Support Vector Machines & Neural Networks

Santiago Velasco

February 22, 2025

This report condenses the results obtained after analyzed the performance of Support Vector Machines (SVMs) and Neural Networks in a non-linearly separable binary classification problem. In specific the classes 5 and 7 from Fashion-MNIST (dataset usually used as benchmark to measure the performance of several machine learning models) were used to perform the classification. To cut down computational time, the training and testing sets were reduced to 6000 and 2400 examples respectively.

First: Linear SVMs.

In this part a logarithmically range of 10 values for the regularization parameter (C) were selected following the formula $C_n = C_o \times \beta^n \ where \ C_o = 0.001, \beta = 4$. First, various Support vector classification (SVC) instances with linear kernel were trained on the noisy training set using the different C values, then its score was measured using the validation set.

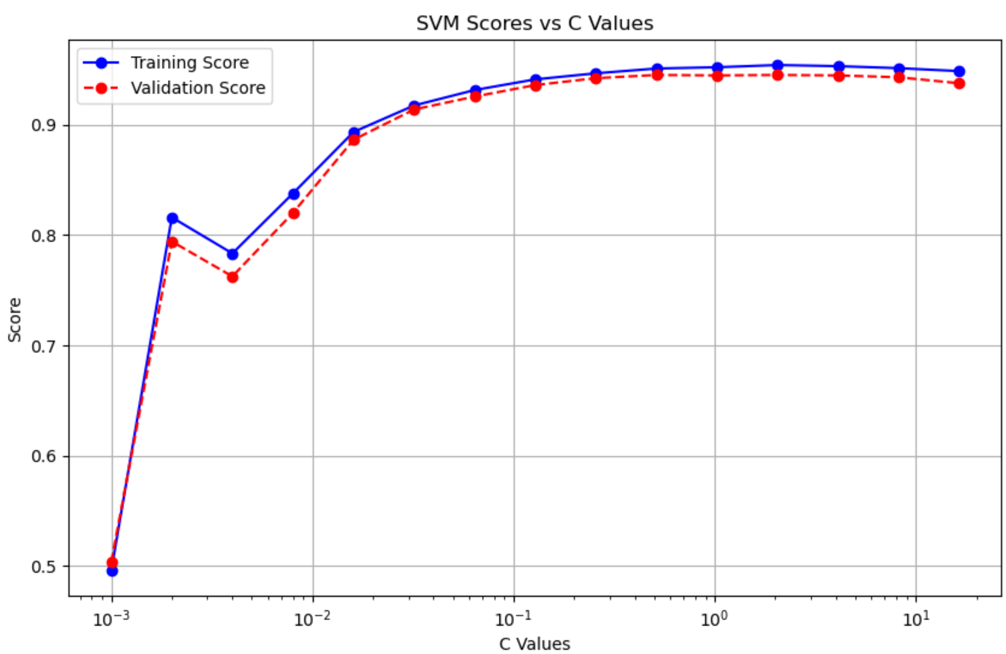

Figure 01. Linear SVM accuracy scores for different values of the regularization parameter C.

The above figure shows a very similar trend line for logarithmically increasing values of C. Showing that there is not much performance loss when linear SVMs are tested on new data, however it is worth mentioning that the best score is achieved when the C value is either 1.024 or 4.096, those values got an score of 0.952 and 0.953 on the training set respectively and the same score in the validation set, 0.9445833333333333.

Later, the regularization parameter values with the best performance were selected (0.512, 1.024, 2.048, 4.096, 8.192) and trained using the k-fold cross-validation technique.

