University of Nottingham
UK | CHINA | MALAYSIA

Assessed Coursework 3: Support Vector Machines

COMP3009 Machine Learning

Team name: Omega

Introduction

This laboratory is about training SVM for regression and a classification data-set. Also, inner-cross validation is used for parameter optimization, and to compare the performance of the SVM models with ANN and Decision trees. The Matlab function ttest2 is used to prove statistically the significance of the difference of the results of the cross-validation for the models ANN, decision trees, and SVM.
For the classification method, the objective is to predict the use or not of a contraception method depending on mostly categorical variables related to their education, religion, age, the standard of living, and the number of children.
The original classification set had three categories (three types of contraceptive methods), but this was transformed into a binary problem (use or not of contraception), for that reason the ANN classification problem is excluded from the comparison.
The objective of the regression model is to predict the number of bikes rented per hour depending on variables related to the weather condition such as temperature, humidity, wind speed, and so on, other variables considered by the model are season and holidays.

a. Train SVMs with linear kernel

Linear Kernel SVM was the simplest method to train using the parameter BoxConstraint equal to 1. Only for the regression, it was necessary to tune the parameter Epsilon. In section a.2 the influence of the parameter Epsilon is studied.

a.1. Linear kernel for classification

This data is obtained running the file ClasificationLinearKernel.m

Number of support vectors	1065
% support vectors	72.3014

Fold	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average**
Accuracy	0.716	0.649	0.764	0.649	0.743	0.622	0.709	0.703	0.655	0.681	0.689
Precision	0.923	0.831	0.907	0.859	0.895	0.788	0.871	0.905	0.847	0.890	0.872
Recall	0.706	0.667	0.743	0.646	0.726	0.638	0.723	0.679	0.604	0.637	0.677

Figure 1 - 10 fold inner cross validation for linear kernel classification SVM.

a.2. Linear kernel for regression

a.2.1. The Sensibility of the value of epsilon

The data is obtained running the file EpsilonSensibility.m
As we see in figure 2 and figure 3, a high value for Epsilon signifies high values of the error, conversely, when Epsilon is high the number of support vectors decreases. This is understandable because Epsilon defines the margin of tolerance for the error and support vectors are by definition points that touch the boundaries of the margin.

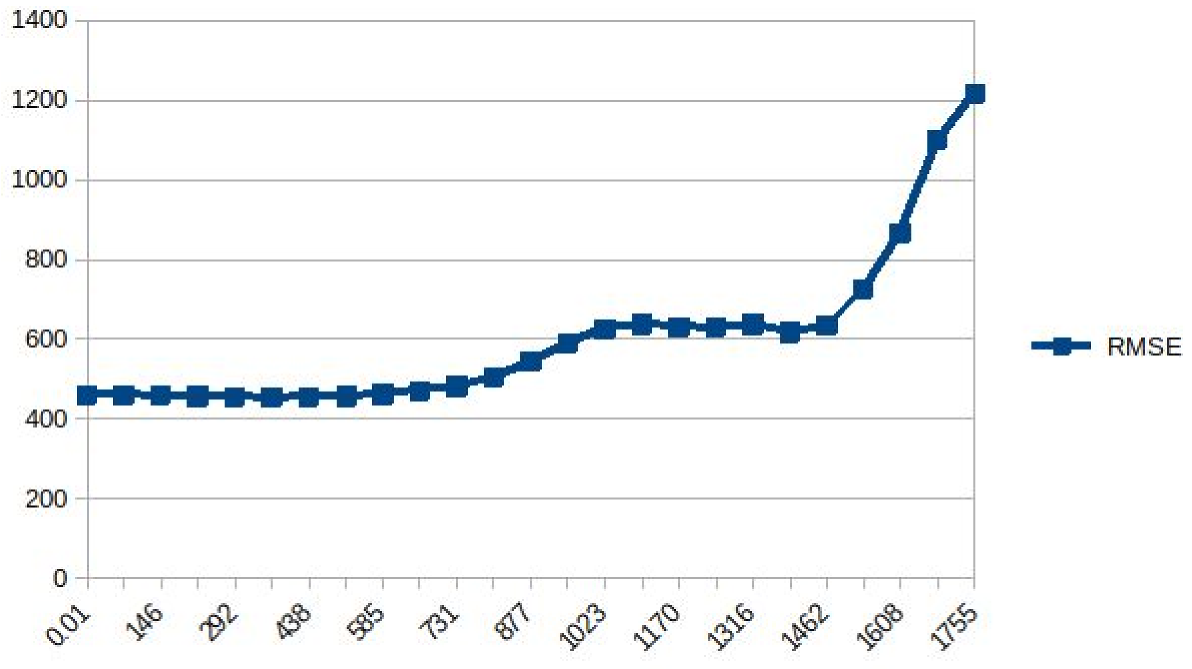

Figure 2 - Relation between Epsilon and RMSE.

