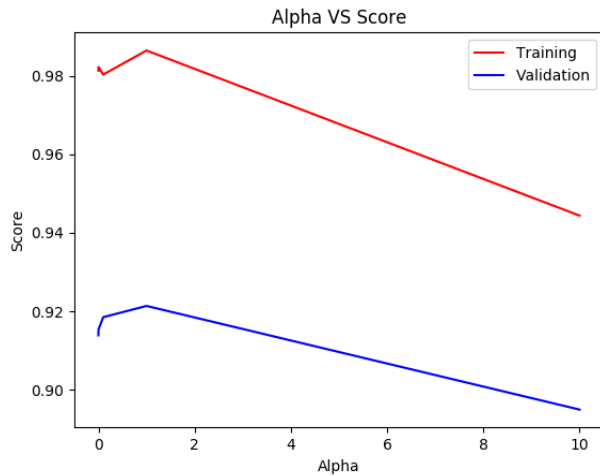|	
Nour ElDin Khaled	31-7974
Kareem Mokhtar	31-609
Narihan Ellaithy	31-8027

Without PCA:

MLP:

We have tried changing the number of hidden layers as well as the number of neurons in each layer, also we have tried changing the regularization coefficient (Lambda/alpha).
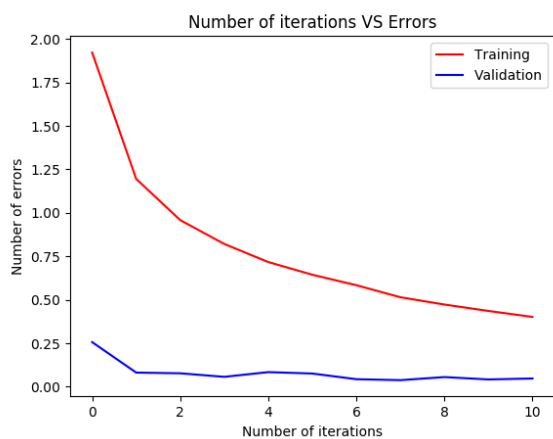The values used were:-
Hidden Layers: [(300), (400), (300, 300), (300, 400), (400, 300), (400, 400)]
Alpha: [0.00001, 0.001, 0.1, 1, 10]

As shown in the above plots, The best score was produced when we used The hidden layers with index 1 (which is (400)), and alpha value of 1.

So we built a MLP with those parameters. Here is the training and validation curves (left), and the classification report of the model (right):-

	precision	recall	f1-score	support
1.0	0.90	0.99	0.95	496
2.0	0.93	0.94	0.93	471
3.0	0.99	0.87	0.93	420
4.0	0.95	0.85	0.89	491
5.0	0.88	0.95	0.91	532
6.0	1.00	1.00	1.00	537
avg / total	0.94	0.94	0.94	2947

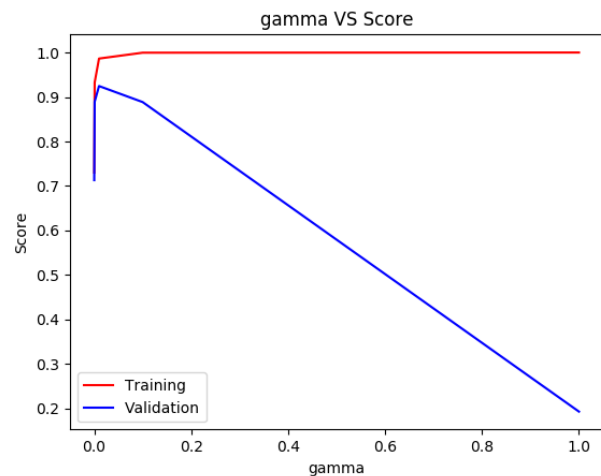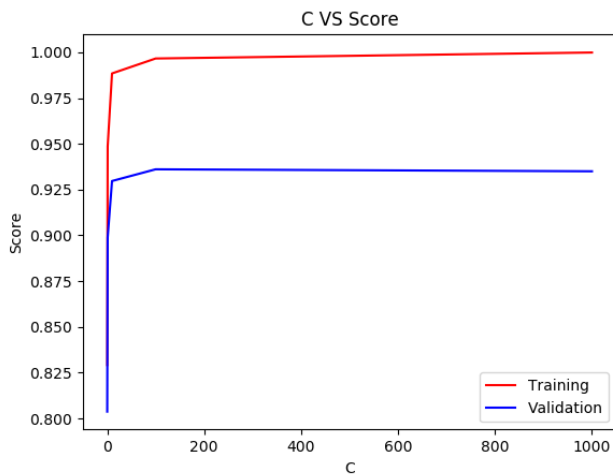As you can see, the accuracy of the MLP is 94%.

