

## ML Project 2 Report - Team 6

Team Members: Ying Zhu, Hang Jin, Abishek Sethuraman, Ricardo Eiris Pereira

### Results for 5 fold Validation

#### 1. Support Vector Machine - Overall confusion matrices for SVM classifier:

	One	Two	Three	Four	Five
One	0.9788	0.003	0.0034	0.0004	0.0014
Two	0.007	0.9756	0.0024	0.0186	0.0004
Three	0.0074	0.0022	0.986	0.0012	0.0074
Four	0.0026	0.0184	0.0016	0.9718	0.0054
Five	0.004	0.0008	0.0064	0.0078	0.9854

Number of Support Vectors for each fold:

Fold 1	[ 90 108 99 125 101]
Fold 2	[ 88 116 105 132 100]
Fold 3	[ 88 121 98 125 103]
Fold 4	[ 80 114 101 131 97]
Fold 5	[ 85 121 101 135 100]

#### 2. Relevance Vector Machine -Overall confusion matrices for RVM classifier:

	One	Two	Three	Four	Five
One	0.953	0.0085	0.011	0.0025	0.0058
Two	0.0101	0.8934	0.0054	0.097	0.0012
Three	0.013	0.004	0.9424	0.0086	0.0284
Four	0.0105	0.092	0.0108	0.862	0.0248
Five	0.0132	0.0022	0.0305	0.03	0.94

Number of Relevance Vectors for each fold: 748 for each fold

### 3. Gaussian Process - Overall confusion matrices for GP classifier:

	One	Two	Three	Four	Five
One	0.971	0.0082	0.0092	0.0028	0.0042
Two	0.0096	0.898	0.0044	0.0738	0.0012
Three	0.009	0.0042	0.9582	0.0066	0.0346
Four	0.0052	0.07	0.0094	0.8872	0.0234
Five	0.0052	0.0194	0.0192	0.029	0.9368

## Experiment Observations

**Support Vector Machine - Advantages:** (1) The regularization parameter “C” provides insight on overfitting. (2) Uses convex optimisation which translate to having no local minima. (3) Flexibility and robustness once training parameters are established.

**Disadvantages:** (1) Non-probabilistic estimations for the classifier. (2) Prone to overfitting as it is difficult to properly select the regularisation and kernel parameters.

**Relevance Vector Machine - Advantages:** (1) Utilizes bayesian inference to provide probabilistic classification, similarly to GP, but opposite from SVM. (2) Complexity parameters not required to avoid overfitting. (3) Convergence is faster compared to the other methods. **Disadvantages:** (1) Computationally expensive as it requires the inversion of a sparse matrix. (2) Requires the use of Prune strategies which can increase the chances of overfitting.

**Gaussian Process - Advantages:** (1) Utilizes bayesian inference to provide probabilistic classification, similarly to RVM, but opposite from SVM. (2) Output only depends on Kernel hyperparameters, facilitating accurate prediction. **Disadvantages:** (1) Computationally expensive as it scales cubically with the number of data points.