CS 464: Introduction to Machine Learning

Spring 2024

Homework 2

Question 1

Q. 1.1

The goal of principal component analysis (PCA) is to reorient the axes to capture the most amount of variance using the least amount of dimensions. In this case, we assume that variance is positively correlated with usefulness. So, when the proportion of variance explained (PVE) of a principal component (PC) is large, we understand that that PC is more useful compared to other PCs. PVE of a PC is an important metric because if we have a total variance explained target we can select PCs starting from the one that explains the most amount of variance and keep selecting the PCs with the next largest PVE values until the total PVE is larger than or equal to our target.

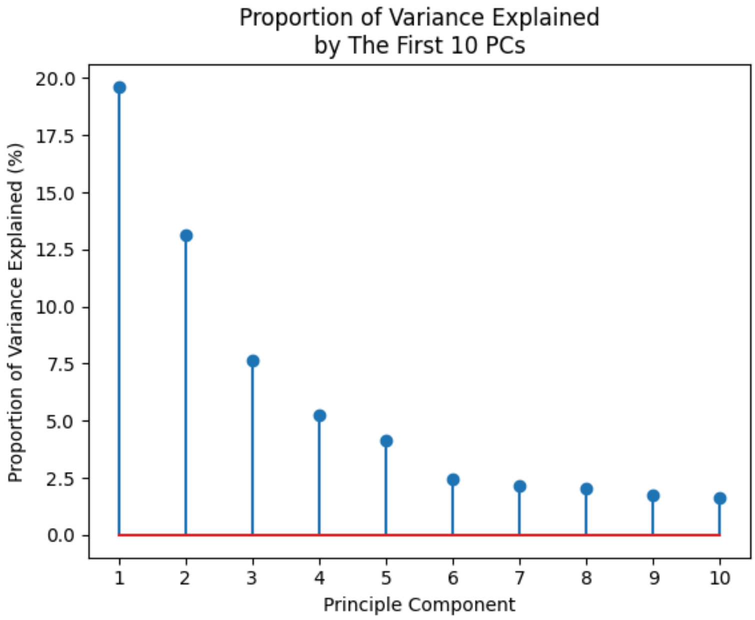

Figure 1: Proportion of variance explained (PVE) by the first 10 principal components.

Q. 1.2

The principal components represent **directions** that contain the most amount of variance. The images of principal components represent the **patterns** that capture the most amount of variance. They can also be understood as the "building blocks" of each individual image.

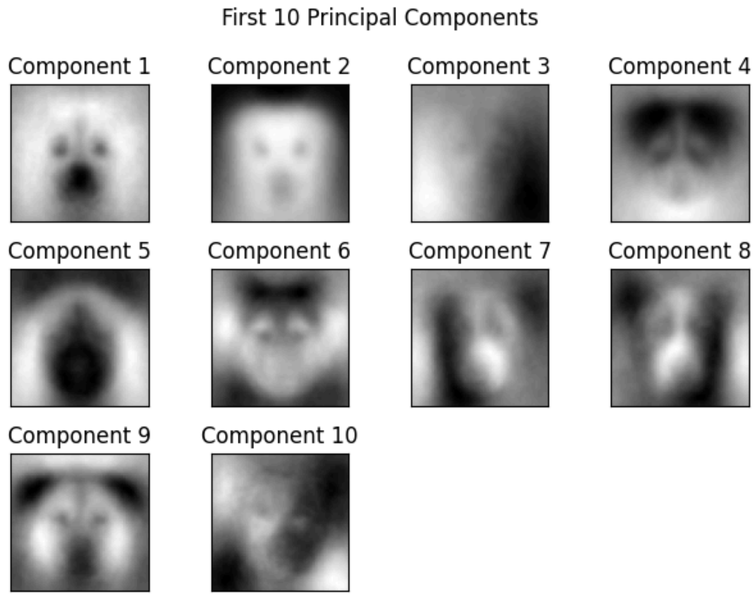

Figure 2: Images of the first 10 PCs.

Q. 1.3

The first step of image compression using PCA is projecting the image to the principal component space. Then, the dot product of the projection and the transpose of the PCs result in the compressed image.

The quality of the compressed image increases as the number of PCs used for compression (k) increases. Since each PC explains a portion of total variance, as more PCs are considered, a larger portion of the total variance is represented. If we use all PCs, we get a lossless reconstruction of the original image.