<u>SVM</u>:
We have tried changing the C and gamma parameters.
The values used were:-
C: [0.1, 1, 10, 100, 1000]
gamma: [1, 0.1, 0.01, 0.001, 0.0001]

As shown in the above plots, The best score was produced when we C was 100 (or 1000), and gamma was 0.01.

So we built a SVC with those parameters. Here is the classification report of the model:-

|	precision	recall	f1-score	support
1.0	0.96	0.98	0.97	496
2.0	0.94	0.97	0.95	471
3.0	0.99	0.94	0.97	420
4.0	0.98	0.91	0.94	491
5.0	0.92	0.98	0.95	532
6.0	1.00	1.00	1.00	537
avg / total	0.97	0.96	0.96	2947

As you can see, the accuracy of the SVC is 96%.

- The SVC performed better (had better accuracy) than the MLP, as the SVC tries it's hardest to make the data linearly separable by transforming the data to higher dimension if they are non-linearly separable in the current dimension, thus ensuring the best performance. However it can be computationally expensive that is why we introduce the C parameter, so that we can be content with a classifier that is not perfect (doesn't have 100% accuracy), in exchange of saving some computational time.

With PCA (Number of components is 5):

MLP:
We tried changing the same parameters as the MLP without PCA, and we used the same values.
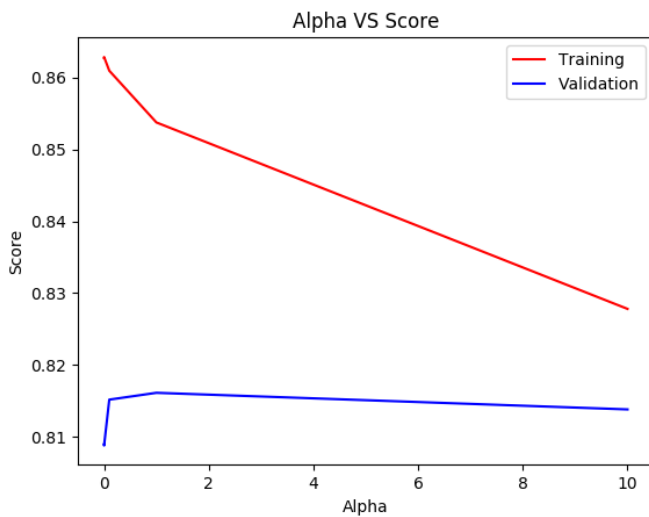
As shown in the above plots, The best score was produced when we used The hidden layers with index 1 (which is (400)), and alpha value of 1.