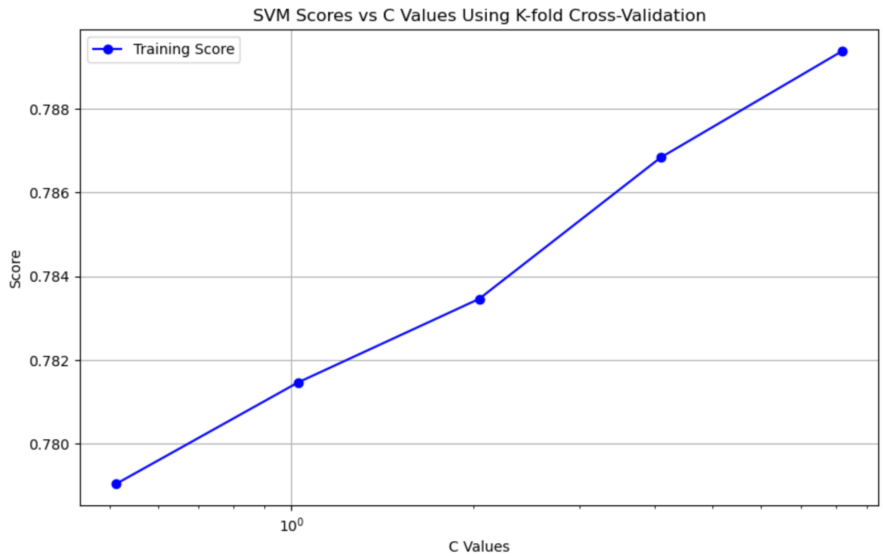

Figure 02. Linear SVM accuracy scores for the best 5 regularization parameter values.

This figure shows an increasing trendline of the performance for higher C values, however the differences between scores are almost neglectable.

Finally, a list of models with regularization parameters values of 0.512, 1.024, 2.048, 4.096, 8.192, 16.384 were trained using the full training set (test + validation) and tested on test set.

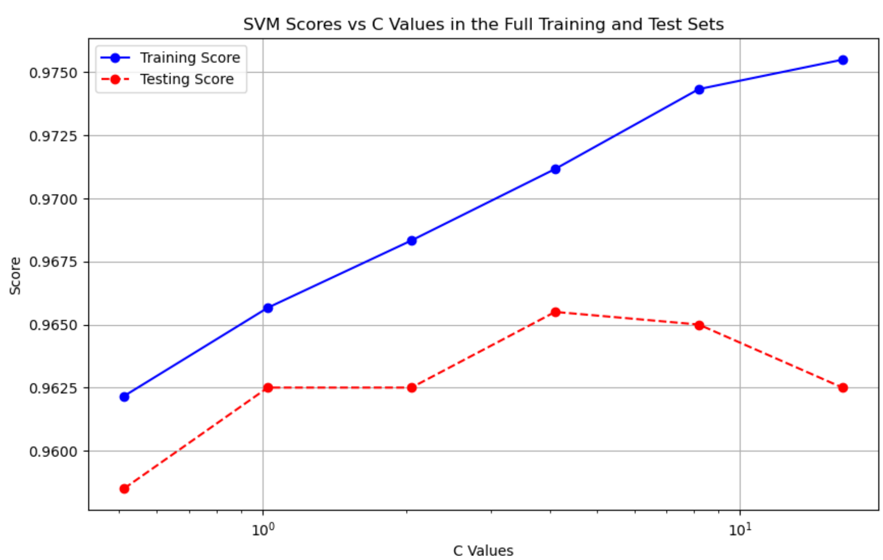

Figure 03. Linear SVM for the best 6 regularization parameter values

Figure 03 shows an increasing trend line similar to Figure 02 in the training set, however it is interesting that the best score in the testing set corresponds to the C value of 4.096.

Gaussian SVMs.

In this section SVC instances were still being used, however its kernel was changed to “rbf” which stands for radial basis function or also known as Gaussian. This is a huge change in the way the models approach classification, a linear kernel only considers the cosine similarity between points, while gaussian computes the similarity of two point in an infinite-dimensional space to build the support vectors.

In that context, the following regularization parameter values were tested using Kfold Cross-Validation 0.001, 0.004, 0.016, 0.064, 0.256, 1.024, 4.096, 16.384, 65.536, 262.144.

