

A2: Classification with SVM, BP and MLR

Report

Sergi Salido Cubero

sergi.salido@estudiants.urv.cat

Contents

Description of the implementation (languages, tools used, etc.)	3
Execution instructions.....	3
SVM execution	3
BP execution	3
MLR execution	3
Implementation decisions	3
Description and link to the selected dataset	3
Comments on cross-validation and results.....	4
Classification results, including plots	4
SVM, ring dataset.....	4
SVM, bank dataset	5
SVM, banknote dataset.....	5
BP, ring dataset	6
BP, bank dataset	7
BP, banknote dataset.....	7
MLR, ring dataset	7
MLR, bank dataset	8
MLR, banknote dataset.....	8
Evaluation of the classifications.....	9
SVM, ring dataset.....	9
Cross validation error.....	9

Test error.....	9
SVM, bank dataset	9
Cross validation error.....	9
Test error.....	9
SVM, banknote dataset.....	9
Cross validation error.....	9
Test error.....	9
BP, ring dataset	10
Cross validation error.....	10
Test error.....	10
BP, bank dataset	10
Cross validation error.....	10
Test error.....	10
BP, banknote dataset.....	10
Cross validation error.....	10
Test error.....	11
MLR, ring dataset	11
Cross validation error.....	11
Test error.....	11
MLR, bank dataset	11
Cross validation error.....	11
Test error.....	11
Percentage classification error obtained from test set: 10.679611650485436%	11
MLR, banknote dataset.....	11
Cross validation error.....	11
Test error.....	12
Discussion and interpretation of the results	12
Results by dataset	12
SVM results	12
BP results	12
MLR results	12

Description of the implementation (languages, tools used, etc.)

To code all the implementations I have used **Jupyter Notebooks** (IPYNB) and **Python**.

The main Python libraries used are:

- For SVM implementation I have used SVC from sklearn library.
- For BP implementation I have used TensorFlow/Keras libraries.
- For MLR implementation I have used LinearRegression from sklearn library.

I have used VS Code as IDE on Windows, and Google Colab to implement and execute BP.

Execution instructions

SVM execution

Open the folder in VS Code on Windows.

Run **classification_SVM_ring.ipynb** file.

Run **classification_SVM_bank.ipynb** file.

Run **classification_SVM_banknote.ipynb** file.

BP execution

Open the files in Google Colab.

Run **classification_BP_ring.ipynb** file.

Run **classification_BP_bank.ipynb** file.

Run **classification_BP_banknote.ipynb** file.

MLR execution

Open the folder in VS Code on Windows.

Run **classification_MLR_ring.ipynb** file.

Run **classification_MLR_bank.ipynb** file.

Run **classification_MLR_banknote.ipynb** file.

Implementation decisions

I followed all the recommendations from the Dr.

More explanations, details and the references can be found on the notebooks.

Description and link to the selected dataset

I have selected the **banknote authentication** dataset from [UCI Machine Learning Repository](#).

Description: Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes.

Data Set Characteristics:	Multivariate	Number of Instances:	1372	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	5	Date Donated	2013-04-16
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	375990

Data Set Information: Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tool were used to extract features from images.

Attribute Information:

1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)
5. class (integer)

Link: <https://archive.ics.uci.edu/ml/datasets/banknote+authentication>

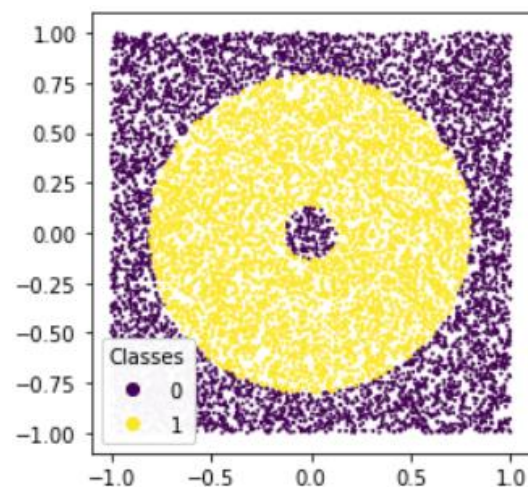
Comments on cross-validation and results