So we built a MLP with those parameters. Here is the training and validation curves (left), and the classification report of the model (right):-
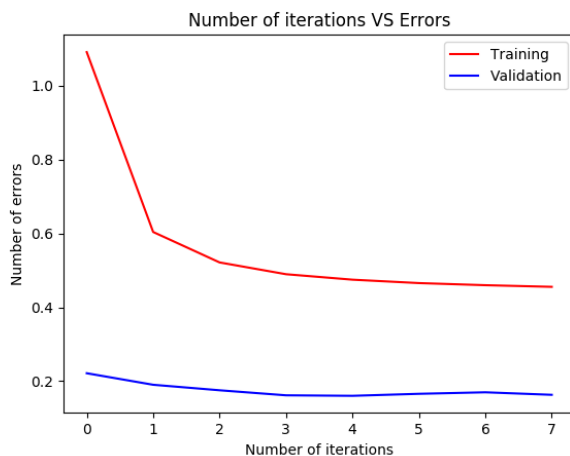

```
               precision    recall  f1-score   support

         1.0       0.73      0.92      0.81       496
         2.0       0.89      0.78      0.83       471
         3.0       0.76      0.64      0.70       420
         4.0       0.84      0.53      0.65       491
         5.0       0.68      0.92      0.78       532
         6.0       1.00      0.99      0.99       537

  avg / total       0.82      0.80      0.80      2947
```
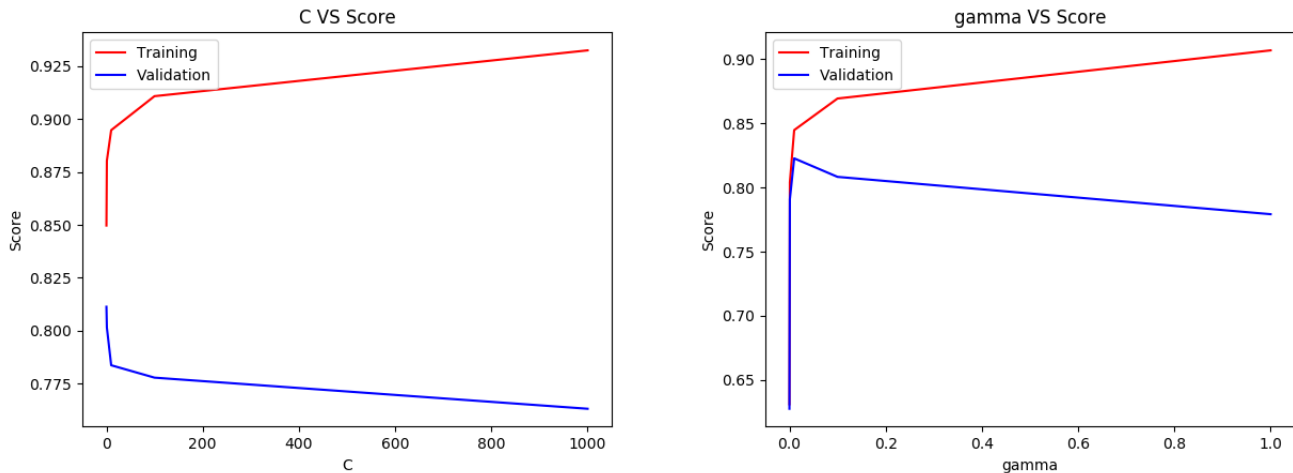
As you can see, the accuracy of the MLP is 80%, which is really bad, indicating that with just 5 principle components, a lot of important features were eliminated.

<u>SVM</u>:
We tried changing the same parameters as the SVM without PCA, and we used the same values.

As shown in the above plots, The best score was produced when we C was 0.1, and gamma was 0.01.

So we built a SVC with those parameters. Here is the classification report of the model:-

	precision	recall	f1-score	support
1.0	0.72	0.92	0.81	496
2.0	0.90	0.88	0.89	471
3.0	0.82	0.59	0.68	420
4.0	0.95	0.43	0.59	491
5.0	0.65	0.98	0.78	532
6.0	1.00	0.99	1.00	537
avg / total	0.84	0.81	0.80	2947

As you can see, the accuracy of the SVC is also 80%, ensuring that 5 principle components is just not enough.
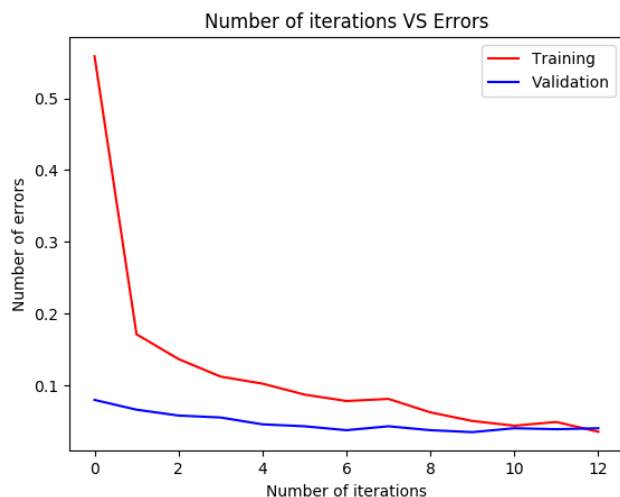
With PCA (Number of components is 50):

