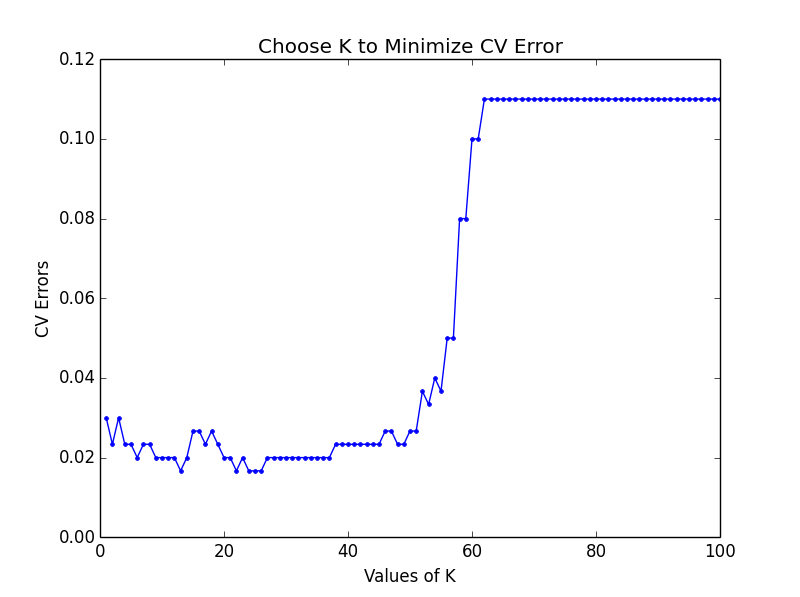
**Machine Learning Section 6100**

**HW11**

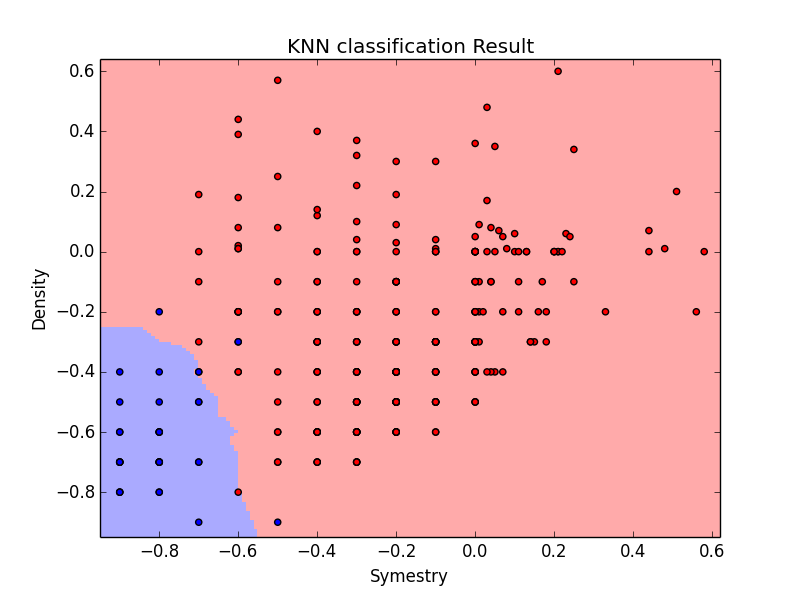
**Pu Zhang (661536015)**

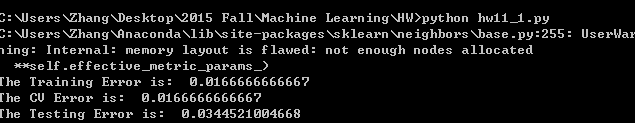
1. **KNN**
2. I used 10 folds cross validation instead of Leave-one-out for computational efficiency. Also the value of k is bounded by 0 to 100 as the input size is 300 so is unreasonable. Below is the plot of versus the value of k.



I would choose as it minimizes the CV error.

