Bilkent University

CS 464 Introduction to Machine Learning Course
Fall 2020
Homework 2

Ömer Faruk Akgül
21703163 CS

21.04.2020

Question 1.1

a) Mini-Batch Gradient Ascent with batch-size = 25

Maximum accuracy is obtained when **learning rate = 0.000095**

a.1) Inception v3

Accuracy :	0.80
Precision :	0.74
Recall :	0.93
Negative predictive rate :	0.91
False positive rate:	0.33
False discovery rate:	0.26
f1 score is:	0.82
f2 score is:	0.88

Confusion Matrix

186	65
14	135

According to the results of measurements, the algorithm used for this part of the question is a good predictor of class 1. That means when the data belongs to class 1, it has a higher chance to be correctly classified. That is why recall is really high. However, the algorithm is not good at predicting class 0 data. Thus, it has precision rate that is remarkably lower than accuracy and recall. Moreover, it is more meaningful to trust this algorithm, when it predicts class 0 because it's NPV is 91%. Also, it is clear that f2 score is higher than f1 score because f2 favors the recall rate more.

a.2) Hog

Accuracy :	0.67
Precision :	0.67
Recall :	0.66
Negative predictive rate :	0.66
False positive rate:	0.33
False discovery rate:	0.33
f1 score is:	0.66
f2 score is:	0.66

Confusion Matrix

131	65
69	135

Measurements obtained with Hog Feauters shows consistency because all measurement ways resulted in approximately the same result. That means predictions made upon the features extracted with Hog do not have tendency to show better performance when a test data belongs to a particular class.

When the same mini-batch algorithm is used with the features obtained by Hog, all the values decreased dramatically. It may have two reasons. The first one is that Inception features possess more features than Hog so it fits better to the real data. The second reason might be Hog is not as good as Inception to extract features.

b) Stochastic Gradient Ascent

b.1) Inception v3

Accuracy :	0.87
Precision :	0.83
Recall :	0.93
Negative predictive rate :	0.92
False positive rate:	0.19
False discovery rate:	0.17
f1 score is:	0.88
f2 score is:	0.91

Confusion Matrix

186	38
14	162

Since stochastic gradient ascent algorithm updates weights for each training data, it fits better than mini-batch gradient ascent. Though predictions are better than mini-batch when test data belongs to 0 labeled class, the algorithm is still better predictor of class 1. When it tries to classify a data that actually belongs to class 0, more mistaken decisions are made. F2 score is higher because algorithm favors recall.

b.2) Hog

Accuracy :	0.67
Precision :	0.67
Recall :	0.66
Negative predictive rate :	0.66
False positive rate:	0.33
False discovery rate:	0.33
f1 score is:	0.66
f2 score is:	0.66

Confusion Matrix

131	65
69	135

Although using stochastic gradient ascent algorithm improves accuracy for Inception features, it could not give better results with Hog features than mini-batch algorithm. Confusion matrix is the same so the discussion made for mini-batch with Hog features is valid for stochastic gradient ascent with hog features.

Since stochastic gradient ascent updates values for each training data, it is quiet inefficient in terms of calculation speed.

Question 1.2

a) Batch Gradient Ascent

Though question asks using the same learning rate, using 0.095 gave much more better result.

a.1) Inception

Accuracy :	0.88
Precision :	0.87
Recall :	0.89
Negative predictive rate :	0.89
False positive rate:	0.13
False discovery rate:	0.12
f1 score is:	0.88
f2 score is:	0.89

Confusion Matrix

178	25
22	175

Since all the measurement values are close to each other, it is reasonable to say that any measurement can be used to talk about algorithm's efficiency. Algorithm does not have obvious tendeny to classify one of the classes better. Since the algorithm classified class 1 better at very small amout, it seems recall and so f2 score are higher but there is no considerable difference.

Corresponding Indices of the Greatest Weights

[1870 1995 124 139 346 994 1784 1949 1929 219]

The Most Important Indices

[1170 1435 939 825 1049 787 247 334 1417 231]

Let one of the feature be measured in terms of meter. If it is measured in terms of centimeters, it's weight would be 100 times itself. Therefore, there is not a correlation between weight and its importance as displayed.

a.2) Hog

Accuracy :	0.66
Precision :	0.67
Recall :	0.65
Negative predictive rate :	0.66
False positive rate:	0.33
False discovery rate:	0.34
f1 score is:	0.66
f2 score is:	0.65

Confusion Matrix

129	65
71	135

Corresponding Indices of the Greatest Weights

[5 0 103 322 30 156 59 316 291 106]

The Most Important Indices

[146 281 209 74 226 23 241 88 313 59]

As explained, there is no correlation.