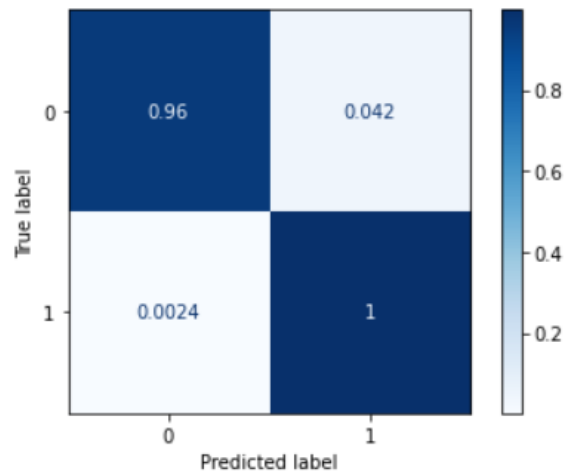
Method used: K-fold

Number of folds used: 4

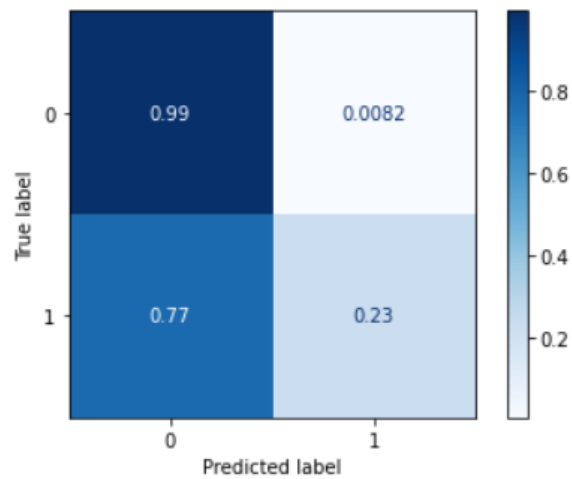
Classification results, including plots

SVM, ring dataset

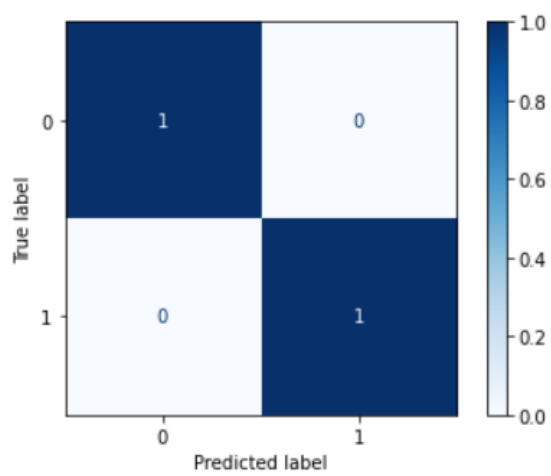




SVM, bank dataset

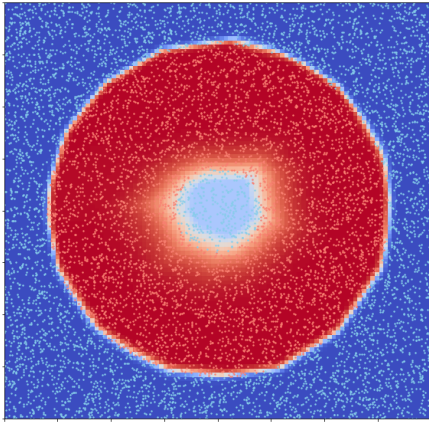


SVM, banknote dataset

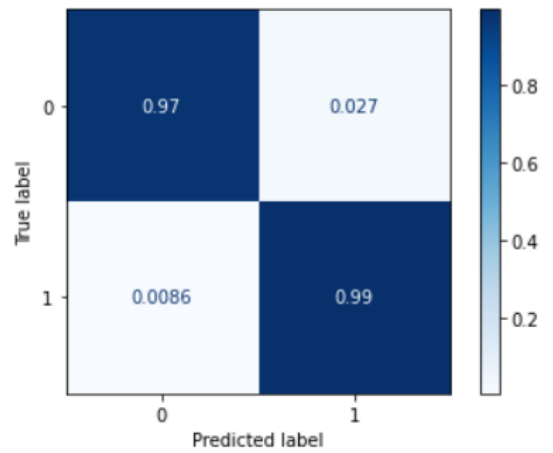
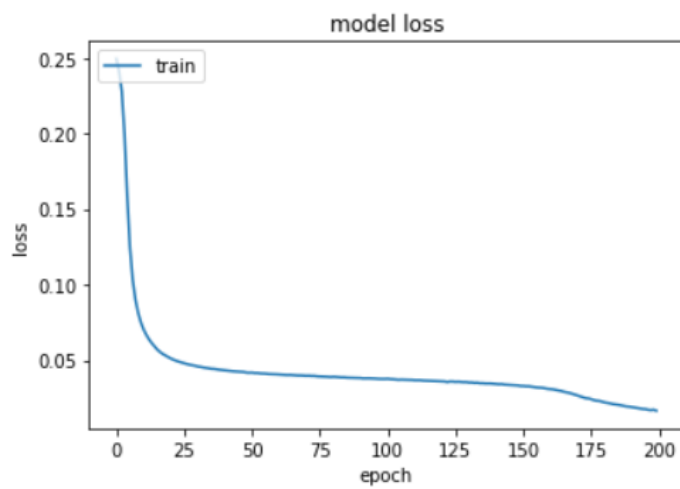
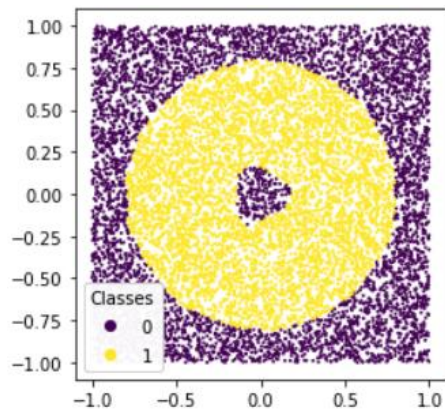


