

Machine Learning – Project3

- Keerti Keerti (KXK190012)

In this project , worked on the Wisconsin Breast Cancer dataset. There are 569 examples, each labeled as 0 or 1. SVM classifiers and scikit-learn is used here. The challenge is to select the free parameters to maximize the accuracy.

1. A classifier trained with 8% of the data using a polynomial kernel.

a. Here the model is trained over

"C=100,kernel='poly',degree=2,coef0=300,gamma=.9"

b. The penalty factor: C is set to 100 making the model misclassify less as much as possible, kernel is polynomial of degree 2 with coef0=300 and gamma = 0.9 making the decision boundary wider instead of making it compressed and dependent on each datapoint.

2. A classifier trained with 8% of the data using a rbf kernel.

a. Here the model is trained over

"C=1000,kernel='rbf',random_state=0,coef0=300,gamma=0.1"

b. Here the important factor is gamma as RBF calculates based on $\exp(-\gamma|u-v|^2)$ thus making gamma as the main factor. Here the gamma chosen is 0.1 and penalty factor as 1000.

3. A classifier trained with 16% of the data using a polynomial kernel.

a. Here the model is trained over

"C=100,kernel='poly',degree=2,coef0=300,gamma=.9"

b. The penalty factor: C is set to 100 making the model misclassify less as much as possible, kernel is polynomial of degree 2 with coef0=300 and gamma = 0.9 making the decision boundary wider instead of making it compressed and dependent on each datapoint.

4. A classifier trained with 16% of the data using a rbf kernel.

a. Here the model is trained over

"C=1000,kernel='rbf',random_state=0,coef0=300,gamma=0.1"

b. Here the important factor is gamma as RBF calculates based on $\exp(-\gamma|u-v|^2)$ thus making gamma as the main factor. Here the gamma chosen is 0.1 and penalty factor as 1000.

Comparison of classifier trained on 8% of data using polynomial and rbf kernel

Classifier	Seed = 2	5	-5	50	-50	100	1000
Polynomial	92.97	92.79	90.51	90.68	87.52	85.24	78.74
RBF	92.27	86.46	90.69	91.56	87.17	79.09	80.67

Comparison of classifier trained on 16% of data using polynomial and rbf kernel

Classifier	Seed = 2	5	-5	50	-50	100	1000
Polynomial	93.49	82.24	92.26	92.67	86.82	89.28	92.61
RBF	90.16	84.53	93.14	88.40	91.38	88.75	92.79

Here we can see the varying accuracy values based on the seeds.

It clearly shows that the on an average all classifiers are classifying the data with an average accuracy of 85-90%.