MLP:
We tried changing the same parameters as the MLP without PCA, and we used the same values.
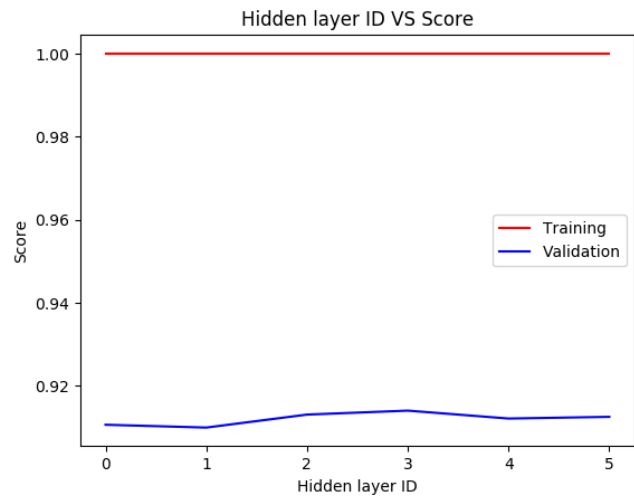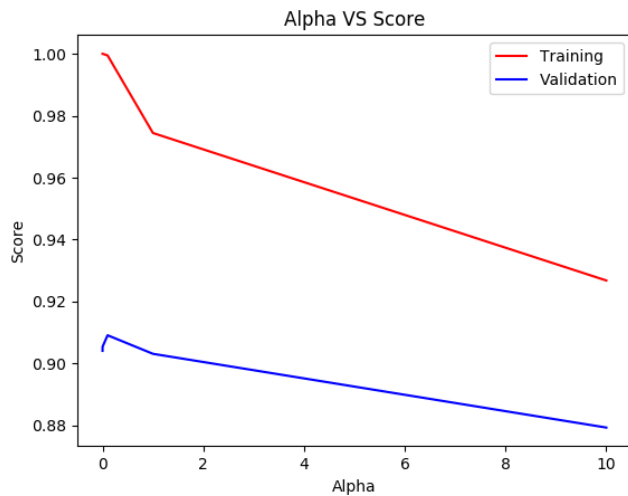
As shown in the above plots, The best score was produced when we used The hidden layers with index 3 (which is (300, 400)), and alpha value of 0.01.

So we built a MLP with those parameters. Here is the training and validation curves (left), and the classification report of the model (right):-
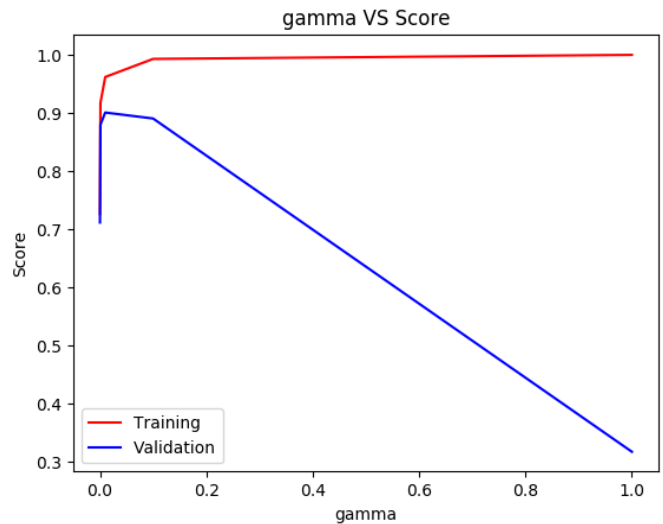

	precision	recall	f1-score	support
1.0	0.93	0.96	0.95	496
2.0	0.90	0.92	0.91	471
3.0	0.94	0.88	0.91	420
4.0	0.91	0.88	0.90	491
5.0	0.90	0.92	0.91	532
6.0	1.00	1.00	1.00	537
avg / total	0.93	0.93	0.93	2947

As you can see, the accuracy of the MLP is 93%.

<u>SVM</u>:
We tried changing the same parameters as the SVM without PCA, and we used the same values.

