Project 2 Report

Team 10: Bhavesh Poddar, Nanda Kishore Sunki, Vineeth Chennapalli, Chuji Luo

1 Member Roles

	TrainMyClassifier			MyCrossValidate	TestMyClassifier			MyConfusionMatrix
	RVM	SVM	GPR		RVM	SVM	GPR	
B. Poddar		1				1		
N. K. Sunki			1				1	
V. Chennapalli	1				1			
C. Luo				1				1

2 Observations

In this project, all the three algorithms were trained using all-pairs methods. Because of time limit, we performed 5-fold cross validation only on a random subset (5000 samples) of the dataset provided by Dr. Gader and compared the 3 algorithms based on this subset.

Note that the 'NaN' in confusion matrix represents there is no sample from the unseen class.

2.1 Support Vector Machines

	C_1	C_2	C_3	C_4	C_5	Unseen
C_1	0.99	0.00	0.00	0.00	0.00	0.00
C_2	0.01	0.96	0.00	0.02	0.00	0.01
C_3	0.00	0.00	0.98	0.00	0.00	0.01
C_4	0.00	0.02	0.00	0.96	0.01	0.01
C_5	0.00	0.00	0.00	0.01	0.98	0.00
Unseen	NaN	NaN	NaN	NaN	NaN	NaN

Table 1 Overall Confusion Matrix for SVM. Overall Accuracy = 97.29%. (5,000 samples)

	# of support vectors
1st Fold	366
2nd Fold	362
3rd Fold	353
4th Fold	355
5th Fold	358

Table 2 Number of Support Vectors for Each Fold (5,000 samples)

2.2 Relevance Vector Machines

	C_1	C_2	C_3	C_4	C_5	Unseen
C_1	0.71	0.05	0.08	0.13	0.00	0.03
C_2	0.11	0.68	0.03	0.13	0.01	0.04
C_3	0.02	0.08	0.66	0.04	0.16	0.04
C_4	0.10	0.14	0.07	0.61	0.03	0.05
C_5	0.01	0.01	0.13	0.05	0.77	0.04
Unseen	NaN	NaN	NaN	NaN	NaN	NaN

Table 3 Overall Confusion Matrix for RVM. Overall Accuracy = 68.41%. (5,000 samples)

	# of relevance vectors
1st Fold	
2nd Fold	
3rd Fold	
4th Fold	
5th Fold	

Table 4 Number of Relevance Vectors for Each Fold (5,000 samples)

2.3 Gaussian Process Regression

	C_1	C_2	C_3	C_4	C_5	Unseen
C_1	1.00	0.00	0.00	0.00	0.00	0.00
C_2	0.00	0.96	0.01	0.03	0.00	0.00
C_3	0.04	0.00	0.89	0.00	0.07	0.00
C_4	0.00	0.01	0.00	0.95	0.03	0.00
C_5	0.01	0.00	0.00	0.00	0.98	0.00
Unseen	NaN	NaN	NaN	NaN	NaN	NaN

Table 5 Overall Confusion Matrix for GPR. Overall Accuracy = 95.63% (5,000 samples)

3 Conclusion

Accuracy: Based on 5000 random samples, the accuracy for SVM, RVM and GPR are respectively 97.29%, 68.41% and 95.63%. Thus, SVM is the best.

Speed: From both time complexity and our experiments, SVM is the fastest algorithm.

Hence, SVM is the best algorithm in terms of accuracy and speed. To end this report, we provided the results of SVM for the whole dataset (25,000 samples) below.

	C_1	C_2	C_3	C_4	C_5	Unseen
C_1	0.99	0.00	0.00	0.00	0.00	0.00
C_2	0.00	0.97	0.00	0.02	0.00	0.00
C_3	0.01	0.00	0.98	0.00	0.01	0.00
C_4	0.00	0.01	0.00	0.97	0.01	0.01
C_5	0.00	0.00	0.00	0.00	0.99	0.00
Unseen	NaN	NaN	NaN	NaN	NaN	NaN

Table 6 Overall Confusion Matrix for SVM. Overall Accuracy = 98.18% (25,000 samples)

	# of support vectors
1st Fold	1143
2nd Fold	1201
3rd Fold	1182
4th Fold	1183
5th Fold	1182

Table 7 Number of Support Vectors for Each Fold. (25,000 samples)