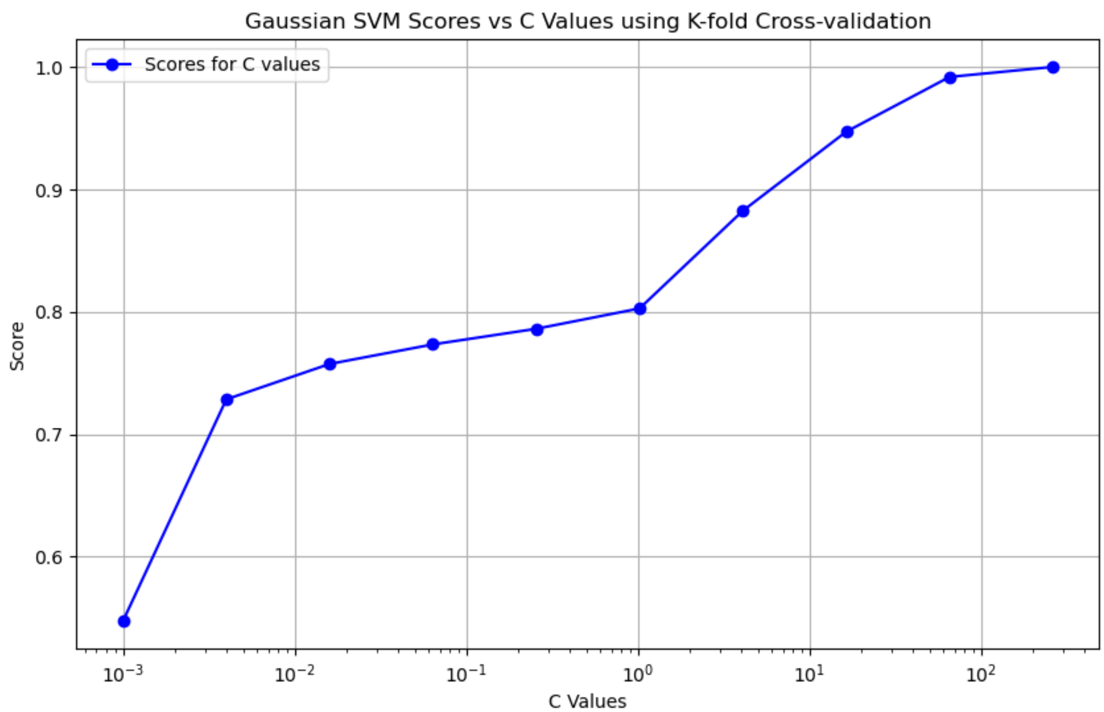

Figure 04. Gaussian SVM accuracy scores for different values of C using K-fold Cross-Validation as training technique

Above is shown that for higher values of C, better is the performance that we get. Even getting a perfect score for the value 262.144. This is because the model is looking to minimize the risk function where C is the coefficient of the sum of the loss function values, as C gets closer to 0, smaller are the values learned by the model, leading to underfitting.

Next, it was explored the changes on performance for distinct pairs of regularization values and kernel coefficients (C, γ), gamma controls the scale of the difference between two points, that is calibrating how much influence each pair of point has in deciding the classification region size.

