Selected Topics in Computer Science – 1 P R O J E C T (C O V E R S H E E T)

Team Number : 17

Team Information :

	ID [Ordered by ID]	Full Name [In Arabic]	Grade
1	202000135	اسماء عربي محمد عبدالمجيد	
2	202000218	بيشوى عصام فاروق	
3	202000236	چمان سميرصلاح الدين محمد خطاب	
4	202000587	عمر احمد عبدالرحيم محمد	
5	202000685	ماجد ممدوح محمد حلمي	
6	202001050	يارا محمد احمد عبداللطيف	
7			

Table of Contents

Logistic Regression on a numerical dataset

a) General Information on dataset :

1) Name of dataset used :

- Health care:Heart attack possibility.

2) Number of classes and their labels :

- Number of classes : 13 columns , 302 rows.
- Number of labels : 1 column (Target column) , 302 rows.
- In this case we have a binary classification.

3) The total number of samples in dataset :

- 303 and after dropping the duplicated data it becomes 302.

4) Number of samples used in training, validation, and testing:

- Training : 180
- Validation : 61
- Testing : 61

B) Implementation details:

Hyperparameters used in the model :

- Using the gradient function we reduce the cost function from (0.6931471805599453) to (0.3482652701775918) .
- When we bulit a model to get the loss curve we had:
  1. No. epochs = 100
  2. Batch size = 64
  3. Optimizer : adam

C) Results details:

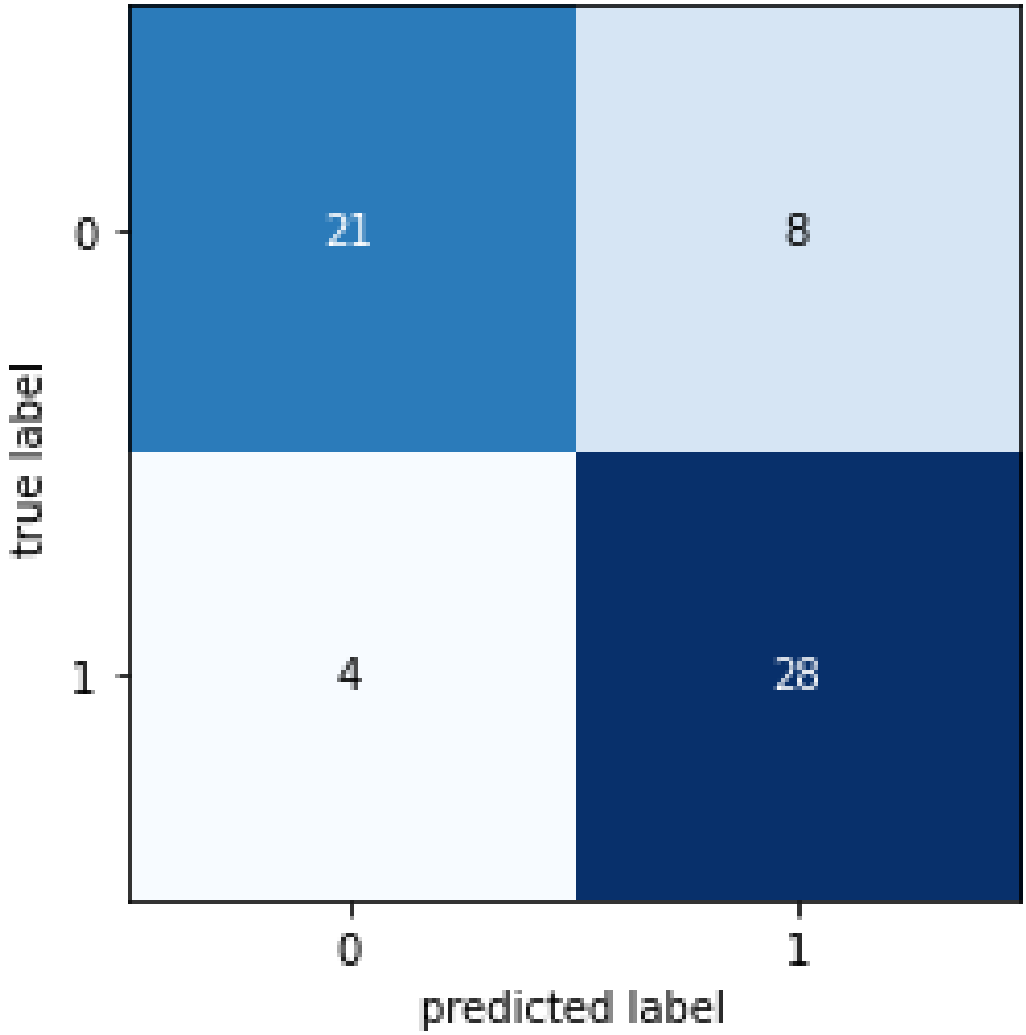
1) Accuracy :