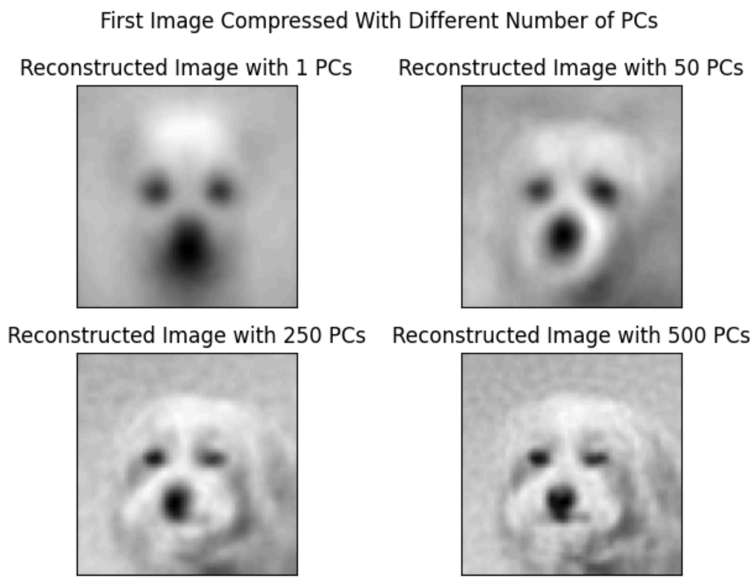

Figure 3: First image of the dataset compressed using various amounts of PCs

Question 2

Q. 2.1

Figure 4 is obtained using batch gradient ascent to maximize the log-likelihood of the logistic function. All weights are initialized to 0. Training is run for 1000 iterations. Learning rates 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} are validated. According to the average F1 scores, the best learning rate is found to be $\mathbf{10^{-4}}$.

Figure 4: F1 score at various learning rates.

The following test set performance metrics belong to the model trained with learning rate 10^{-4}.

		Actual Classes	
		Positive	Negative
Predicted Classes	Positive	**TP:** 48	**FP:** 5
	Negative	**FN:** 2	**TN:** 45

Table 1: Confusion matrix for test set. (full-batch)

- Accuracy: 0.93
- Precision (binary): 0.9056603773584906
- Precision (macro): 0.9315535929345644
- Precision (micro): 0.93
- Recall (binary): 0.96
- Recall (macro): 0.9299999999999999
- Recall (micro): 0.93
- NPV (binary): 0.9574468085106383

- NPV (macro): 0.9315535929345644
- NPV (micro): 0.93
- FPR (binary): 0.1
- FPR (macro): 0.07
- FPR (micro): 0.07
- FDR (binary): 0.09433962264150944
- FDR (macro): 0.06844640706543557
- FDR (micro): 0.07
- F1 (binary): 0.9320388349514563
- F1 (macro): 0.9299369432489241
- F1 (micro): 0.93
- F2 (binary): 0.9486166007905138
- F2 (macro): 0.9297738874397914
- F2 (micro): 0.93

Q. 2.2

In training the mini-batch and stochastic models, the same values were used. All weights are initialized to 0. Training is run for 1000 iterations. Learning rate is set to 10^{-4}. The following test set performance metrics belong to the model trained using mini-batch gradient ascent. The performance metrics are **identical** to the full-batch model and stochastic models.

		Actual Classes	
		Positive	Negative
Predicted Classes	Positive	**TP:** 48	**FP:** 5
	Negative	**FN:** 2	**TN:** 45

Table 2: Confusion matrix for test set. (mini-batch)

- Accuracy: 0.93
- Precision (binary): 0.9056603773584906
- Precision (macro): 0.9315535929345644
- Precision (micro): 0.93
- Recall (binary): 0.96
- Recall (macro): 0.9299999999999999
- Recall (micro): 0.93
- NPV (binary): 0.9574468085106383
- NPV (macro): 0.9315535929345644
- NPV (micro): 0.93
- FPR (binary): 0.1
- FPR (macro): 0.07
- FPR (micro): 0.07
- FDR (binary): 0.09433962264150944
- FDR (macro): 0.06844640706543557
- FDR (micro): 0.07

- F1 (binary): 0.9320388349514563
- F1 (macro): 0.9299369432489241
- F1 (micro): 0.93
- F2 (binary): 0.9486166007905138
- F2 (macro): 0.9297738874397914
- F2 (micro): 0.93

The following test set performance metrics belong to the model trained using stochastic gradient ascent. The performance metrics are **identical** to the full-batch model and mini-batch models.