As shown in the above plots, The best score was produced when we C was 10, and gamma was 0.01.

So we built a SVC with those parameters. Here is the classification report of the model:-

	precision	recall	f1-score	support
1.0	0.92	0.93	0.92	496
2.0	0.87	0.91	0.89	471
3.0	0.95	0.89	0.92	420
4.0	0.95	0.88	0.91	491
5.0	0.90	0.95	0.92	532
6.0	1.00	1.00	1.00	537
avg / total	0.93	0.93	0.93	2947

As you can see, the accuracy of the SVC is 93%.

With PCA (Number of components is 200):

MLP:
We tried changing the same parameters as the MLP without PCA, and we used the same values.

As shown in the above plots, The best score was produced when we used The hidden layers with index 1 (which is (400)), and alpha value of 0.00001.

So we built a MLP with those parameters. Here is the training and validation curves (left), and the classification report of the model (right):-
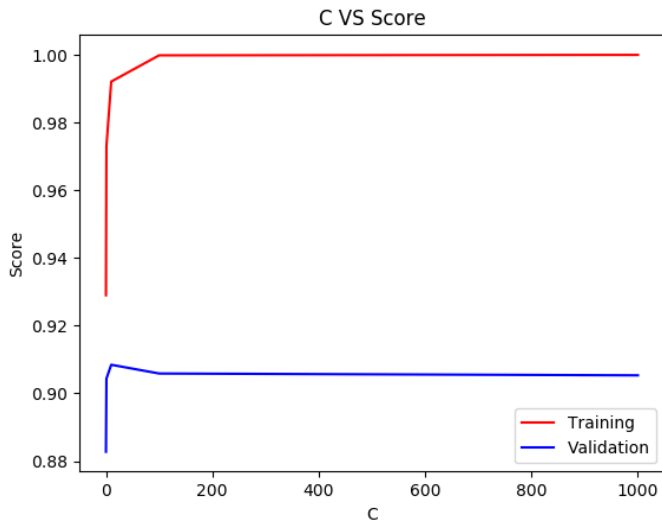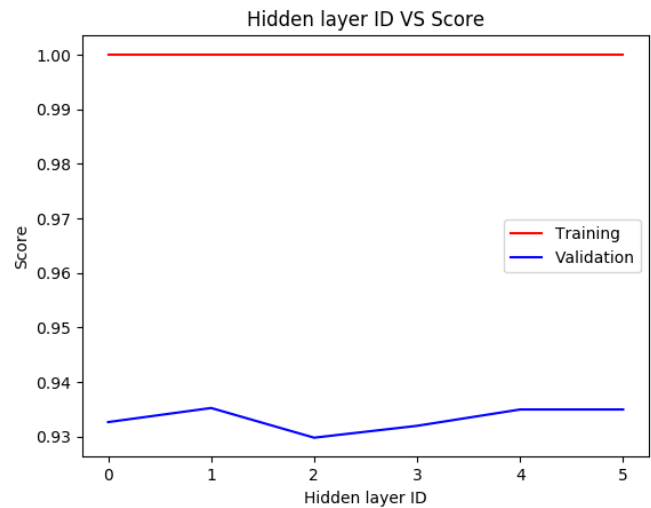

	precision	recall	f1-score	support
1.0	0.96	1.00	0.98	496
2.0	0.97	0.96	0.97	471
3.0	1.00	0.97	0.98	420
4.0	0.96	0.89	0.92	491
5.0	0.88	0.97	0.92	532
6.0	1.00	0.97	0.98	537
avg / total	0.96	0.96	0.96	2947

As you can see, the accuracy of the MLP is 96%.