- 84.76821192052981%

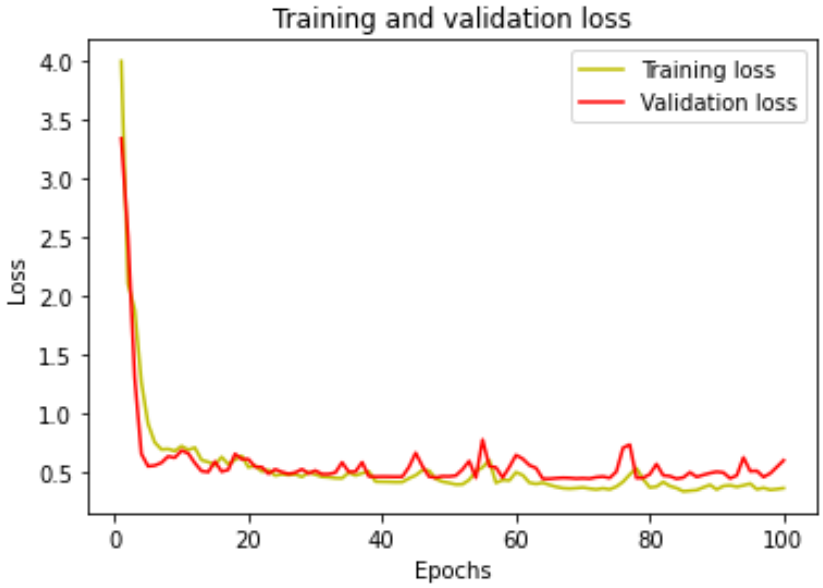
2) Confusion Matrix :

- [[21 8]
  [4 28]]

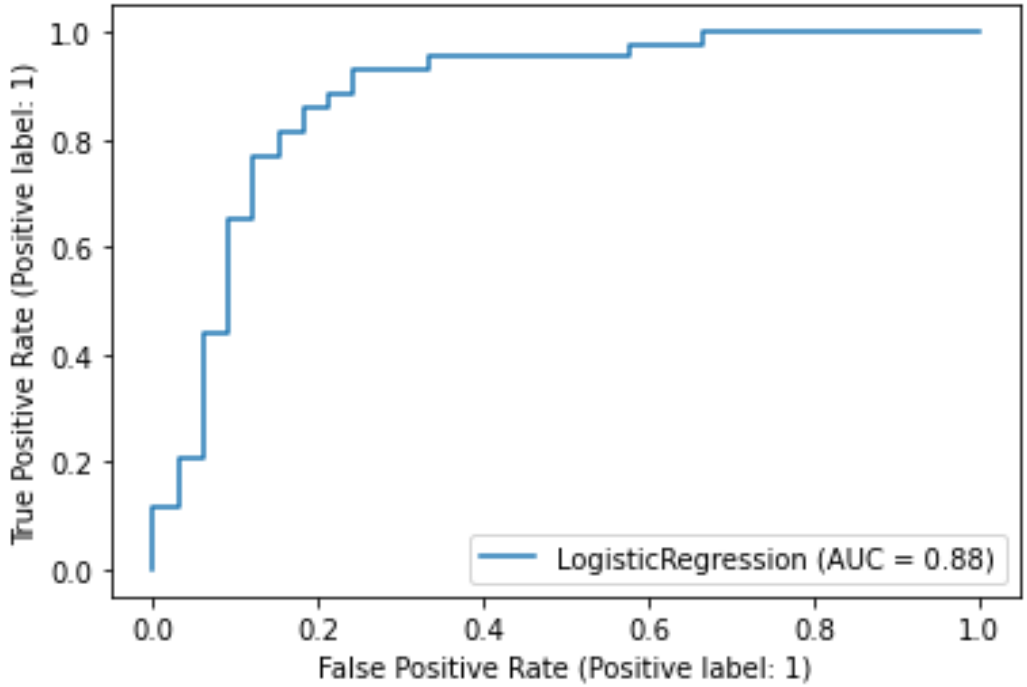
- Plot of confusion matrix



3) Loss Curve



4) ROC Curve



ANN on images dataset

a) General Information on dataset :

1) Name of dataset used :

- Fruits 360

2) Number of classes and their labels :

- Number of classes : 38 folder , 19,000.
- Number of labels : 19,000.
- In this case we have a binary classification.