As before, Hog features do not possess enough information to ease the classification when logistic regression is used. Results are approximately the same as previous algorithms.

Since batch gradient ascent algorith updates values once for all training set, it is quite efficient though it's accuray is a bit lower than stochastic gradient descent.

Question 1.4

Soft margin SVM model with linear kernel

a) Inception

$C \in \{10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}\}$
Mean of Accuracy for C = 0.01 is 0.8875
Mean of Accuracy for C = 0.1 is 0.87
Mean of Accuracy for C = 1 is 0.8675
Mean of Accuracy for C = 10 is 0.8675
Mean of Accuracy for C = 100 is 0.8675

Thus, C = 0.01 is chosen.

Confusion Matrix

182	25
18	175

Accuracy: 0.8925
Precision: 0.879
Recall: 0.91

b) Hog

$C \in \{10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}\}$
Mean of Accuracy for C = 0.01 is 0.67
Mean of Accuracy for C = 0.1 is 0.669
Mean of Accuracy for C = 1 is 0.669

Mean of Accuracy for C = 10 is 0.661
Mean of Accuracy for C = 100 is 0.6485

Thus, C = 0.01 is chosen.

Confusion Matrix

115	46
85	154

Accuracy: 0.6725
Precision: 0.7143
Recall: 0.575

Inception features resulted in better prediction performance than Hog. For both extraction method, smallest value of C, 0.01, gave the best accuracy.

Question 1.5

Hard margin SVM with radial basis function (rbf) kernel

a) Inception

$\gamma \in \{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{6}\}$
Mean of Accuracy for $\gamma = 2^{-4}$ is 0.58
Mean of Accuracy for $\gamma = 2^{-3}$ is 0.52
Mean of Accuracy for $\gamma = 2^{-2}$ is 0.51
Mean of Accuracy for $\gamma = 2^{-1}$ is 0.50
Mean of Accuracy for $\gamma = 2^{0}$ is 0.50
Mean of Accuracy for $\gamma = 2^{1}$ is 0.50
Mean of Accuracy for $\gamma = 2^{4}$ is 0.50

Chosen $\gamma = 2^{-4}$

Confusion Matrix

200	171
0	29

Accuracy: 0.57
Precision: 0.54
Recall: 1.0

As seen, when the test data belongs to class 1, it is always classified correctly. However, when the test data belongs to class 0, it is generally misclassified. Thus, it is not reasonable to choose recall as a performance metric for this algorithm.

b) Hog

$\gamma \in \{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{6}\}$
Mean of Accuracy for $\gamma = 2^{-4}$ is 0.68
Mean of Accuracy for $\gamma = 2^{-3}$ is 0.67
Mean of Accuracy for $\gamma = 2^{-2}$ is 0.68
Mean of Accuracy for $\gamma = 2^{-1}$ is 0.70
Mean of Accuracy for $\gamma = 2^{0}$ is 0.71
Mean of Accuracy for $\gamma = 2^{1}$ is 0.69
Mean of Accuracy for $\gamma = 2^{4}$ is 0.50

Chosen $\gamma = 1$

Confusion Matrix

136	58
64	142

Accuracy: 0.69
Precision: 0.71
Recall: 0.64

This is the first time Hog features gave better performance values than Inception features. Moreover, while Inception data predicts class 1 datas better, Hog data does not show such tendency.

Question 1.6

Soft margin SVM with radial basis function (rbf) as kernel

a) Inception

$(C, \gamma) \in \{10^{-2}, 1, 10^{2}\} \times \{2^{-2}, 2^{1}, 2^{6}\}$

$C = 10^{-2}$
$\gamma = 2^{-2}$ accuracy = 0.5
$\gamma = 2$ accuracy = 0.84
$\gamma = 2^{-6}$ accuracy = 0.5

$C = 1$
$\gamma = 2^{-2}$ accuracy = 0.51
$\gamma = 2$ accuracy = 0.50
$\gamma = 2^{-6}$ accuracy = 0.50

$C = 10^{2}$
$\gamma = 2^{-2}$ accuracy = 0.51
$\gamma = 2$ accuracy = 0.50
$\gamma = 2^{-6}$ accuracy = 0.50