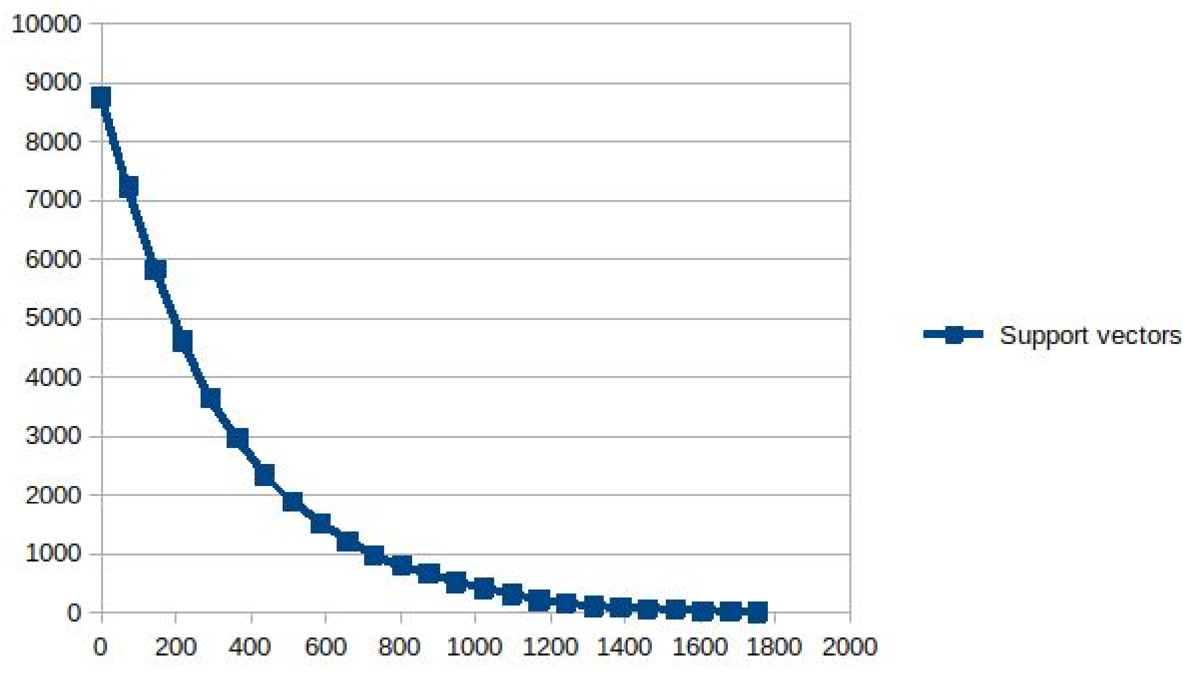

Figure 3 - Relation between Epsilon and support vectors.

a.2.2. 10-fold validation for the linear kernel regression model

This data is obtained running the file RegressionLinearKernel.m

Fold	1	2	3	4	5	6	7	8	9	10	Average
RMSE	454.8	489.8	428.9	472.6	432.7	469.6	498.9	451.1	486.9	425.0	461.0
Epsilon	1.00	82.42	0.50	0.07	4.33	82.42	0.07	82.42	82.42	82.42	41.8
Support Vectors	7877	6323	7884	7884	7821	6332	7884	6312	6336	6331	7098.4
Support Vectors%	89.92	72.18	90.00	90.00	89.28	72.28	90.00	72.05	72.33	72.27	81.0

Figure 4 - 10 fold inner cross-validation for linear kernel regression SVM.

b.1 Classification model:

b.1.1 RBF kernel SVM classification

This data was obtained running the file ClassificationRBFkernel.m

Fold	1	2	3	4	5	6	7	8	9	10	Average
Box Constraint	128.0	2048.0	128.0	2.00	2.00	512.0	2.00	0.03	0.50	8.00	283.05
sigma	5.66	5.66	22.63	1.41	5.66	22.63	5.66	0.00	1.41	5.66	7.64
Kernel Scale	8.00	8.00	32.00	2.00	8.00	32.00	8.00	0.00	2.00	8.00	10.80
Accuracy	0.72	0.73	0.72	0.71	0.74	0.63	0.76	0.55	0.70	0.74	0.70
Precision	0.86	0.83	0.89	0.82	0.91	0.88	0.88	1.00	0.77	0.85	0.87
Recall	0.74	0.77	0.70	0.71	0.72	0.58	0.71	0.55	0.71	0.76	0.70
Suport Vectors	799	758	877	1007	892	821	884	1276	1087	835	923.60
Suport Vectors %	54.24	51.46	59.54	68.36	60.56	55.74	60.01	86.63	73.79	56.69	62.70

