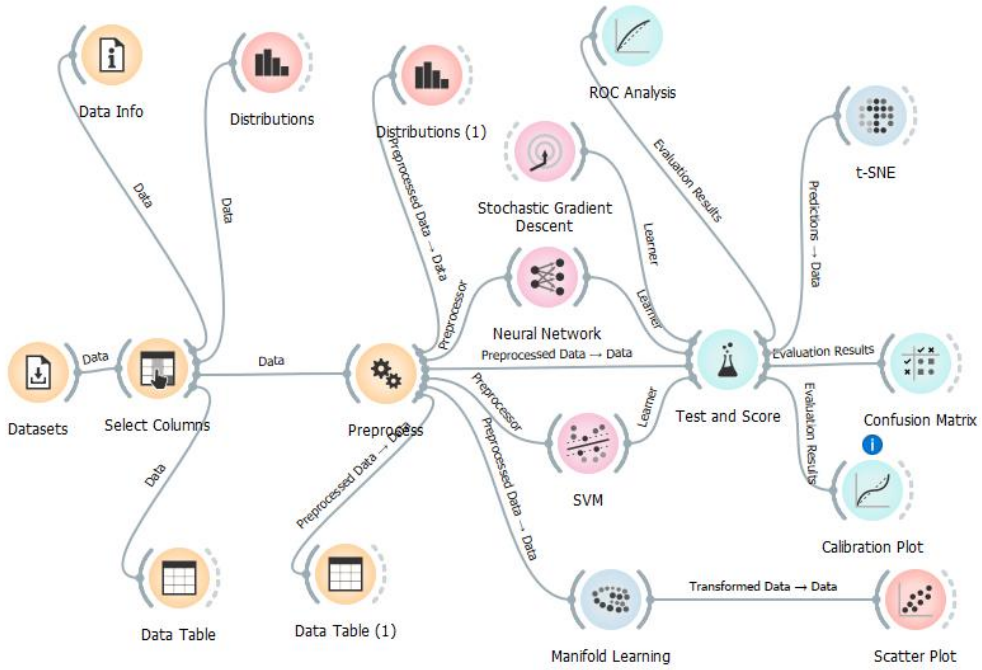
Anastasia Psarou – Assignment 5 – Machine Learning

This is my dataflow:



1. **Find the best method loss function for SGD regression.**

In the SGD widget we are going to use different kinds of loss functions that have to do with regression. Finally, with the use of Test and Score widget we are going to check the results that occur.

- Squared Loss

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.160	0.028	0.967	0.968	0.968	0.968	0.968	1.113	0.967

- Huber

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.171	0.028	0.967	0.968	0.968	0.968	0.968	1.113	0.967

- ε insensitive

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.163	0.026	0.966	0.966	0.966	0.967	0.966	1.163	0.966

- Squared ε insensitive

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.171	0.030	0.966	0.966	0.966	0.967	0.966	1.163	0.966

I believe that the best loss function is Squared Loss as the majority of evaluation results have a better value in comparison with the other loss functions. Although, we need to mention the results of the other loss functions are very close.

2. **Find the best activation function and the best optimization scheme for NN classifiers.**

As far as the best activation function for NN classifiers is concerned:

- Identity

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.345	0.039	0.987	0.965	0.965	0.966	0.965	0.107	0.968
SGD	0.134	0.024	0.954	0.963	0.963	0.964	0.963	1.264	0.946
Neural Network	2.744	0.041	0.994	0.969	0.969	0.969	0.969	0.094	0.964

- Logistic

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.345	0.039	0.987	0.965	0.965	0.966	0.965	0.107	0.968
SGD	0.134	0.024	0.954	0.963	0.963	0.964	0.963	1.264	0.946
Neural Network	5.963	0.037	0.994	0.968	0.968	0.968	0.968	0.092	0.963

- Tanh

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.345	0.039	0.987	0.965	0.965	0.966	0.965	0.107	0.968
SGD	0.134	0.024	0.954	0.963	0.963	0.964	0.963	1.264	0.946
Neural Network	3.907	0.044	0.994	0.969	0.969	0.969	0.969	0.094	0.964

- ReLu

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.345	0.039	0.987	0.965	0.965	0.966	0.965	0.107	0.968
SGD	0.134	0.024	0.954	0.963	0.963	0.964	0.963	1.264	0.946
Neural Network	9.869	0.047	0.993	0.966	0.966	0.967	0.966	0.090	0.966

From the above results it is obvious that most parameters except time have really close results between each other. The only parameter that has a lot of different values is train time. So, in order to determine the best activation time I am going to take into consideration the training time. The activation function that exhibits the lower time is Identity. So, I believe that Identity is the best activation function for NN classifiers.