BP, ring dataset

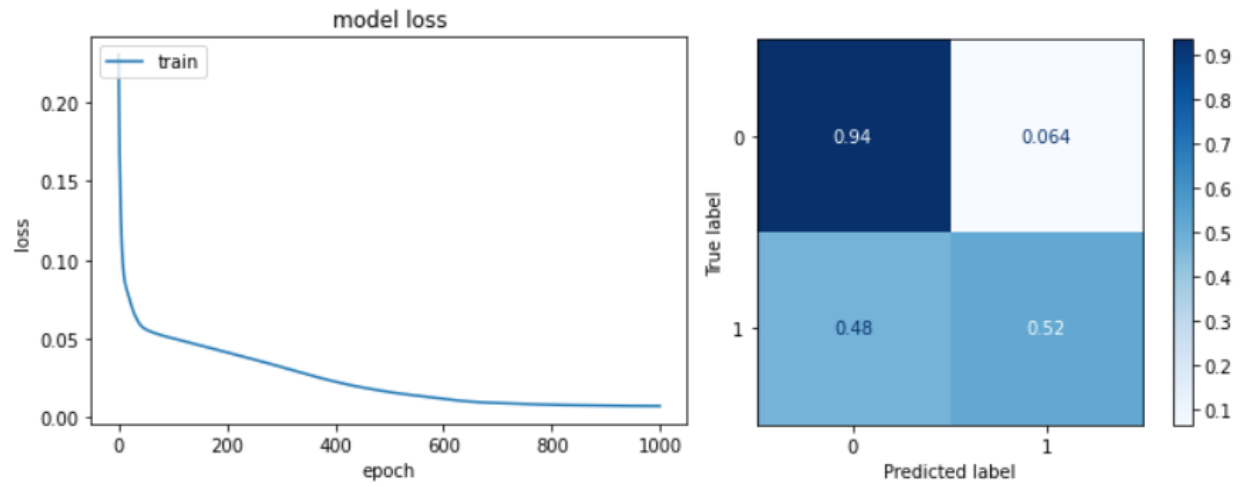
Tensorflow



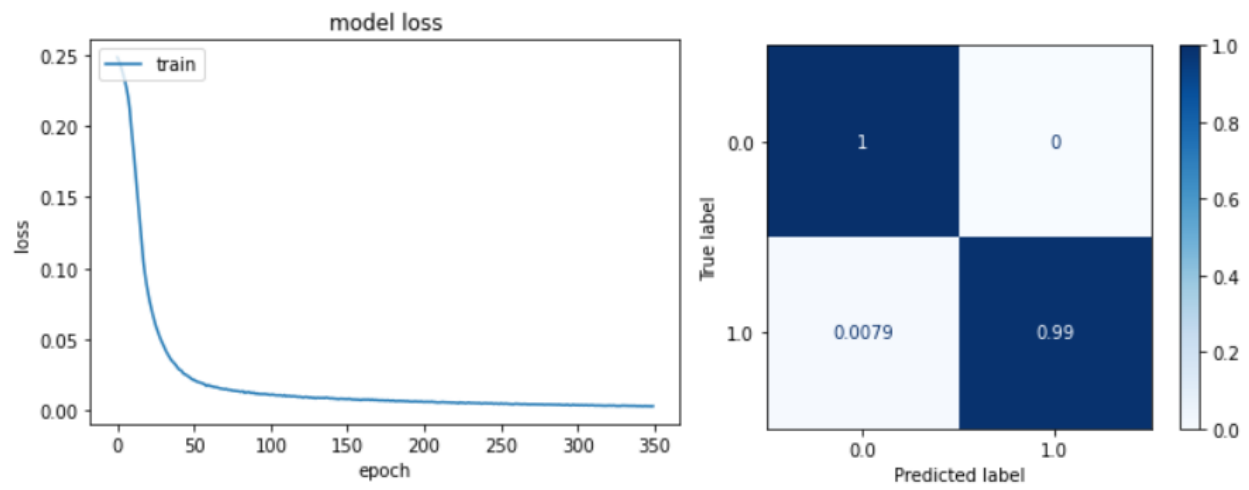
Keras



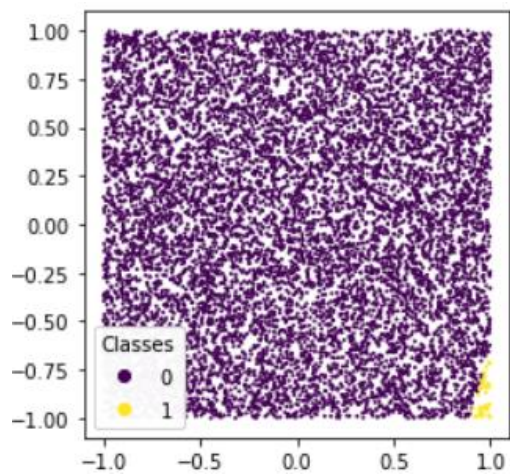
BP, bank dataset

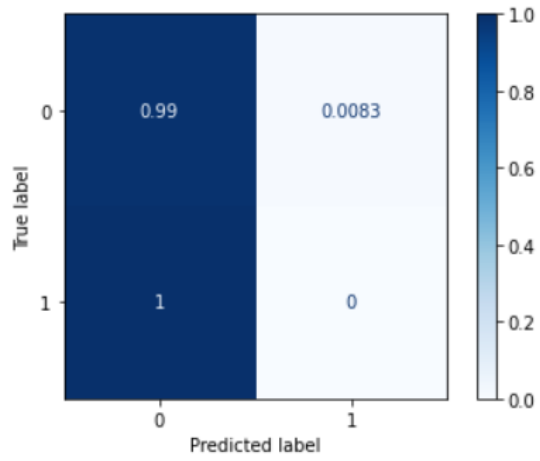


BP, banknote dataset

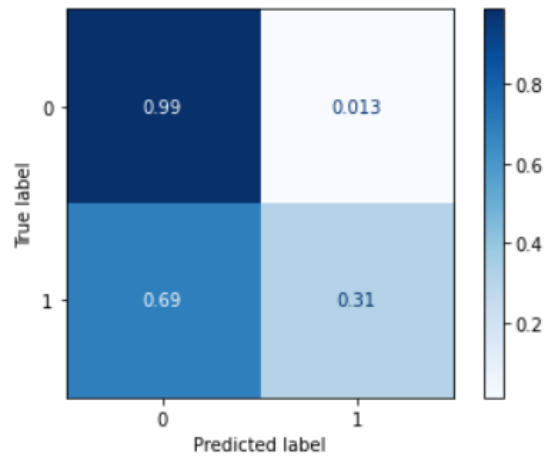


MLR, ring dataset

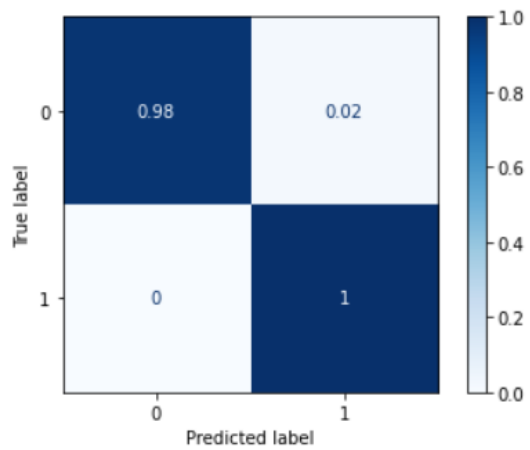




MLR, bank dataset



MLR, banknote dataset



Evaluation of the classifications

SVM, ring dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 2.76%
- Fold 2 - 2.6%
- Fold 3 - 3.1199999999999997%
- Fold 4 - 2.92%

Mean percentage classification error obtained from cross validation: 2.8499999999999996%

Test error

Percentage classification error obtained from test set: 2.36%

SVM, bank dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 9.344660194174757%
- Fold 2 - 10.558252427184465%
- Fold 3 - 10.436893203883495%
- Fold 4 - 9.842041312272174%

Mean percentage classification error obtained from cross validation: 10.045461784378722%

Test error

Percentage classification error obtained from test set: 9.223300970873787%

SVM, banknote dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 0.0%
- Fold 2 - 0.0%
- Fold 3 - 0.0%
- Fold 4 - 0.0%

Mean percentage classification error obtained from cross validation: 0.0%

Test error

Percentage classification error obtained from test set: 0.0%

BP, ring dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 1.72%
- Fold 2 - 3.56%