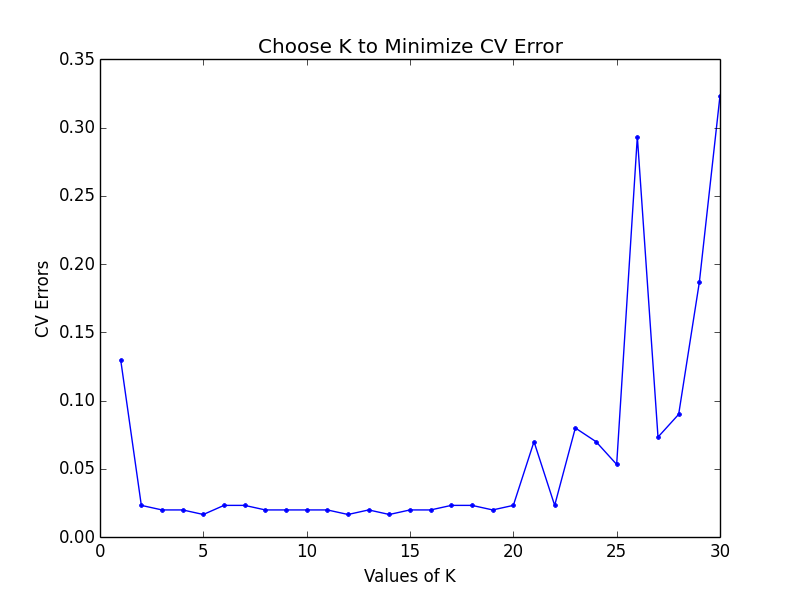
1. When , the KNN classification boundary is as below:





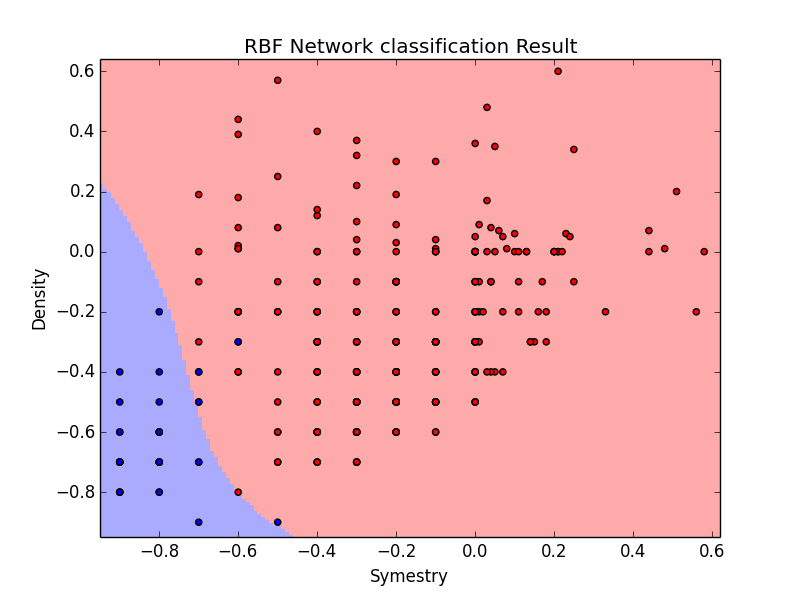
The in-sample error is **0.0167**, CV error is **0.0167**.

1. The testing error is **0.0345**.
2. **RBF Network**
3. I used 10 folds cross validation instead of Leave-one-out for computational efficiency. The number of centers k is bounded by 0 to 30.

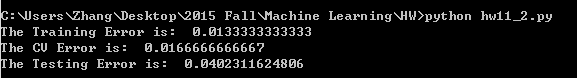


I would choose to minimize the CV error.

1. The RBF Network classification boundary when is as below:



Where the in-sample error is **0.0133**, the CV error is **0.0167**.



1. The testing error is **0.0402**.
2. **Compare linear regression, KNN and RBF network**

The following table summarizes the classification error of the testing data with the three approaches:

1. Linear Regression with 8th order polynomial transform and regularization selected by CV (LR)
2. KNN with k selected by CV
3. RBF Network with the number of centers selected by CV

|  |  |
| --- | --- |
| Approach | Testing Error |
| LR | 0.0893 |
| KNN | 0.0345 |
| RBF Network | 0.0402 |

It can be concluded that Linear Regression gives the largest classification error, while KNN and RBF Network performs similarly, KNN gives a better testing result.

Here are some Insights:

1. Linear regression is computationally efficient as we have the pseudo-inverse algorithm to solve for the weights. Also it does not require computing any distance. However, since it is a parametric method, it is less flexible and gives larger classification errors.
2. KNN is computationally demanding as distances are required for classification. Since it is nonparametric, it is more flexible to deal with. The choice of K requires CV. With a proper value of K, it gives better classification accuracy than LR.
3. RBF Network is based on the centers of clusters. It gives good classification result and is somehow robust to outliers.