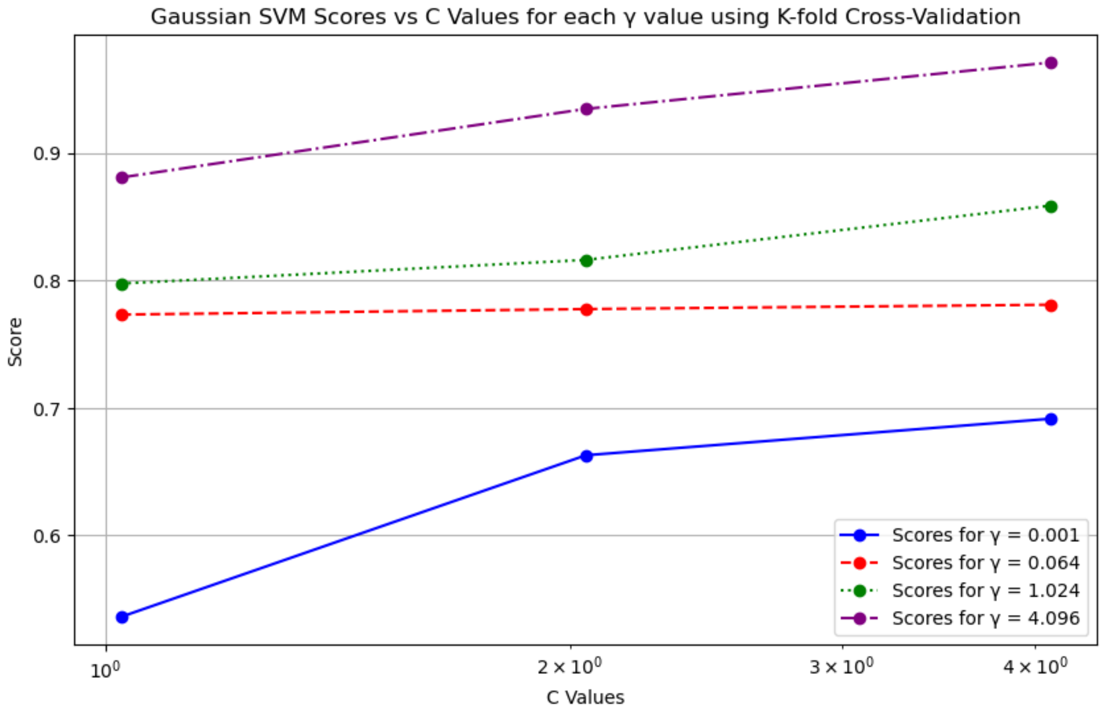

Figure 05. Gaussian SVM accuracy scores for different pairs of C & γ using Cross-Validation

This experiment concluded that for this data set, using cross-validation, bigger pair values combinations yield better performances.

The previous results led to questioning if this positive correlation between performance and higher parameter values is true. For that reason, it was experimented a larger pair of values in the full training set and its results were validated using the test set.