C = [0.01] and **γ = [2]** are chosen

Confusion Matrix

170	34
30	166

Accuracy: 0.84
Precision: 0.83
Recall: 0.85

When the optimum parameters are chosen, the model predicts well. Since accuracy, precision and recall are close to each other, any performance metric can be selected.

b) Hog

$(C, \gamma) \in \{10^{-2}, 1, 10^{2}\} \times \{2^{-2}, 2^{1}, 2^{6}\}$

$C = 10^{-2}$

$\gamma = 2^{-2}$ accuracy = 0.66
$\gamma = 2$ accuracy = 0.67
$\gamma = 2^{-6}$ accuracy = 0.51

$C = 1$

$\gamma = 2^{-2}$ accuracy = 0.69
$\gamma = 2$ accuracy = 0.70
$\gamma = 2^{-6}$ accuracy = 0.50

$C = 10^{2}$

$\gamma = 2^{-2}$ accuracy = 0.69
$\gamma = 2$ accuracy = 0.71
$\gamma = 2^{-6}$ accuracy = 0.50

C = 100 and γ = 2 are chosen

Confusion Matrix

147	64
53	136

Accuracy: 0.71
Precision: 0.70
Recall: 0.74

Though Inception features predicted well just for selected C and gamma values, predictions made with Hog features are good for most of the selected c, gamma pairs. However, optimum accuracy obtained with Hog feature is still lower than that of Inception.

Question 1.7

Soft margin SVM with radial basis function(rbf) as kernel and one-vs-all

a) Inception

$(C, \gamma) \in \{10^{-2}, 1, 10^{2}\} \times \{2^{-2}, 2^{1}, 2^{6}\}$

$C = 10^{-2}$

$\gamma = 2^{-2}$	accuracy = 0.11
$\gamma = 2$	accuracy = 0.58
$\gamma = 2^{-6}$	accuracy = 0.10

$C = 1$

$\gamma = 2^{-2}$	accuracy = 0.12
$\gamma = 2$	accuracy = 0.39
$\gamma = 2^{-6}$	accuracy = 0.10

$C = 10^{2}$

$\gamma = 2^{-2}$	accuracy = 0.12
$\gamma = 2$	accuracy = 0.39
$\gamma = 2^{-6}$	accuracy = 0.10

$C = 0.01$ and $\gamma = 2$ are chosen

Confusion matrix

```
[[23  1  5  5  0  0  2  1  0  3]
 [ 0 20  0  2  6  0  0  3  8  1]
 [11  1 13  8  1  1  0  1  1  3]
 [ 4  4  5 21  2  0  1  0  3  0]
 [ 0  6  1  2 25  0  0  0  5  1]
 [ 3  0  0  0  0 32  1  0  1  3]
 [ 1  2  0  0  0  2 27  4  0  4]
 [ 1  2  1  0  0  3  1 22  8  2]
 [ 2 10  0  0  1  1  1  2 21  2]
 [ 3  2  0  2  0  2  1  0  2 28]]
```

Class Based Accuracy: 0.58

Micro Average of Precision:	0.59	Macro Average of Precision:	0.59
Micro Average of Recall:	0.58	Macro Average of Recall:	0.58
Micro Average of F1:	0.58	Macro Average of F1:	0.58

b) Hog

(C, γ) ∈ $\{10^{-2}, 1, 10^{2}\} \times \{2^{-2}, 2^{1}, 2^{6}\}$

C = 10^{-2}

- **γ = 2^{-2}** accuracy = 0.28
- **γ = 2** accuracy = 0.27
- **γ = 2^{-6}** accuracy = 0.18

C = 1

- **γ = 2^{-2}** accuracy = 0.33
- **γ = 2** accuracy = 0.36
- **γ = 2^{-6}** accuracy = 0.11

C = 10^{2}

- **γ = 2^{-2}** accuracy = 0.32
- **γ = 2** accuracy = 0.36
- **γ = 2^{-6}** accuracy = 0.11

C = 100 and γ = 2 are chosen

Confusion Matrix

```
[[21  2  4  2  0  4  3  3  0  1]
 [ 0  8  3  4 14  3  1  0  7  0]
 [ 6  5  5  7  5  2  3  2  4  1]
 [ 4  4  2 11  9  1  3  2  1  3]
 [ 2  6  0  3 20  0  1  2  5  1]
 [ 3  1  1  5  3 16  6  1  0  4]
 [ 2  1  0  6  1  2 20  4  2  2]
 [ 2  1  1  3  2  4  5 12  9  1]
 [ 1  5  3  4  7  1  2  6 11  0]
 [ 0  2  1  1  3  6  0  0  7 20]]
```

Class Based Accuracy: 0.36

Micro Average of Precision:	0.36	Macro Average of Precision:	0.36
Micro Average of Recall:	0.36	Macro Average of Recall:	0.36
Micro Average of F1:	0.36	Macro Average of F1:	0.36

Since each subclass contains equal number of labels, micro and macro averages are the same.