As far as the best optimization scheme is concerned we are going to check the results in Test and Score widget from the different number of solvers of NN classifiers.

- L – BFGS - B

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.113	0.028	0.993	0.971	0.971	0.971	0.971	0.217	0.965
Neural Network	0.564	0.046	0.994	0.969	0.969	0.969	0.969	0.095	0.962

- SGD

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.113	0.028	0.993	0.971	0.971	0.971	0.971	0.217	0.965
Neural Network	3.701	0.042	0.995	0.968	0.968	0.968	0.968	0.092	0.961

- Adam

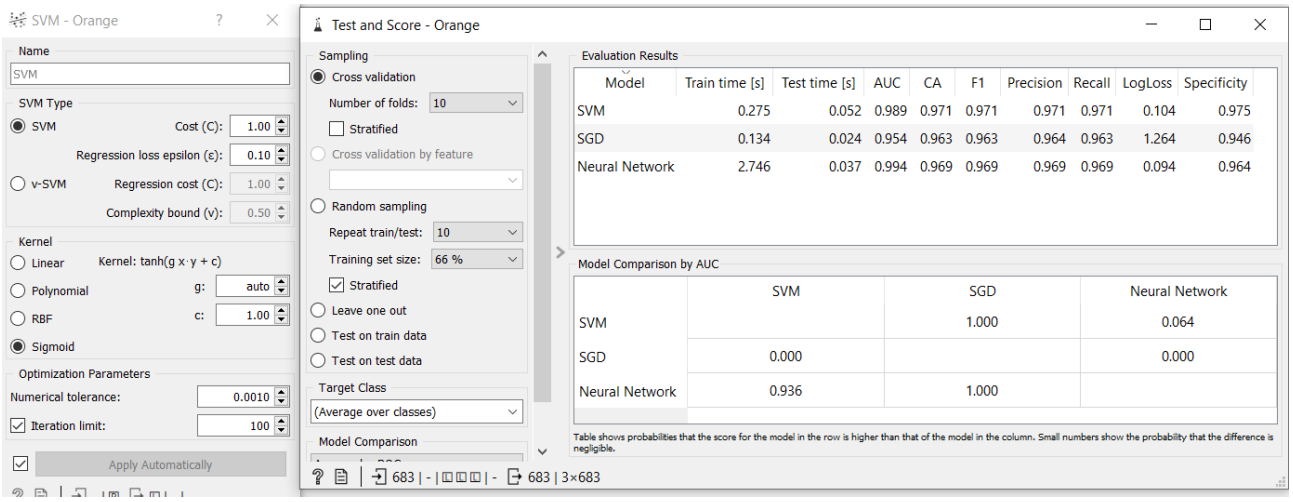
Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
SVM	0.275	0.052	0.989	0.971	0.971	0.971	0.971	0.104	0.975
SGD	0.113	0.028	0.993	0.971	0.971	0.971	0.971	0.217	0.965
Neural Network	2.780	0.042	0.994	0.969	0.969	0.969	0.969	0.094	0.964

We can see that all the results, except from the train time, have approximately the same value. So, I am also here going to take into consideration the Train Time. It is obvious that the train time of L – BFGS – B method is a lot smaller compared to the train time of the other methods. That is why I believe that L – BFGS – B is the best optimization scheme for NN classifiers.