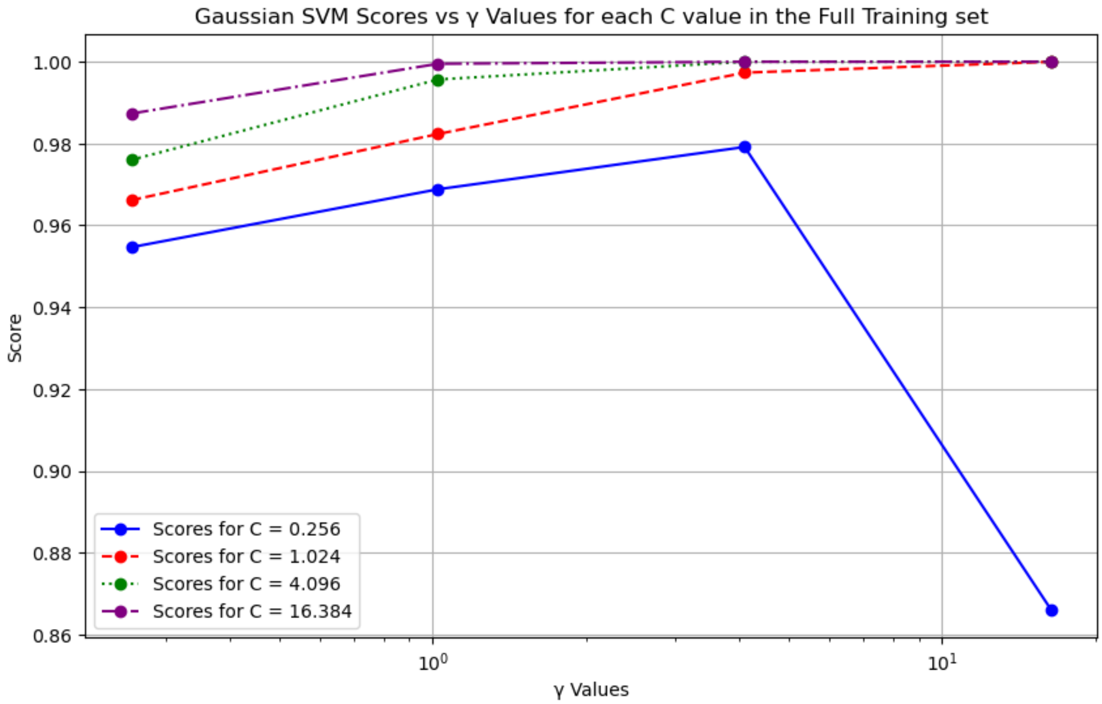

Figure 06. Gaussian SVM accuracy scores for different pairs of C & γ in the Full Training Set

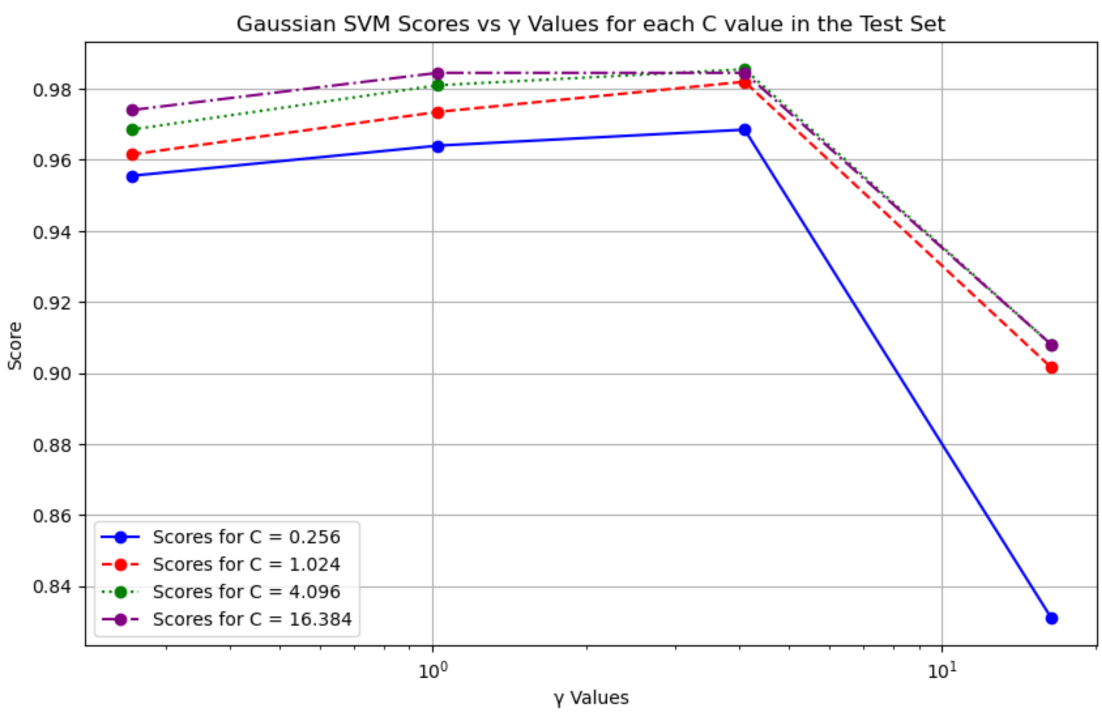

Figure 07. Gaussian SVM accuracy scores for different pairs of C & γ in the Test Set

For this set of values, the positive correlation maintains until the pair (4.096, 16.384). For higher γ values, the performance in the test set falls even less than the small initial pair values, due to overfitting because that is not the case in the training set.

MLPCs.

Finally, MLPClassifiers instances (Multi-Layer Perceptron) were set up with different configuration for their regularization parameter (α), activation functions, hidden layers, epochs and learning rates. MLPCs are a pipeline of neurons layers where each neuron contains a perceptron and a non-linear activation function.