<u>SVM</u>:
We tried changing the same parameters as the SVM without PCA, and we used the same values.
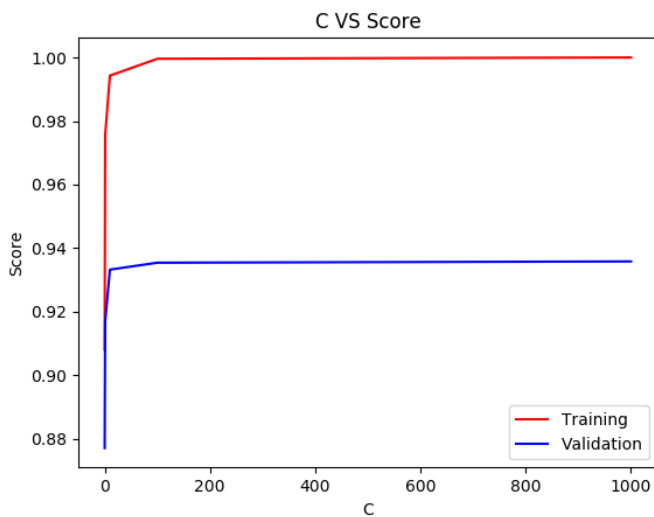
As shown in the above plots, The best score was produced when we C was 100 (or 1000), and gamma was 0.01.

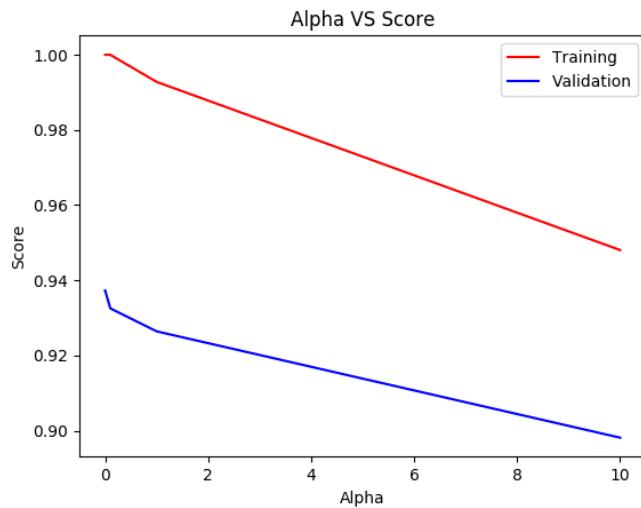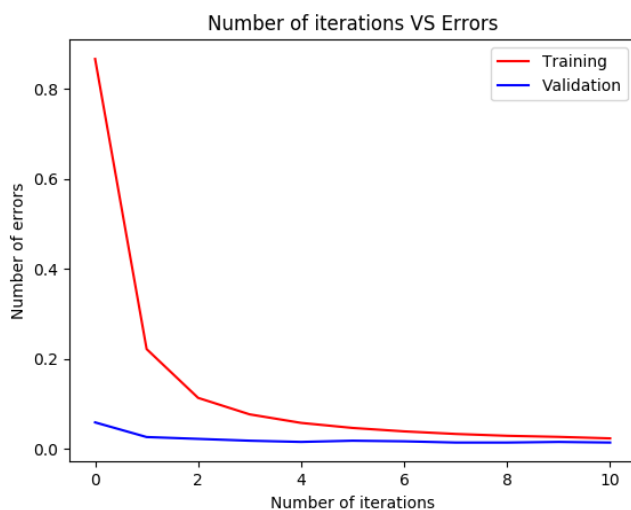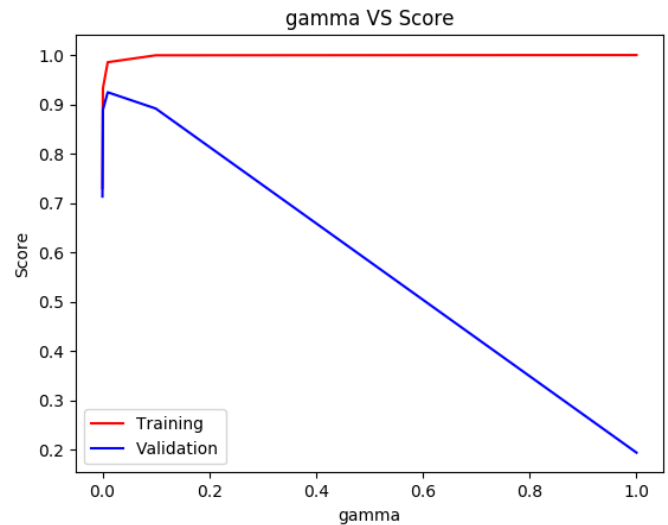So we built a SVC with those parameters. Here is the classification report of the model:-

	precision	recall	f1-score	support
1.0	0.95	0.98	0.96	496
2.0	0.94	0.95	0.95	471
3.0	0.99	0.94	0.96	420
4.0	0.98	0.90	0.94	491
5.0	0.92	0.98	0.95	532
6.0	1.00	1.00	1.00	537
avg / total	0.96	0.96	0.96	2947

As you can see, the accuracy of the SVC is 96%.