- Fold 3 - 2.2399999999999998%
- Fold 4 - 1.92%

Mean percentage classification error obtained from cross validation:

- 2.36% (+- 0.7172168430816442)

Test error

Percentage classification error obtained from test set: 1.8499999999999999%

BP, bank dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 10.800970873786406%
- Fold 2 - 10.315533980582524%
- Fold 3 - 10.194174757281553%
- Fold 4 - 9.842041312272174%

Mean percentage classification error obtained from cross validation:

- 10.288180230980664% (+- 0.34335820757562435)

Test error

Percentage classification error obtained from test set: 11.04368932038835%

BP, banknote dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 1.8181818181818181%
- Fold 2 - 0.0%
- Fold 3 - 0.0%
- Fold 4 - 0.7299270072992701%

Mean percentage classification error obtained from cross validation:

- 0.637027206370272% (+- 0.7442049303330632)

Test error

Percentage classification error obtained from test set: 0.36363636363636365%

MLR, ring dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 - 55.52%
- Fold 2 - 47.4%
- Fold 3 - 54.16%
- Fold 4 - 49.16%

Mean percentage classification error obtained from cross validation:

- 51.559999999999995% (+- 3.372951230006151)

Test error

Percentage classification error obtained from test set: 47.11%

MLR, bank dataset

Cross validation error

Results of cross validation:

- Fold 1: 0.3338306
- Fold 2: 0.28591392
- Fold 3: 0.33778509
