Support Vector Machines (SVM) Tuning

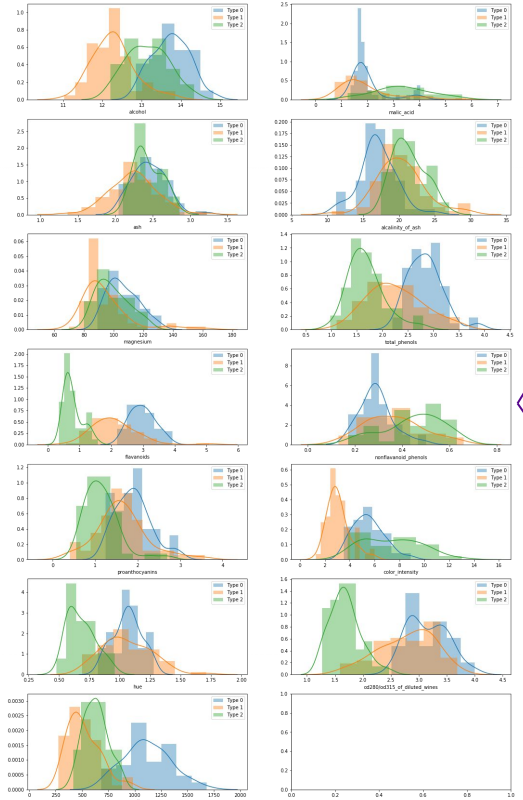
Comparing a Baseline SVM with a Manually Modified SVM and GridSearchCV Resulting SVM to Solve A Multiclass Classification Problem

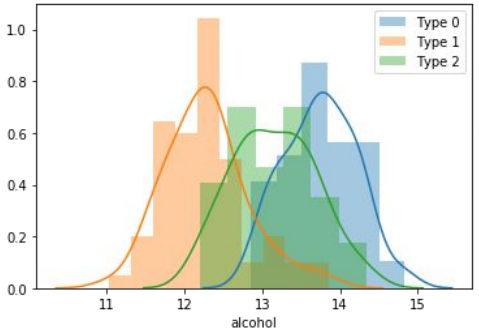
جامعة نيويورك أبوظبي

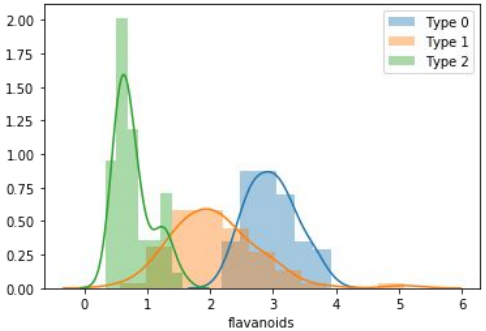
NYU | ABU DHABI

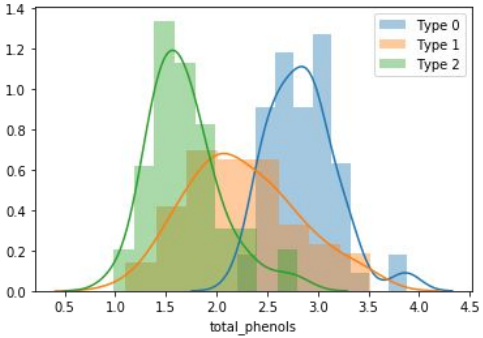
Mini Project 2: Multiclass Classification with SVM

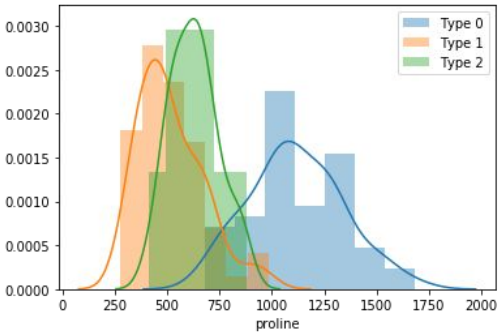
Understanding the Data











Data Checking and Preprocessing

- ➔ Confirmed no missing data
- ➔ Splitting the data into train/test (70/30)
- ➔ Converting to a Pandas dataframe for easier access