As in previous models, inception results are much more better that Hog results.

Question 1.8

Hard margin SVM with polynomial kernel

a) Inception

$(d, \gamma) \in \{3, 5, 7\} \times \{2^{-2}, 2, 2^{6}\}$

d = 3

- **$\gamma = 2^{-2}$** accuracy = 0.68
- **$\gamma = 2$** accuracy = 0.68
- **$\gamma = 2^{-6}$** accuracy = 0.68

d = 5

- **$\gamma = 2^{-2}$** accuracy = 0.67
- **$\gamma = 2$** accuracy = 0.67
- **$\gamma = 2^{-6}$** accuracy = 0.67

d = 7

- **$\gamma = 2^{-2}$** accuracy = 0.63
- **$\gamma = 2$** accuracy = 0.63
- **$\gamma = 2^{-6}$** accuracy = 0.63

d = 3 and γ = 0.25 are chosen

Confusion Matrix

```
[[31  1  6  2  0  0  0  0  0  0]
 [ 0 30  0  1  2  0  1  1  4  1]
 [10  1 17  8  1  1  1  1  0  0]
 [ 1  0  6 28  3  0  1  0  1  0]
 [ 0  4  0  1 31  0  0  2  2  0]
 [ 0  0  0  0  0 35  2  1  2  0]
 [ 0  1  2  0  0  1 31  3  0  2]
 [ 0  0  2  2  0  0  2 33  1  0]
 [ 0  3  1  0  3  0  1  4 24  4]
 [ 1  1  2  0  0  2  1  0  2 31]]
```

Class Based Accuracy: 0.73

Micro Average of Precision:	0.73	Macro Average of Precision:	0.73
Micro Average of Recall:	0.73	Macro Average of Recall:	0.73
Micro Average of F1:	0.73	Macro Average of F1:	0.73

b) Hog

$(d, \gamma) \in \{3, 5, 7\} \times \{2^{-2}, 2, 2^{6}\}$

d = 3

$\gamma = 2^{-2}$ accuracy = 0.35
$\gamma = 2$ accuracy = 0.34
$\gamma = 2^{-6}$ accuracy = 0.34

d = 5

$\gamma = 2^{-2}$ accuracy = 0.37
$\gamma = 2$ accuracy = 0.35
$\gamma = 2^{-6}$ accuracy = 0.35

d = 7

$\gamma = 2^{-2}$ accuracy = 0.37
$\gamma = 2$ accuracy = 0.36
$\gamma = 2^{-6}$ accuracy = 0.36

Confusion Matrix

```
[[18  2  5  4  0  4  2  3  0  2]
 [ 1 11  2  4 12  2  2  1  5  0]
 [ 6  5  8  4  6  2  3  2  3  1]
 [ 4  4  3 10  9  1  3  2  1  3]
 [ 3  6  0  3 20  0  0  2  3  3]
 [ 6  1  1  4  3 15  5  1  0  4]
 [ 3  1  0  6  2  1 20  3  2  2]
 [ 4  1  3  3  2  3  5 10  8  1]
 [ 1  5  3  3  6  2  2  6 12  0]
 [ 0  1  1  1  3  6  0  3  4 21]]
```

Class Based Accuracy: 0.36
Micro Average of Precision: 0.36
Micro Average of Recall: 0.36
Micro Average of F1: 0.36
Macro Average of Precision: 0.36
Macro Average of Recall: 0.36
Macro Average of F1: 0.36

Question 1.9

• Which feature extraction method yields better results?

Features extracted using the Inception yields better results for both subclass and superclass classification. However, there is an exception that features extracted with HOG yields better results, which is hard margin svm with radial basis function. Also, for some pairs of C and gamma values HOG yields more consistent results than Inception(i.e accuracy yielded by using Inception was so sensitive to C and gamma values).

• Which model and feature combination performed best/worst?