3) The total number of samples in dataset and its size :

- 25,356 samples
- 100 x 100

4) Number of samples used in training, validation, and testing:

- Training : 19,000
- Testing : 6,356

B) Implementation details:

1) Feature extraction phase :

- *We resized the training and testing images to be 45 x 45*
- *To normalize the array that store the pixels to divided it over 255 to have data between 0's and 1's*

2) Hyperparameters used in the model :

- When we bulit a model we had:
  1. No. epochs = 10.
  2. Optimizer : adam.
  3. 2 hidden layer one has 128 neurons and the other has 64 neurons.
  4. Output layer with 60 neurons.

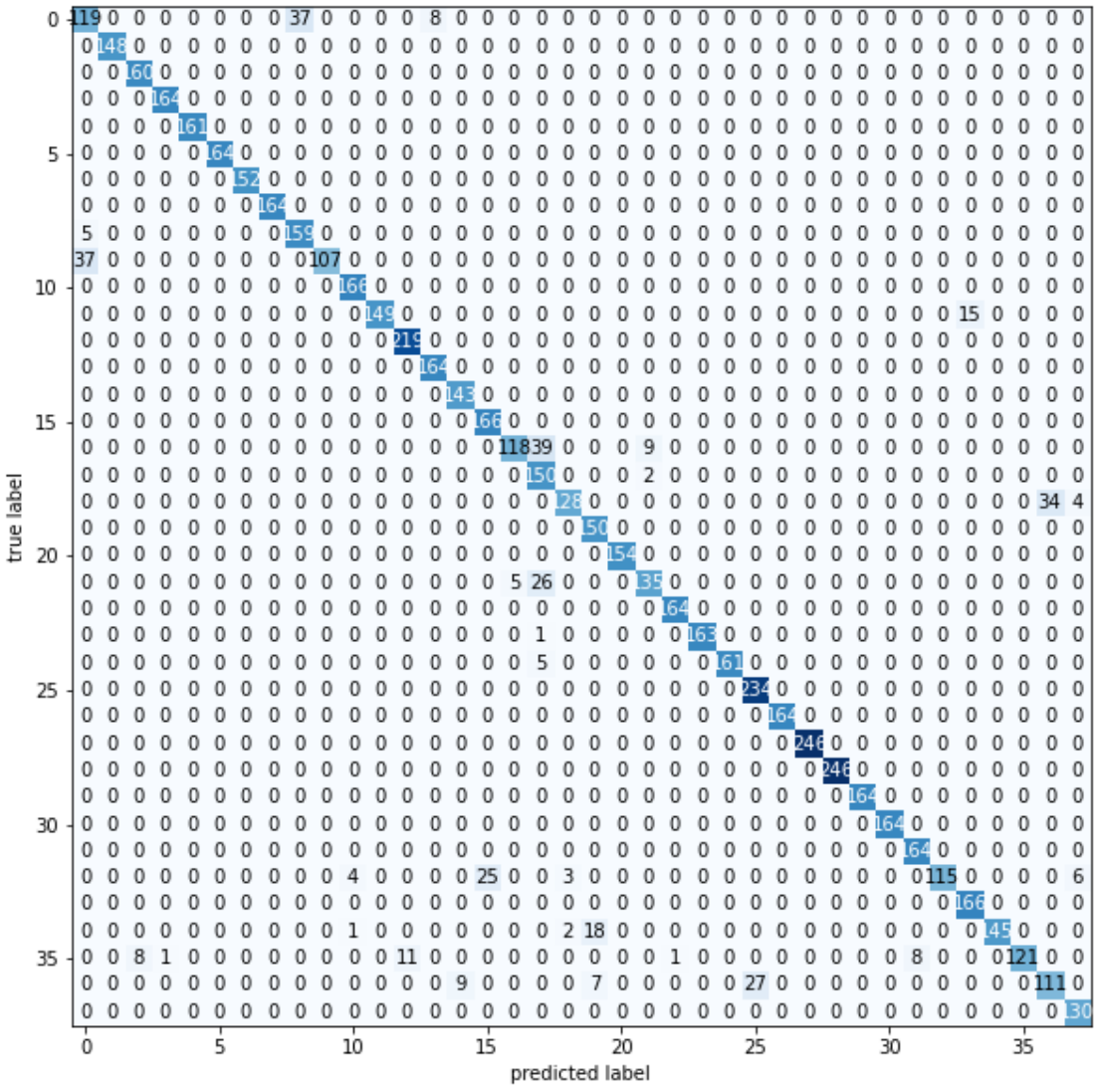
C) Results details:

1) Accuracy :

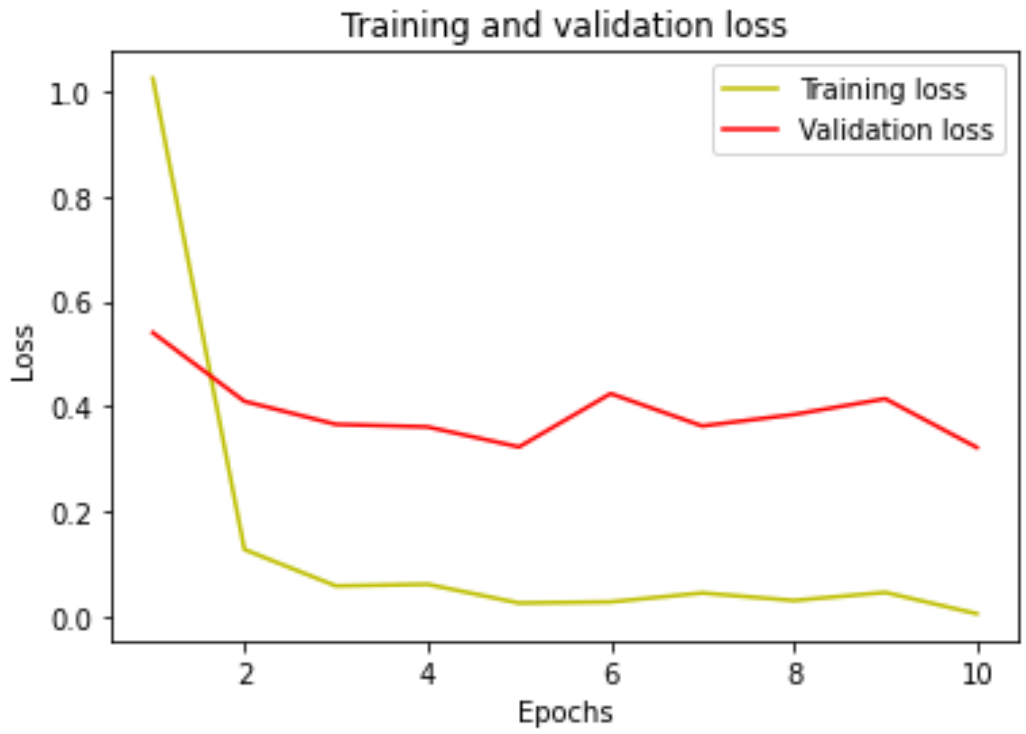
- Test loss : 32.20229%
- Test Accuracy : 94.36752%

2) Confusion Matrix :

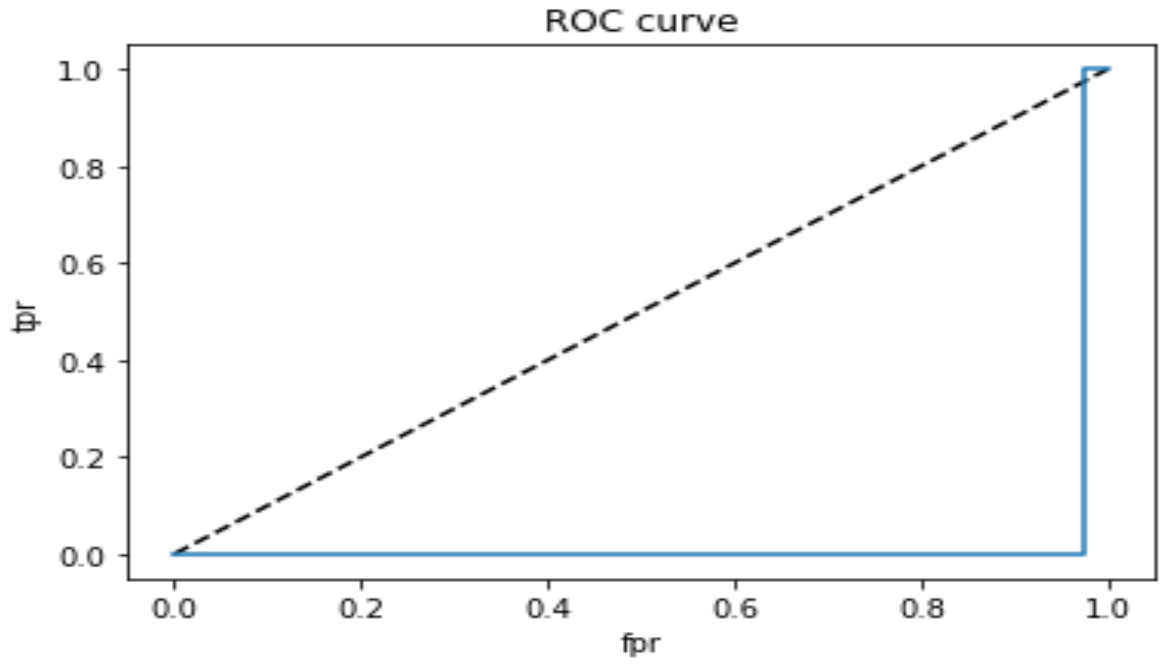
- [[119 0 0 ... 0 0 0]
  [0 148 0 ... 0 0 0]
  [0 0 160 ... 0 0 0]
  ...
  [0 0 8 ... 121 0 0]
  [0 0 0 ... 0 111 0]
  [0 0 0 ... 0 0 130]]
- Plot of confusion matrix