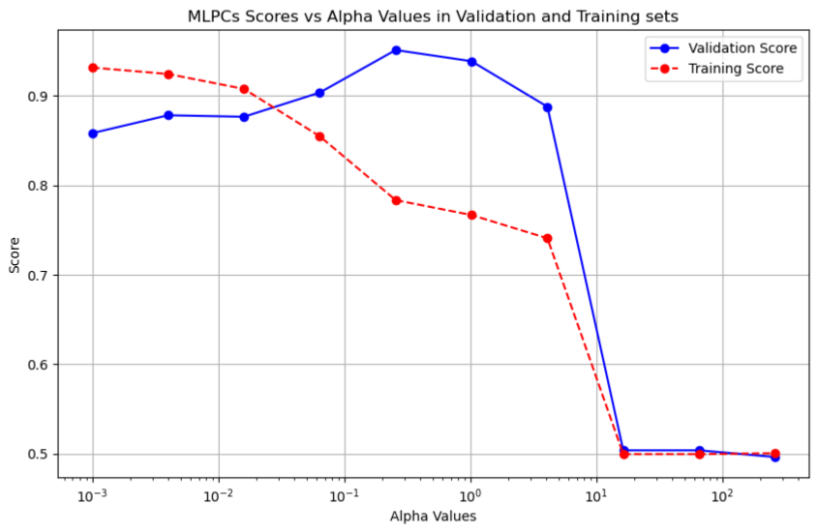

Figure 08. MLPCs accuracy scores for increasing alpha parameter values.

The first experiment was carried out, varying the alpha value parameter and the results are odd, for small values of C the training set obtained a better score, but for values closer to 1 the score in the validation set was much higher.

The second experiment that was carried out was varying the number of hidden layers and activation functions. This was done to check the performance changes of neuron distributions over a variable number of layers and to observe the behaviours of these functions under these configurations.

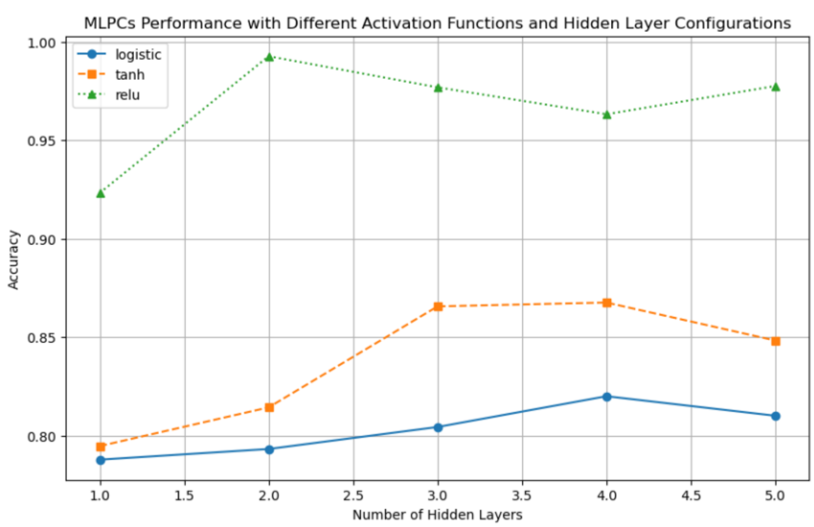

Figure 09. MLPCs accuracy scores for different neuron layers distribution with different activation functions.

The figure shows the results of testing models with the following hidden layer configurations: (100,), (50, 50), (25, 75, 25), (25, 50, 50, 25), and (25, 25, 25, 25). In these configurations, the i^{th} element represents the number of neurons in the i^{th} hidden layer. The models were evaluated using the logistic, tanh, and ReLU activation functions.

To finish MLPC experimentation there were built models that were trained with 50, 100, 200, 300 and 500 epochs; and used "constant", "invscaling" and "adaptive" learning rates.