		Actual Classes	
		Positive	Negative
Predicted Classes	Positive	**TP:** 48	**FP:** 5
	Negative	**FN:** 2	**TN:** 45

Table 3: Confusion matrix for test set. (stochastic)

- Accuracy: 0.93
- Precision (binary): 0.9056603773584906
- Precision (macro): 0.9315535929345644
- Precision (micro): 0.93
- Recall (binary): 0.96
- Recall (macro): 0.9299999999999999
- Recall (micro): 0.93
- NPV (binary): 0.9574468085106383
- NPV (macro): 0.9315535929345644
- NPV (micro): 0.93
- FPR (binary): 0.1
- FPR (macro): 0.07
- FPR (micro): 0.07
- FDR (binary): 0.09433962264150944
- FDR (macro): 0.06844640706543557
- FDR (micro): 0.07
- F1 (binary): 0.9320388349514563
- F1 (macro): 0.9299369432489241
- F1 (micro): 0.93
- F2 (binary): 0.9486166007905138
- F2 (macro): 0.9297738874397914
- F2 (micro): 0.93

Question 3

Q. 3.1

In this part, linear support vector machine (SVM) models with soft margins without any kernels are used. The hyper-parameter C is optimized by selecting the C value that results in the minimum validation error using Monte Carlo cross validation. Possible C values are 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, and 10^{3}.

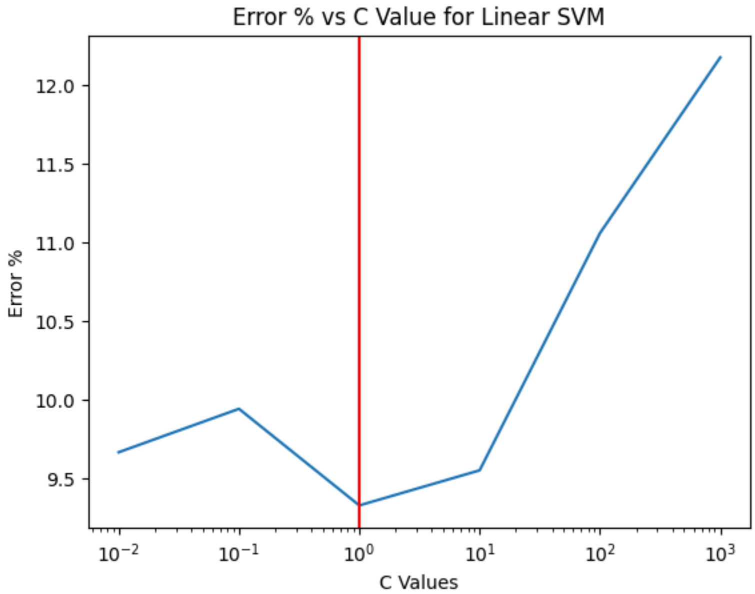

Figure 5: Results of Monte Carlo cross validation.

As seen from figure 5, the most optimal C value is $\mathbf{10^{0}}$. The following test set performance metrics belong to the model trained using 10^{0} as the C value.

		Actual Classes	
		Positive	Negative
Predicted Classes	Positive	**TP:** 45	**FP:** 4
	Negative	**FN:** 5	**TN:** 46

Table 4: Confusion matrix for test set. (linear SVM)

- Accuracy: 0.91
- Precision (binary): 0.9183673469387755
- Precision (macro): 0.9101640656262505
- Precision (micro): 0.91
- Recall (binary): 0.9
- Recall (macro): 0.91
- Recall (micro): 0.91
- NPV (binary): 0.9019607843137255

- NPV (macro): 0.9101640656262505
- NPV (micro): 0.91
- FPR (binary): 0.08
- FPR (macro): 0.09
- FPR (micro): 0.09
- FDR (binary): 0.08163265306122448
- FDR (macro): 0.0898359343737495
- FDR (micro): 0.09
- F1 (binary): 0.9090909090909091
- F1 (macro): 0.90999099909991
- F1 (micro): 0.91
- F2 (binary): 0.9036144578313253
- F2 (macro): 0.9099745595929535
- F2 (micro): 0.91

Q. 3.1

In this part, support vector machine (SVM) models with soft margins and polynomial kernels are used. The hyper-parameters C and d are optimized together by selecting the C and d values that result in the minimum validation error using Monte Carlo cross validation. Possible C values are 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, and 10^{3}. Possible d values are 2, 3, 4, 5.