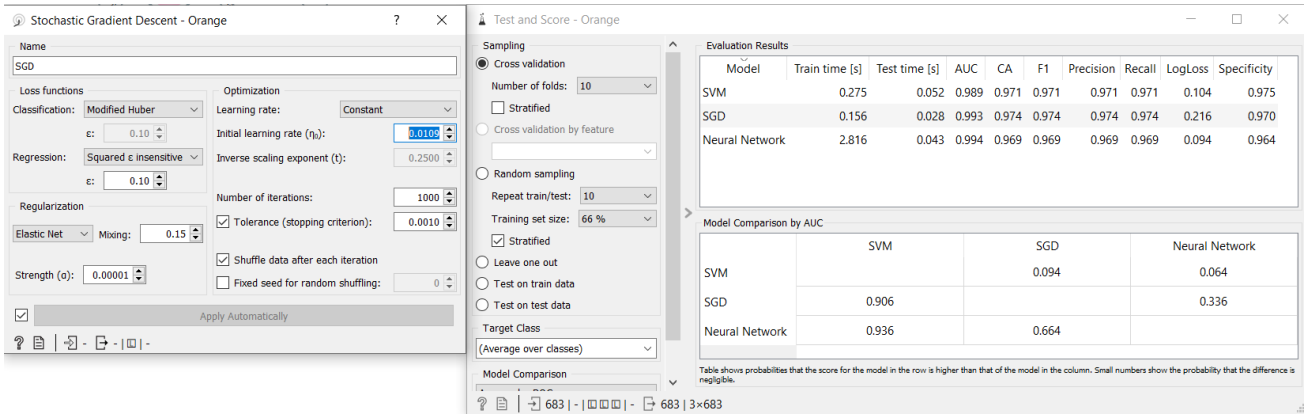
3. Play parameters to get the highest F1 values.

I tried to implement different algorithms in order to find the highest F1 values. We know that F1 has values between 0 and 1. So, the values we will find must be as closest to 1 as possible.

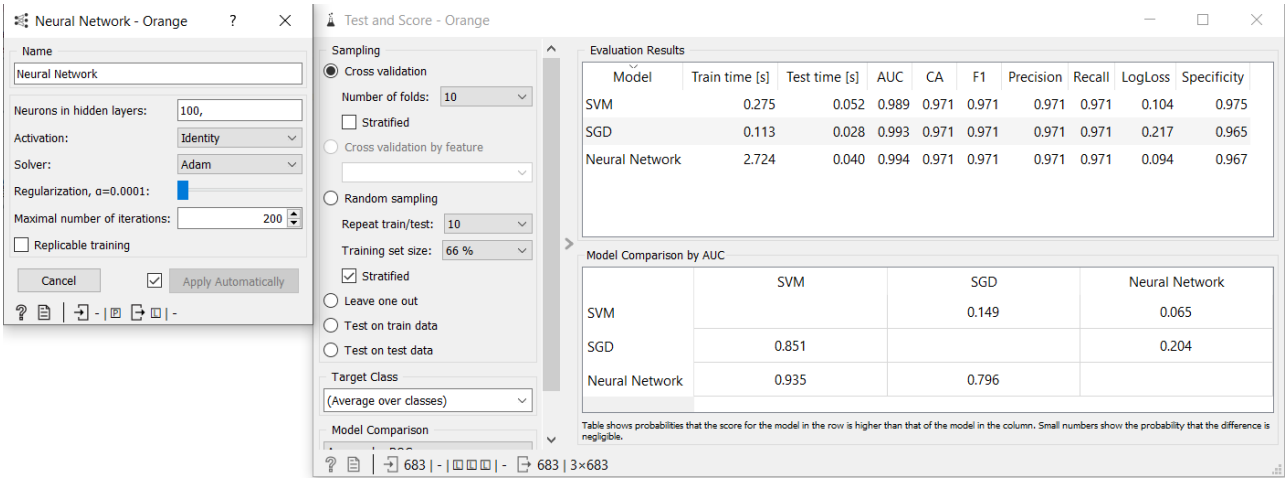
Here using SVM model I tried to achieve the highest value of F1. I used different kinds of kernels and SVM types and the one that produced the highest F1 value is SVM type with Sigmoid kernel. The final value of F1 = 0.971.



As far SGD is concerned I, also, tried to achieve the highest value of F1. In order to achieve that I used different kinds of loss functions that concern classification and regression and also different kinds of optimization learning rates. I also changed the Initial learning rate. So, the highest F1 value I achieved is 0.974.



As far as Neural Network is concerned, the highest value I could achieve is 0.971. To achieve this I implemented the Identity activation and the Adam solver. Also, replicable training was not used.



In conclusion, the highest F1 value I could achieve is 0.974 with the use of SGD.