For superclass classification, **soft margin SVM model with linear kernel** and **Inception** combination performed best. (Accuracy: 0.8925, Precision: 0.879, Recall: 0.91)

For superclass classification, **hard margin SVM with radial basis function (rbf) kernel** and **Inception** combination performed worst. (Accuracy: 0.57, Precision: 0.54, Recall: 1.0)
(There are some combinations that performed worse in terms of recall metric but the worst is thought as overall performance.)

For subclass classification, **hard margin SVM with polynomial kernel** and **Inception** combination performed best.(73% of the all subclass labels are predicted correctly).

For subclass classification, **soft margin SVM with radial basis function(rbf) as kernel and one-vs-all** and **HOG** combination performed worst.(only 36% of the all subclass labels are predicted correctly).

The effect of C on the decision boundary of SVM.

A smaller value of C allows to ignore points close to boundary and increase the margin, while a large value of C a large penalty is assigned to errors and margin errors. The large value of C causes data to overfit training samples. Therefore, small values of C yielded better test accuracies.
For Inception features, the accuracy falls from 89% to 86% as C increases.
For Hog features, the accuracy falls from 67% to 64% as C increases.

The effect of γ on the decision boundary of SVM with RBF kernel.

γ can be thought of as the ‘spread’ of the kernel and therefore the decision region. When gamma is low, the ‘curve’ of the decision boundary is very low and thus the decision region is very broad. When gamma is high, the ‘curve’ of the decision boundary is high, which creates islands of decision-boundaries around data points. Therefore, high values of gamma cause overfitting to the training set. Whereas, very low values underfit. Thus, optimal value should be found on validation set and then testing should be done.
For Inception features, increasing the value of gamma caused accuracy to fall from 58% to 50%. Thus, it is obvious that model overfits to the training data.
For Hog features, accuracy flactuates between 68% and 71% as gamma goes from 2^{-4} to 2^{1}. However, when gamma is 2^{4}, accuracy falls to 50% that shows the overfitting of model to training data.

The effect of d on the decision boundary of SVM with polynomial kernel.

When d is low, the degree of the decision boundary is not enough to represent the boundary and thus misclassification is high. When d is high, the degree of the decision boundary is more sensitive to each training data, which creates islands of decision-boundaries around data points. Therefore, high values of d cause overfitting to the training set. Whereas, very low values underfit. Thus, optimal value should be found on validation set and then testing should be done.
For Inception features, increasing the value of d caused accuracy to fall from 68% to 63%. Thus, it is obvious that model overfits to the training data as d increases.
For Hog features, increasing the value of d caused accuracy to rise from 34% to 37%. Thus, it is obvious that model improves as d increases.

The effects of different (C, γ) pairs on the decision boundary and tolerance of SVM.

For Inception, γ = 2 always yields the best result, that means the other values of γ either underfits or overfits. However, increasing C causes accuracy to fall so overfits. Choosing the optimal (C, γ) pair yields the best result.

For HOG, γ = 2 yields the best result, except while C = 10^{-2}. However, increasing C causes accuracy to fall so the model overfits the training data.

The effects of different (d, γ) pairs on the decision boundary of SVM with polynomial kernel.

For Inception, changing the value of gamma did not have considerable effect on the value of accuracy. However, increasing the value of d resulted in decrease of accuracy so higher d means overfitting to the training data.(from 0.68 to 0.63)

For HOG, changing the value of gamma did not have considerable effect on the value of accuracy(more or less 1%). Although increment in the value of d causes an improvement in accuracy, it is not more than 2%. Thus, given (d, γ) values do not have huge effect on the accuracy value.

The effect of batch size to performance and training time of logistic regression models.

Small batch size gets weights "close" to the minimum much faster than the higher batch sizes. Thus, stochastic gradient ascent yielded the highest accuracy. However, when the batch size is small, more update is done, which means the longer training time. Thus, stochastic gradient ascent model consumed more time than mini batch, which is more than full batch.

The use of γ parameter for polynomial kernel.

γ is essential to scale the Gaussians.

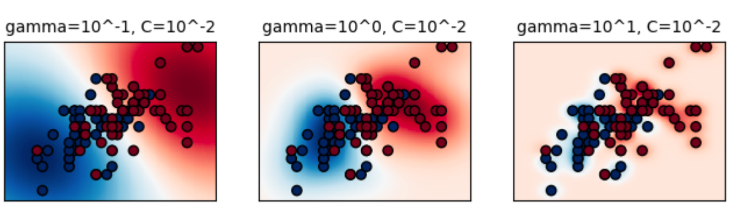

Figure 1

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close'. The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. If gamma is too large, the radius of the area of influence of the support vectors only includes the support vector itself. When gamma is very small, the model is too constrained and cannot capture the complexity or “shape” of the data[1].