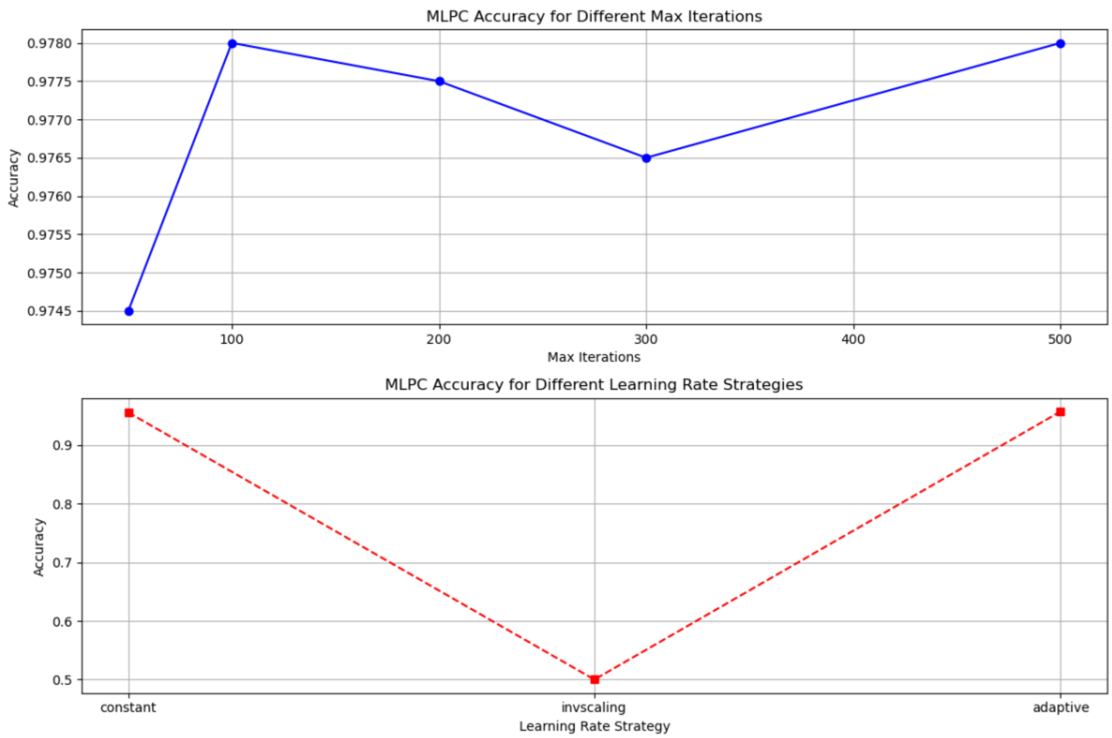

Figure 10. MLPCs accuracy scores for distinct training epochs and learning rates.

The figure above shows that for the number of iterations, the differences in the scores are minimal, 0.0035 being the maximum difference between two scores. In the other hand, the learning rate strategies have several implications in the model performance, an "adaptative" learning rate obtained the highest accuracy score (0.957).

Comparison.

Finally, the three best model configurations that outperformed in the previous experiments (best of Linear SVMs, Gaussian SVMs and MLPCs) were trained on the full not-noisy training set and its accuracy measured on the test set.

Figure 11. MLPCs accuracy scores for distinct training epochs and learning rates.

The previous figure exposes the performance of:

- MLPC that used ReLU as activation function, a hidden layer shape of (50, 50), stochastic descent gradient ("sgd") as solver, adaptative learning, a regularization parameter alpha of 0.256 and was trained with 100 epochs.
- Gaussian SVM with a value of 4.096 for the regularization (C) and gamma parameters.
- Linear SVM with a regularization parameter (C) of 4.096.

The model that performed the best was the Gaussian SVM that obtained an accuracy score of 0.986, followed by the Linear SVM with a score of 0.962 and finally the MLPC obtained 0.954, even though was the one that offered more option for tunning parameters.