Figure 5 - 10 fold inner cross-validation for RBF kernel classification SVM.

b.1.2 Polynomial kernel SVM classification

ClassificationPolynomialKernel.m

Fold	1	2	3	4	5	6	7	8	9	10	Avg.
Box Constraint	0.50	8192.00	512.00	0.03	2048	2048	512	0.13	0.03	0.13	1331.28
Kernel Scale	32.00	0.50	32.00	32.00	2.00	8.00	2.00	32.00	0.50	0.01	14.10
Polynomial Order	6.00	0.50	6.00	10.00	0.50	3.00	3.00	6.00	0.50	0.50	3.60
Support Vectors	918.00	682.00	833.00	912.	947.	1217	1124	942	976	896	944.70
Support Vectors %	62.32	46.30	56.55	61.91	64.29	82.62	76.31	63.95	66.26	60.83	64.13
Accuracy	0.72	0.49	0.58	0.69	0.66	0.45	0.38	0.69	0.75	0.45	0.59
Precision	0.94	0.37	0.48	0.93	0.77	0.26	0.28	0.95	0.90	0.41	0.63
Recall	0.70	0.55	0.69	0.65	0.68	0.62	0.44	0.68	0.74	0.43	0.62

Figure 6 - 10 fold inner cross-validation for polynomial kernel classification SVM.

b.2 Regression model::

b.2.1 RBF kernel SVM regression:

File: RegressionRbfKernel.m

Fold	1	2	3	4	5	6	7	8	9	10	Avg.
RMSE	394.48	454.34	383.22	409.86	399.98	419.18	443.22	377.66	409.05	380.34	407.13
Epsilon	82.42	4.33	82.42	82.42	82.42	4.33	82.42	2.00	82.42	1.00	50.62
Box Constraint	2048.0	512.00	512.00	512.00	512.00	512.00	512.00	2048.0	2048.0	512.00	972.80
sigma	22.63	22.63	22.63	22.63	22.63	22.63	22.63	22.63	22.63	22.63	22.63
Kernel Scale	32.00	32.00	32.00	32.00	32.00	32.00	32.00	32.00	32.00	32.00	32.00
Support Vectors	5454.0	7682.0	5521.0	5523.0	5528.0	7707.0	5526.0	7818.0	5475.0	7812.0	6404.6
Support Vectors %	62.26	87.69	63.03	63.05	63.11	87.98	63.08	89.25	62.50	89.18	73.11

Figure 7 - 10 fold inner cross-validation for RBF kernel regression SVM.

b.2.2 Polynomial kernel SVM regression:

File: RegressionPolynimialKernel.m

Fold	1	2	3	4	5	6	7	8	9	10	Avg.
RMSE	371.4	393.47	389.67	390.61	282.26	20103.7	653.2	276.79	526.23	559.48	2394.68
Epsilon	0.010	0.010	82.415	4.330	82.415	0.010	4.330	82.415	4.330	0.010	26.028
Box Constraint	512	2048	0.031	512	0.031	0.03125	0.031	0.031	128	0.031	320.02
Kernel Scale	32	32	0.5	32	8	0.00049	32	32	32	32	23.25
Support Vectors	94	94	74	92	82	77	94	79	94	98	87.80
Support Vectors %	89.52	89.52	70.48	87.62	78.10	73.33	89.52	75.24	89.52	93.33	83.62

Figure 8 - 10 fold inner cross-validation for polynomial kernel regression SVM.

c. Compare performance between methods

The performance between methods was evaluated by comparing the averages of the 10-fold cross-validation. The statistical significance was tested using the ttest2 Matlab function for means. The p-values of this test can be used as a decision boundary for the rejection of the null hypothesis. A low value of p means a low probability of rejection. The file ttest_2.m contains the code for the execution of the tests. The variables p_c and p_r correspond to the p-values for classification and regression respectively.

c.1. Compare classification methods

Fold	1	2	3	4	5	6	7	8	9	10	Avg.
Decision trees	0.730	0.642	0.635	0.541	0.919	0.919	0.730	0.189	0.777	0.872	0.695
Linear kernel SVM	0.716	0.649	0.764	0.649	0.743	0.622	0.709	0.703	0.655	0.681	0.689
Polynomial kernel SVM	0.723	0.486	0.581	0.689	0.662	0.446	0.378	0.689	0.750	0.447	0.585
RBF kernel SVM	0.723	0.730	0.723	0.709	0.743	0.628	0.757	0.554	0.696	0.738	0.700

Figure 9 - Compare classification methods using 10 fold cross-validation.

p values	Decision trees	Linear kernel SVM	Polynomial kernel SVM	RBF kernel SVM
Decision trees	1.000	0.930	0.194	0.948
Linear kernel SVM	0.930	1.000	0.042	0.656
Polynomial kernel SVM	0.194	0.042	1.000	0.030
RBF kernel SVM	0.948	0.656	0.030	1.000