What are the advantages and disadvantages of logistic regression and SVM models compared to each other?

SVM is a more complex model (non-linear model) than logistic regression (linear model). It may provide more accuracy, but may suffer from overfitting.

SVM will not work well for large amount of data (say a million data points). But logistic regression is fine (logistic regression will suffer more on number of features/columns instead of number of rows).

SVM has kernel methods which can classify features by mapping data in higher dimensions using orthogonal projections and RBF kernels.

Logistic regression is more sensitive to outliers, hence SVM performs better in presence of outliers.

Question 1.10

In what cases NPV, FPR, FDR and F1 score would be more informative compared to accuracy, precision, recall alone?

If most of the data belongs to negative class, predicting all new data as negative yields a high accuracy. However, NPV is a better metric for the given situation.

If the distribution of data is like given below, predicting most of the new data as positive yields a high accuracy. However, most of the negative data is classified incorrectly. Thus, FPR can be considered.

90	10
3	1

If the distribution of data is like given below, most of the new data as negative yields a high accuracy. However, if prediction is positive, it is incorrect with 50%. Thus, FDR can be considered.

10	10
3	100

If the distribution of data is like given below, recall is 100%. However, if prediction is positive, it is probably incorrect. Thus, F1 score can be considered, which is 0.29.

10	50
0	40

Accuracy is a reliable metric for subclass and superclass classification tasks because there are equal number of labels. However, some models have a tendency to predict one of the classes more correctly. For such cases, other metrics should be examined, as well.

Question 2.1

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

[2]

Total time for SVD based PCA is **35.39 seconds**. When images are reconstructed using svd, MSE is calculated as **0.0509299** using the equation above. SVD based implementation is much more faster than that of covariance based implementation. Though it is faster, it yields the same MSE.

Question 2.2

Total time for covariance matrix based implementation of PCA is **166.31 seconds**. When images are reconstructed using all PC's MSE is calculated as **0.0509299**. As seen result is quite close to the that of SVD. However, implementation is much more slower. Also, to calculate MSE, complex numbers need to be converted real numbers by splitting their imaginary parts. Thus, it is more reasonable to use SVD.

Question 2.3

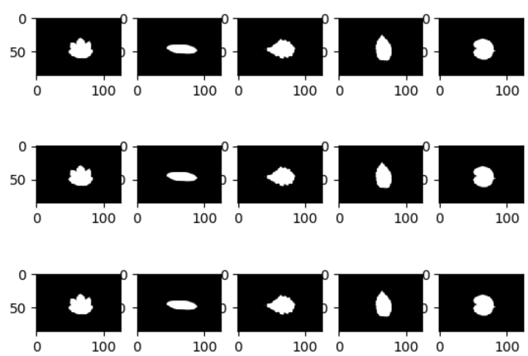

Figure 2

Since both methods have similar MSE, constructed images are nearly the same. However, covariance based implentation consumes 4 times more time that that of SVD. Also, covariance based implementation uses complex numbers, that causes error when converted to real number.

MSE is not normalized. So, other methods like SSIM and FSIM can be treated more understandable than the MSE. This is due to the fact that MSE is absolute error, however, SSIM and FSIM are giving perception and saliency-based errors. If noise level is increasing, then the recovery quality of output image is also deteriorating. So, we can conclude that SSIM and SSIM are comparatively better than MSE and PSNR metrics from human visual perspective.

Question 2.4

Expected running times

Slowest: (iii)n samples << n features .
Middle: (ii) n samples $\sim$ n features
Fastest: (i) n samples >> n features

I used the dataset in the previous question to see the effects of feature size on running time. Decreasing the number of features yields a shorter time. The reason is that when there are more features, covariance matrix enlarges and more eigenvalues and more eigenvectors are calculated(for covariance based implementation). This increases the running time.

REFERENCES

[1] “RBF SVM parameters¶,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html. [Accessed: 03-May-2020].

[2] freeCodeCamp.org, “Machine learning: an introduction to mean squared error and regression lines,” *freeCodeCamp.org*, 08-Oct-2018. [Online]. Available: https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/. [Accessed: 03-May-2020].