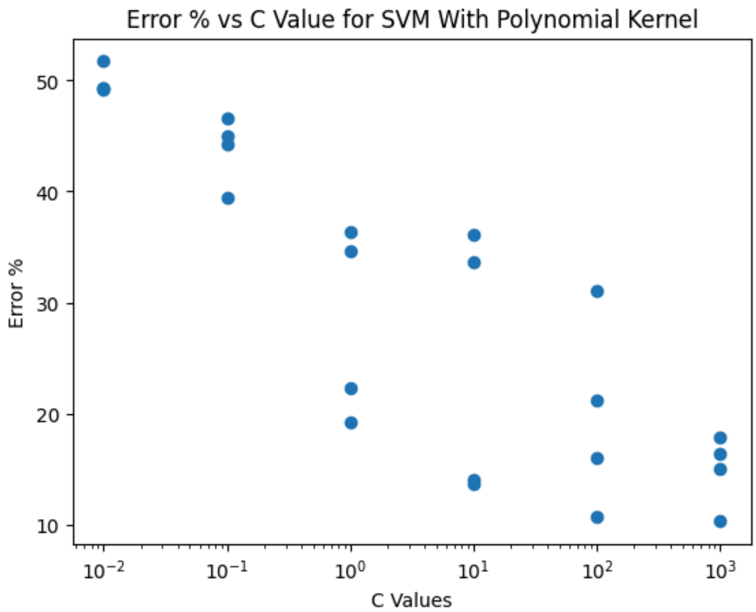

Figure 6: Results of Monte Carlo cross validation for all C and d values indexed by C values.

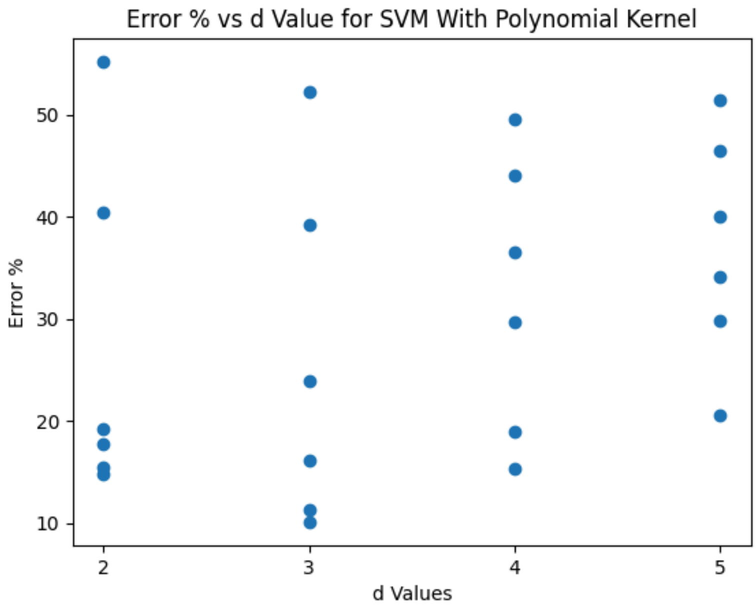

Figure 7: Results of Monte Carlo cross validation for all C and d values indexed by d values.

As seen from figures 6 and 7, the most optimal C and d value combination is $\mathbf{10^3}$ and **3** respectively. The following test set performance metrics belong to the model trained using the optimal C and d values found above.

<table>
<tr><td colspan="2"></td><td colspan="2">Actual Classes</td></tr>
<tr><td colspan="2"></td><td>Positive</td><td>Negative</td></tr>
<tr><td rowspan="2">Predicted Classes</td><td>Positive</td><td>TP: 47</td><td>FP: 6</td></tr>
<tr><td>Negative</td><td>FN: 3</td><td>TN: 44</td></tr>
</table>

Table 4: Confusion matrix for test set. (linear SVM)

- Accuracy: 0.91
- Precision (binary): 0.8867924528301887
- Precision (macro): 0.9114813327980731
- Precision (micro): 0.91
- Recall (binary): 0.94
- Recall (macro): 0.9099999999999999
- Recall (micro): 0.91
- NPV (binary): 0.9361702127659575
- NPV (macro): 0.9114813327980731
- NPV (micro): 0.91
- FPR (binary): 0.12
- FPR (macro): 0.09
- FPR (micro): 0.09
- FDR (binary): 0.11320754716981132

- FDR (macro): 0.08851866720192694
- FDR (micro): 0.09
- F1 (binary): 0.912621359223301
- F1 (macro): 0.9099189270343309
- F1 (micro): 0.91
- F2 (binary): 0.9288537549407114
- F2 (macro): 0.9097710070250116
- F2 (micro): 0.91