Testing Multiple Models on the Iris Dataset

Master in Artificial Intelligence
Machine Learning I 2022-2023

Guijas Bravo, Pedro
Miguélez Millos, Ángel

November 7, 2022

1 Introduction

This report aims to list the results obtained from the training of multiple variants of machine learning models on the famous Iris dataset[1] for the Unit 6 of the Machine Learning I course.

Some combinations of the most important hyperparameters from each type of model have been tested, but a deeper analysis would be necessary to fully achieve a good comprehension of the advantages and disadvantages of each model. The metric selected is the accuracy, but as a cross-validation technique was applied ($k = 10$), the mean and the standard deviation of the accuracy of the models are presented.

In addition, a confusion matrix is also shown, whose model corresponds to the best combination of hyperparameters of the model, in this case trained and tested with the whole dataset. If there are multiple combinations with the highest mean accuracy, then the one with lowest standard deviation is selected. If there is still a tie, then the simplest one is chosen, which corresponds to a higher row on the table. The rows of the matrix correspond to the real classes, while the columns correspond to the predictions.

2 Artificial Neural Networks

For the Artificial Neural Network (ANN), one of the most important hyperparameters to test is the number of hidden layers of the network. In the Figure 1 we show the results obtained for each topology $(n_0, n_1, ..., n_n)$ where n_i is the number of neurons in the i^{th} hidden layer. Note that the number of inputs and outputs is omitted since it is dependent on the problem to tackle and remains the same for all the networks. They are all trained for a maximum of 1000 epochs, with a learning rate of 0.01, early stopping if the accuracy in the validation set (10% of the training dataset) is not improved for 25 epochs, and the results are averaged for each topology by training it 50 times.

Topology	Mean(accuracy)	Std. deviation(accuracy)
(2)	0.9623	0.0333
(4)	0.9640	0.0519
(8)	**0.9647**	0.0344
(2,2)	0.7760	0.0815
(4,2)	0.7796	0.1212
(4,4)	0.8721	0.0980
(8,4)	0.9125	0.0849
(8,8)	0.9476	0.0525

Table 1: Accuracy results for different ANNs configurations.

[1] https://archive.ics.uci.edu/ml/datasets/iris

The confusion matrix for the best ANN configuration, with 8 neurons in one hidden layer, is shown in the Figure 2.

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-versicolor	0	49	1
Iris-virginica	0	1	49

Table 2: Confusion matrix for the best ANN configuration.

3 Support Vector Machines

For the Support Vector Machine (SVM) models, three different kernels have been tested: a linear one, a polynomial with degree 2 or 3 and a radial basis function one. Each one receives a different hyperparameter, as shown in the Figure 3.

Kernel	C	Degree	Gamma	Mean(accuracy)	Std. deviation(accuracy)
linear	0.1	-	-	0.9600	0.0344
linear	**1**	-	-	**0.9667**	0.0351
linear	10	-	-	0.9600	0.0344
polynomial	0.1	2	-	0.8533	0.1288
polynomial	1	2	-	0.8867	0.0996
polynomial	10	2	-	0.8867	0.0632
polynomial	0.1	3	-	0.9600	0.0344
polynomial	1	3	-	0.9600	0.0344
polynomial	10	3	-	0.9133	0.0773
rbf	0.1	-	0.5	0.9400	0.0734
rbf	0.1	-	2	0.9267	0.0584
rbf	0.1	-	8	0.7800	0.1045
rbf	1	-	0.5	0.9600	0.0344
rbf	1	-	2	0.9467	0.0526
rbf	1	-	8	0.9133	0.0632
rbf	10	-	0.5	0.9533	0.0450
rbf	10	-	2	0.9400	0.0492
rbf	10	-	8	0.9200	0.0526

Table 3: Accuracy results for different SVMs configurations.

The confusion matrix for the best SVM configuration, with a linear kernel and $C = 1$, and the best among all the types of models in this experiment, is shown in the Figure 4.

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-versicolor	0	46	4
Iris-virginica	0	1	49

Table 4: Confusion matrix for the best SVM configuration.

4 Decision Trees

The hyperparameter tested for the Decision Trees models is the maximum depth of the tree. The results are shown in the Figure 5.

The confusion matrix for the best Decision Tree configuration, with a max depth equal to 4, is shown in the Figure 6.

Max. depth	Mean(accuracy)	Std. deviation(accuracy)
1	0.6667	0.0000
2	0.9333	0.0444
4	**0.9400**	0.0378
8	0.9400	0.0378
16	0.9400	0.0378

Table 5: Accuracy results for different Decision Trees configurations.

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-versicolor	0	50	0
Iris-virginica	0	1	49

Table 6: Confusion matrix for the best Decision Tree configuration.

5 k-Nearest Neighbours

Finally, for the k-Nearest Neighbour (kNN) models, only the number of neighbours to take into account for classification (k) is tested, as shown in the Figure 7.

k	Mean(accuracy)	Std. deviation(accuracy)
1	0.9533	0.0450
3	0.9467	0.0526
5	**0.9600**	0.0344
7	0.9600	0.0344
9	0.9600	0.0344

Table 7: Accuracy results for different kNNs configurations.

The confusion matrix for the best kNN configuration, with $k = 5$, is shown in the Figure 8.

	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	50	0	0
Iris-versicolor	0	47	3
Iris-virginica	0	4	46

Table 8: Confusion matrix for the best kNN configuration.