3) Loss Curve



4) ROC Curve



SVM on numerical dataset

a) General Information on dataset :

1) Name of dataset used :

- Health care:Heart attack possibility.

2) Number of classes and their labels :

- Number of classes : 13 columns , 303 rows.
- Number of labels : 1 column (Target column) , 303 rows.
- In this case we have a binary classification.

3) The total number of samples in dataset :

- 303 Samples.

4) Number of samples used in training, validation, and testing:

- Training : 241
- Testing : 61

B) Implementation details:

Hyperparameters used in the model :

- When we bulit a model to get the loss curve we had:
    1. No. epochs = 10
    2. Optimizer : adam
    3. Kernel : linear

C) Results details:

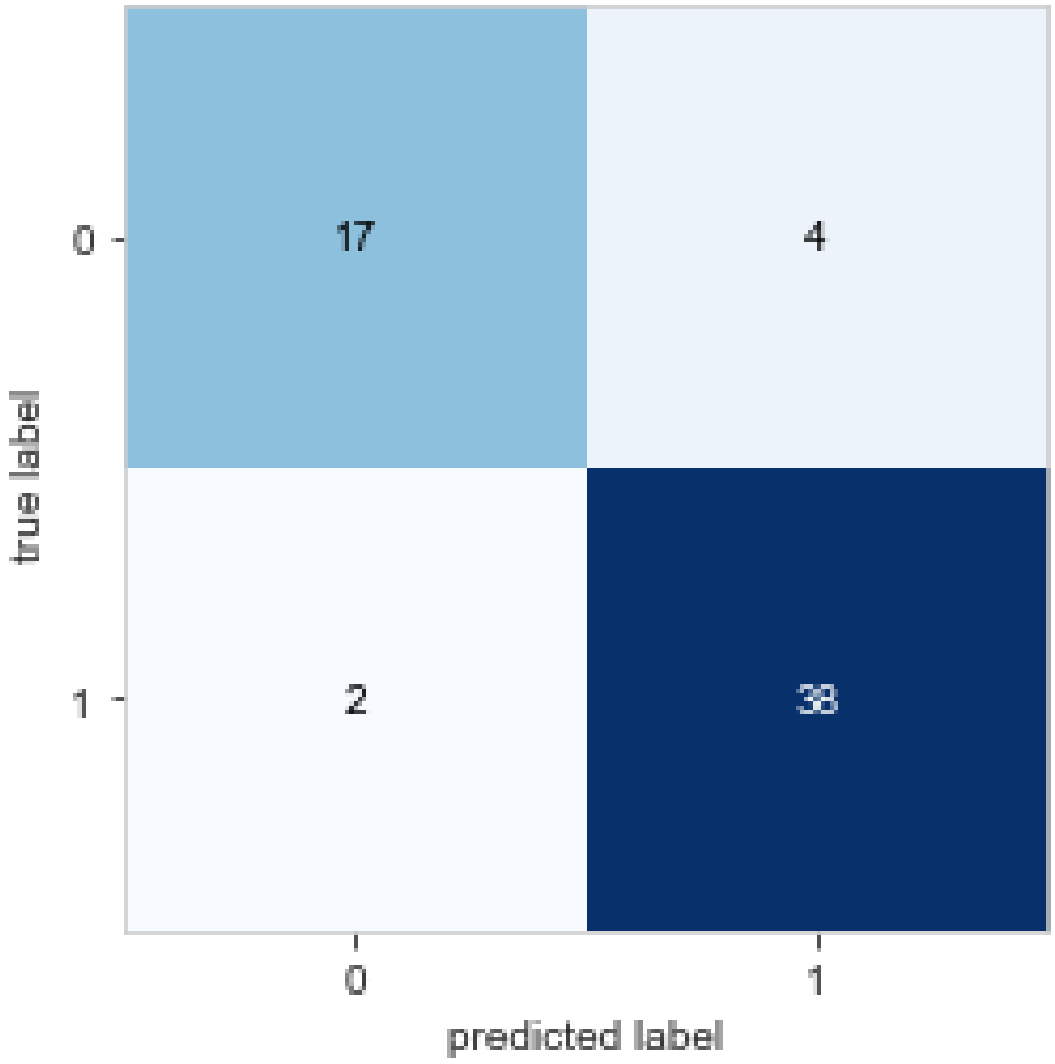
1) Accuracy :