- Fold 4: 0.3485118

Mean of CV scores:

- 0.3265103518321649

Test error

Percentage classification error obtained from test set: 10.679611650485436%

MLR, banknote dataset

Cross validation error

Results of cross validation:

Percentage classification error obtained from validation set per fold

- Fold 1 – 13.853471%
- Fold 2 – 14.54795%
- Fold 3 – 12.270476%
- Fold 4 – 14.246801%

Mean percentage classification error obtained from cross validation:

➤ 13.729674558289073%

Test error

Percentage classification error obtained from test set: 1.090909090909091%

Discussion and interpretation of the results

Results by dataset

Test error	Ring	Bank	Banknote authentication
SVM	2.36%	9.22%	0.0%
BP	1.85%	11.04%	0.36%
MLR	51.56%	10.68%	1.09%

SVM results

SVM classification results for ring dataset are **very good**.

SVM classification results for bank dataset are **regular**.

SVM classification results for banknote authentication dataset are **very good**.

BP results

BP classification results for ring dataset are **very good**.

BP classification results for bank dataset are **regular**.

BP classification results for banknote authentication dataset are **very good**.

MLR results

MLR classification results for ring dataset are **very bad** because the problem is not linearly separable.

MLR classification results for bank dataset are **regular**.

MLR classification results for banknote authentication dataset are **very good**.