With PCA (Number of components is 500):

MLP:

We tried changing the same parameters as the MLP without PCA, and we used the same values.

Alpha VS Score

Hidden layer ID VS Score

As shown in the above plots, The best score was produced when we used The hidden layers with index 5 (which is (400, 400)), and alpha value of 0.00001.

So we built a MLP with those parameters. Here is the training and validation curves (left), and the classification report of the model (right):-
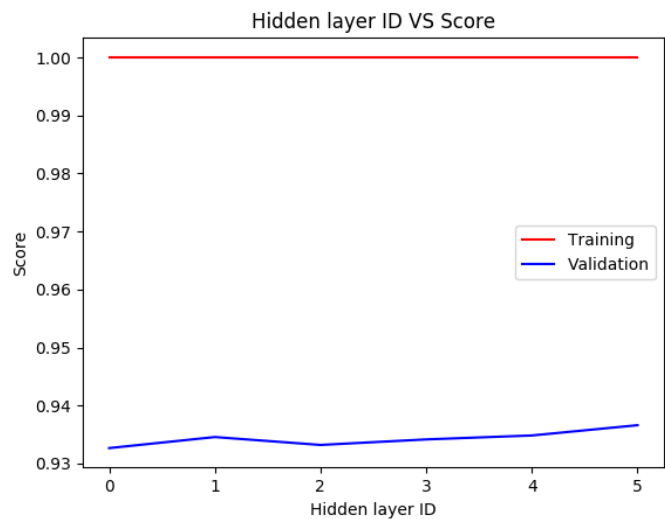
Number of iterations VS Errors


```
              precision    recall  f1-score   support

        1.0       0.95      1.00      0.97       496
        2.0       0.98      0.95      0.96       471
        3.0       1.00      0.97      0.99       420
        4.0       0.98      0.88      0.92       491
        5.0       0.87      0.98      0.92       532
        6.0       1.00      0.97      0.98       537

avg / total       0.96      0.96      0.96      2947
```

As you can see, the accuracy of the MLP is 96%.

<u>SVM</u>:
We tried changing the same parameters as the SVM without PCA, and we used the same values.

As shown in the above plots, The best score was produced when we C was 100 (or 1000), and gamma was 0.01.