- 90.1639344262295 %

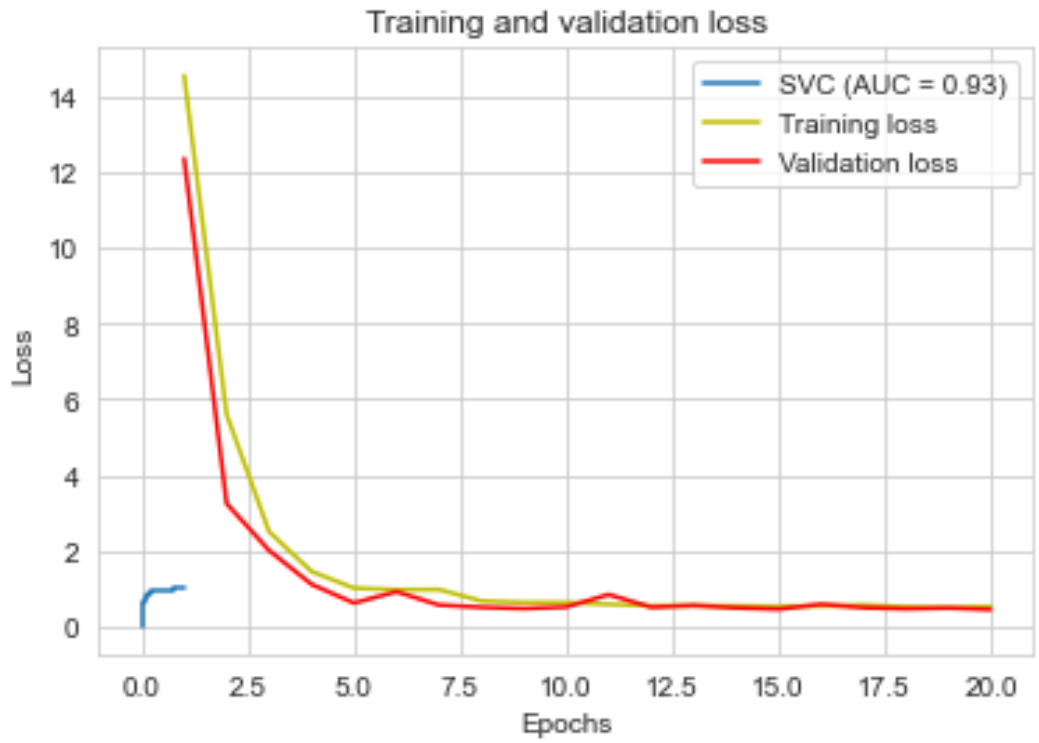
2) Confusion Matrix :

- [[17 4]
  [2 38]]