Figure 10 - p-values (classification methods) for the ttest2 without assuming equal variance. $H_1 :\ mean_1 \neq mean_2$ **, If the p-value is higher than 0.05 ttest2 does not reject** H_0 **with 5% of significance.**

Decision tree presents different performance than Linear, Polynomial, and RBF with p-values of 0.93, 0.194, and 0.948 respectively, Although RBF is different to the linear kernel, there is no statistical evidence of the difference between Polynomial, linear (p = 0.042), and polynomial RBF(p = 0.03).

c.2. Comparison of the regression methods

Regression models	1	2	3	4	5	6	7	8	9	10	Avg.
ANN	277.72	491.54	426.75	543.37	288.69	384.97	403.42	345.87	453.66	353.21	396.92
Regression Trees	418.03	393.47	433.28	411.26	420.72	439.80	428.53	428.41	414.37	413.50	420.14
Linear kernel SVM	454.78	489.80	428.88	472.56	432.66	469.65	498.91	451.15	486.89	424.96	461.02
Polynomial kernel SVM	371.37	393.47	389.66	390.61	282.26	433.97	653.22	276.80	526.23	559.48	427.71
RBF kernel SVM	394.48	454.34	383.22	409.86	399.98	419.18	443.22	377.66	409.05	380.34	407.13

Figure 11 - Compare regression methods using 10 fold cross-validation.

	ANN	Regression Trees	Linear kernel SVM	Polynomial kernel SVM	RBF kernel SVM
ANN	1.000	0.405	0.035	0.516	0.721
Regression Trees	0.405	1.000	0.0004	0.845	0.175
Linear kernel SVM	0.035	0.0004	1.000	0.402	0.0002
Polynomial kernel SVM	0.516	0.845	0.402	1.000	0.602
RBF kernel SVM	0.721	0.175	0.0002	0.602	1.000

Figure 12 - p-values (regression methods) for the ttest2 without assuming equal variance. $H_1 : mean_1 \neq mean_2$, If the p-value is higher than 0.05 ttest2 does not reject H_0 with 5% of significance.

As we see in figure 12, the difference between ANN, Regression Trees, Polynomial Kernel, and RBF kernel is statistically significant. According to ttest2 the mean of the Polynomial Kernel is different from the others. However, there is no statistical evidence of the difference between Linear Kernel SVM, ANN, Regression Trees, and RBF Kernel SVM.

d.1. Question 1

According Matlab documentation KernelScale = 1/sqrt(gamma)), and for the definition of the gaussian kernel, gamma = 1/(2*sigma^2).
https://uk.mathworks.com/matlabcentral/answers/328796-gaussian-kernel-scale-for-rbf-svm
Then sigma = KernelScale/sqrt(2).

Slow values of sigma mean that the kernel will consider further points to fit the decision boundary, in contrast, high values of sigma means that it is going to consider closer points to fit the decision boundary. As a result, a low value of sigma means a more linear boundary.
https://www.youtube.com/watch?v=m2a2K4lprQw

d.2. Question 2

A high value of the parameter C results in a hard margin and a soft margin is given by a low value of C. Decreasing C causes loss of separability but gains stability. The problem with hard margin is that a simple outlier can determine the boundary, making the model oversensitive to the noise data and increasing overfitting. In a hard margin, all the points are classified correctly, which means that if the data is overlapped (not separable data), it is not possible to fit a hard-margin SVM. In simple words, whether the data has overlap or not the consequence of hard-margin is overfitting. If the data is separable, a high value of C is enough to achieve a hard margin but must be detected with all the values of the training set classified correctly.
http://fourier.eng.hmc.edu/e161/lectures/svm/node5.html
https://stackoverflow.com/questions/4629505/svm-hard-or-soft-margins

d.2. Question 3

Sub-sampling before splitting the data means that there could be important information in the subsets that the model will not learn, but by splitting the data in k-fold just before start training, we can surpass this problem by taking folds from all the data.

Conclusion

All the methods studied in this laboratory have the potential capacity to predict, but their effectiveness depends on the right choice of the parameters. Nested cross-validation shows as a powerful tool of Machine Learning to compare models, parameter tuning, and avoid overfitting. Besides, a statistical test for the difference of the means such as ttest2 of Matlab is a good tool to compare the results of the cross-validation.

The highest values of classification performance were given by RBF, Classification Trees, and Linear Kernel with an accuracy of 0.7, 0.695, and 0.689 respectively and these values present a difference statistically significant. For regression the lowest values of RMSE were ANN and RBF with 396.92 and 407.13 respectively, these values also present a statistical significance in their difference.