So we built a SVC with those parameters. Here is the classification report of the model:-
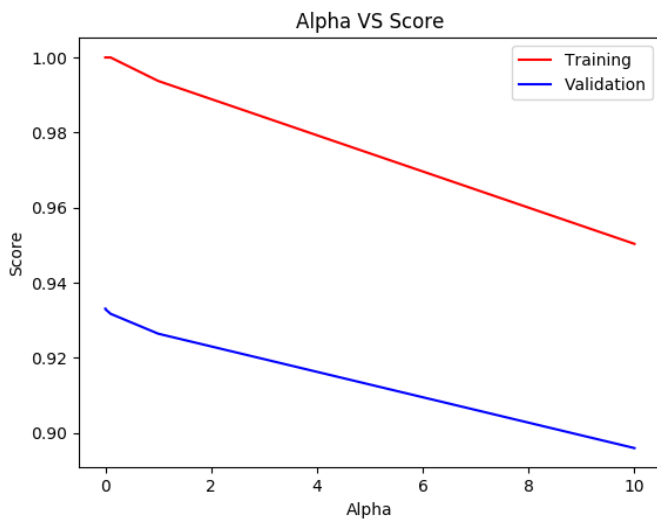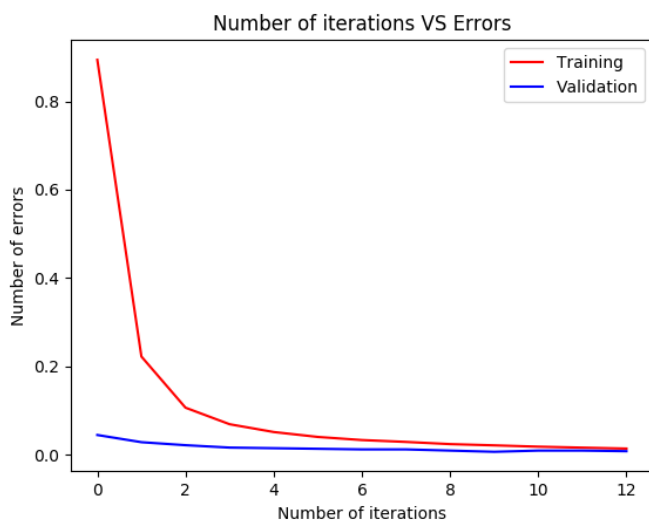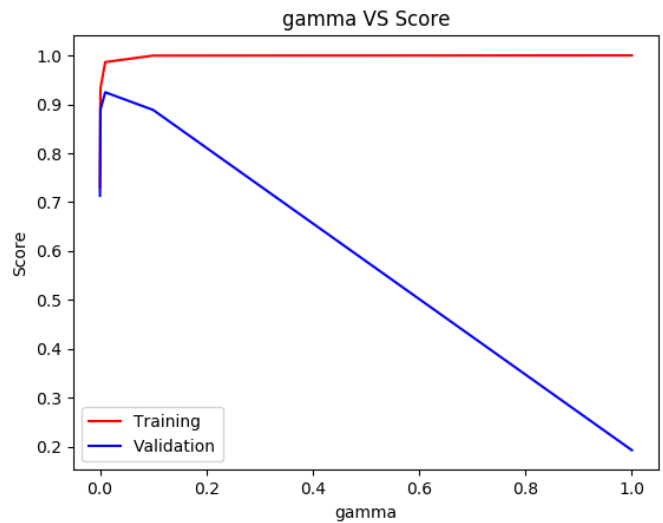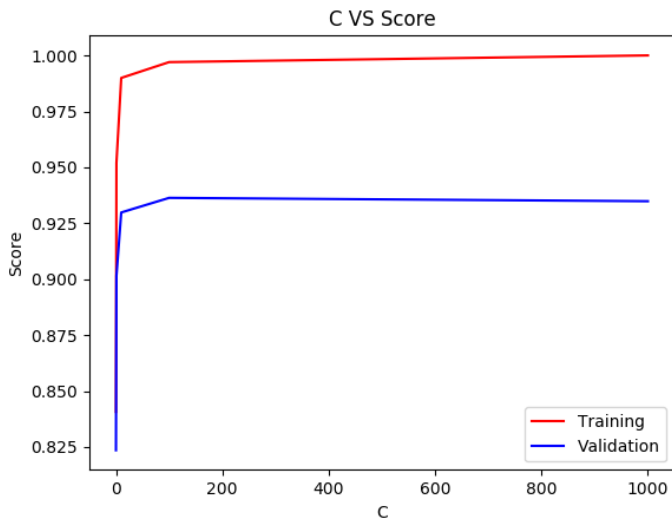
```
              precision    recall  f1-score   support

         1.0       0.96      0.98      0.97       496
         2.0       0.94      0.97      0.95       471
         3.0       0.99      0.94      0.97       420
         4.0       0.98      0.91      0.94       491
         5.0       0.92      0.98      0.95       532
         6.0       1.00      1.00      1.00       537

 avg / total       0.97      0.97      0.97      2947
```

As you can see, the accuracy of the SVC is 97%, which is the best accuracy that we could achieve among all models.

- So the best performance was achieved when using PCA with number of components equal to 500. This is logical as a low number of components can result in lose of important features (as we saw with PCA of 5 components), thus as we increase the number of components the performance increases because we don't eliminate important features. However this also means that we should get the highest accuracy when we use all the features (without PCA), but this is not the case, since some features can be redundant and unnecessary for the model to keep track of and train on.

- So when using PCA with number of components equal to 500, this is a balance between keeping track of the important features and eliminating the redundant features, thus it produces the highest accuracy.