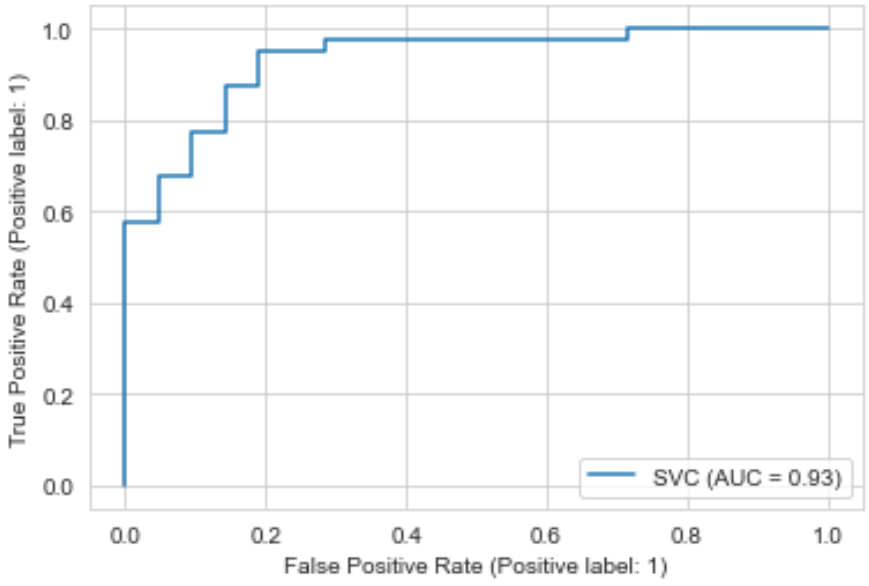
- Plot of confusion matrix



3) Loss Curve



4) ROC Curve



SVM on images dataset

a) General Information on dataset :

1) Name of dataset used :

- Fruits 360

2) Number of classes and their labels :

- Number of classes : 38 folder , 19,000.
- Number of labels : 19,000.
- In this case we have a binary classification.

3) The total number of samples in dataset and its size :

- 25,356 samples
- 100 x 100

4) Number of samples used in training, validation, and testing:

- Training : 19,000
- Testing : 6,356

B) Implementation details:

1) Feature extraction phase :

- *We resized the training and testing images to be 45 x 45*
- *To normalize the array that store the pixels to divided it over 255 to have data between 0's and 1's*

Hyperparameters used in the model :

- When we bulit a model to get the loss curve we had:
    4. No. epochs = 20
    5. Optimizer : adam
    6. Kernel : linear

C) Results details:

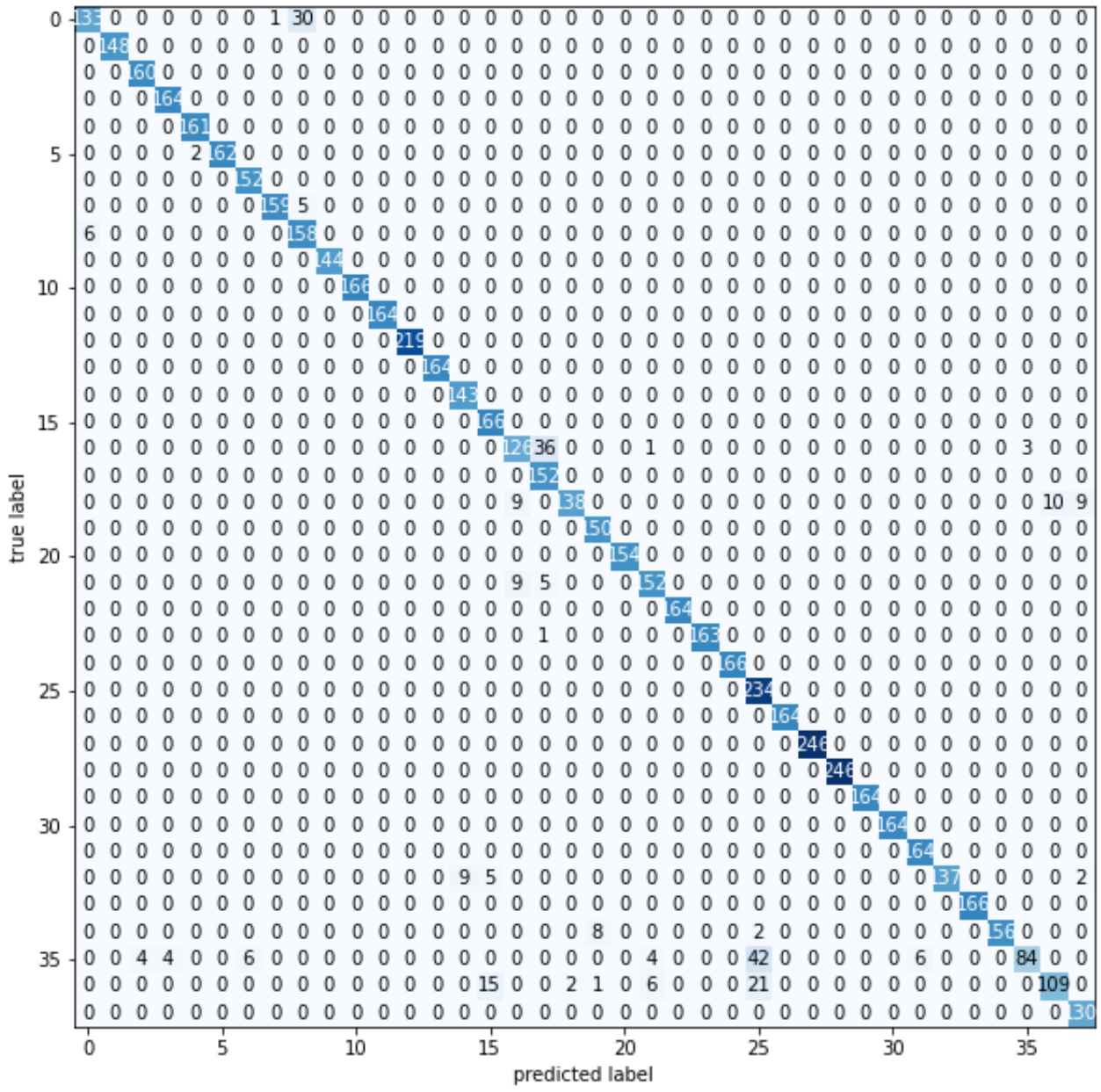
1) Accuracy :

- *95.846644%*

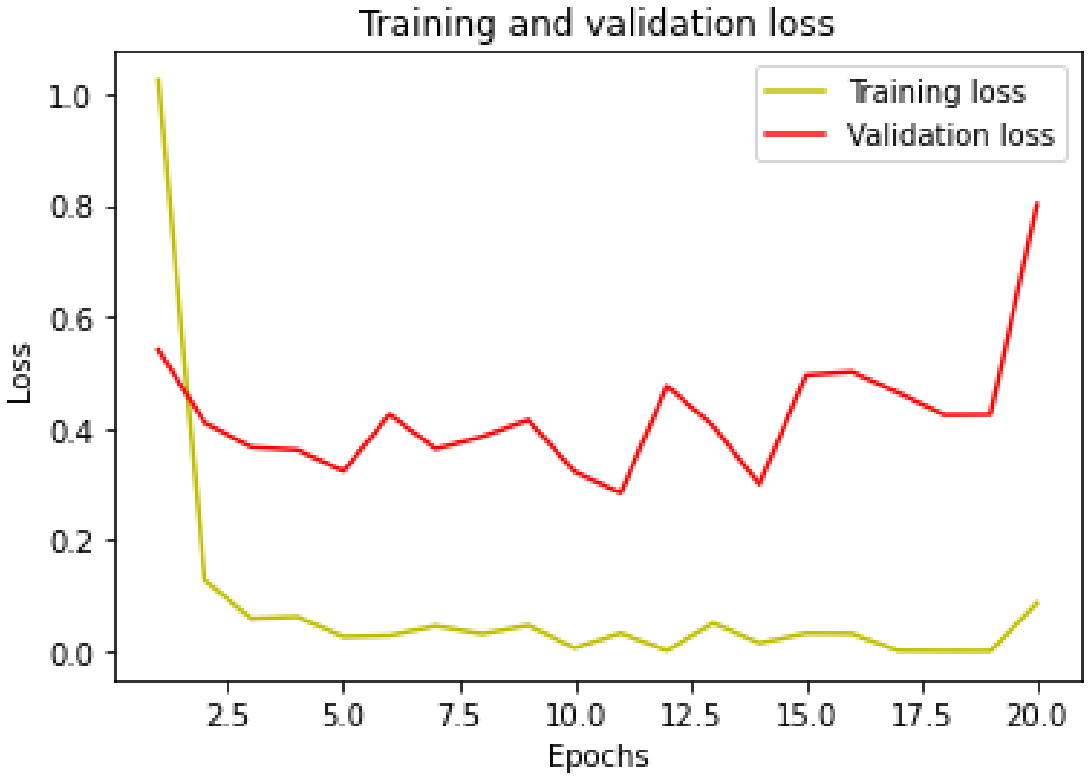
2) Confusion Matrix :

- *[[133 0 0 … 0 0 0]*
  [0 148 0 … 0 0 0]
  [0 0 160 … 0 0 0]
  …
  [0 0 4 … 84 0 0]
  [0 0 0 … 0 109 0]
  [0 0 0 … 0 0 130]]

- Plot of confusion matrix



3) Loss Curve



4) ROC Curve