2.2 Scaling

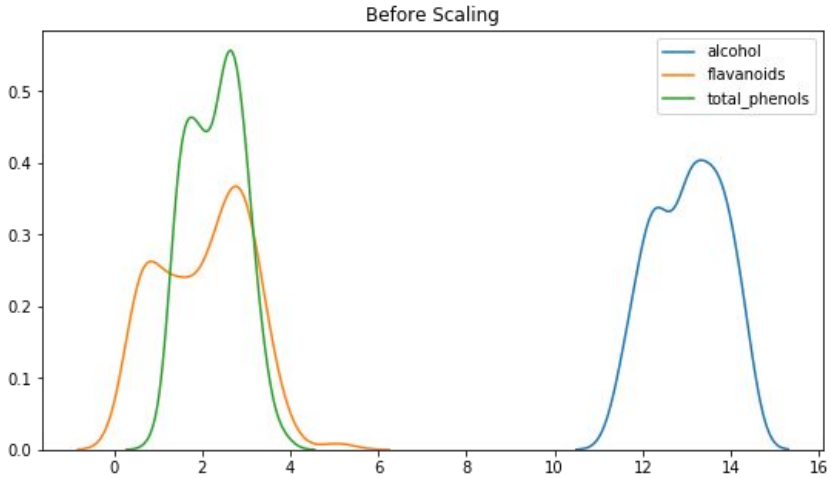
Scaling before applying SVM is very important. Part 2 of Sarle's Neural Networks FAQ Sarle (1997) explains the importance of this and most of considerations also apply to SVM. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation. Because kernel values usually depend on the inner products of feature vectors, e.g. the linear kernel and the polynomial kernel, large attribute values might cause numerical problems. We recommend linearly scaling each attribute to the range $[-1, +1]$ or $[0, 1]$.

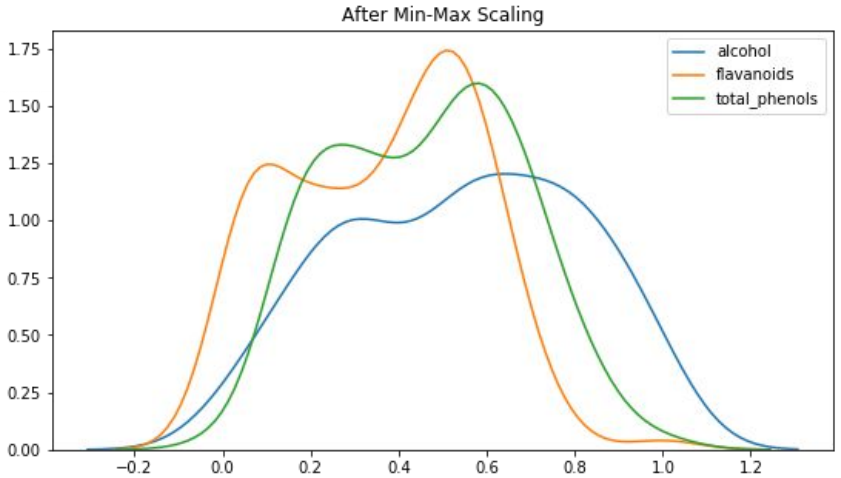
Of course we have to use the same method to scale both training and testing data. For example, suppose that we scaled the first attribute of training data from $[-10, +10]$ to $[-1, +1]$. If the first attribute of testing data lies in the range $[-11, +8]$, we must scale the testing data to $[-1.1, +0.8]$. See Appendix B for some real examples.

Source: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Data Preprocessing

جامعة نيويورك أبوظبي
NYU ABU DHABI





- ➔ Used Min-Max Scaler to avoid certain features from dominating
- ➔ Min-Max Scaler maintains natural skew of the densities
- ➔ Robust and Standard Scaler do not maintain this desired skew

Baseline Model

جامعة نيويورك أبوظبي
NYU ABU DHABI

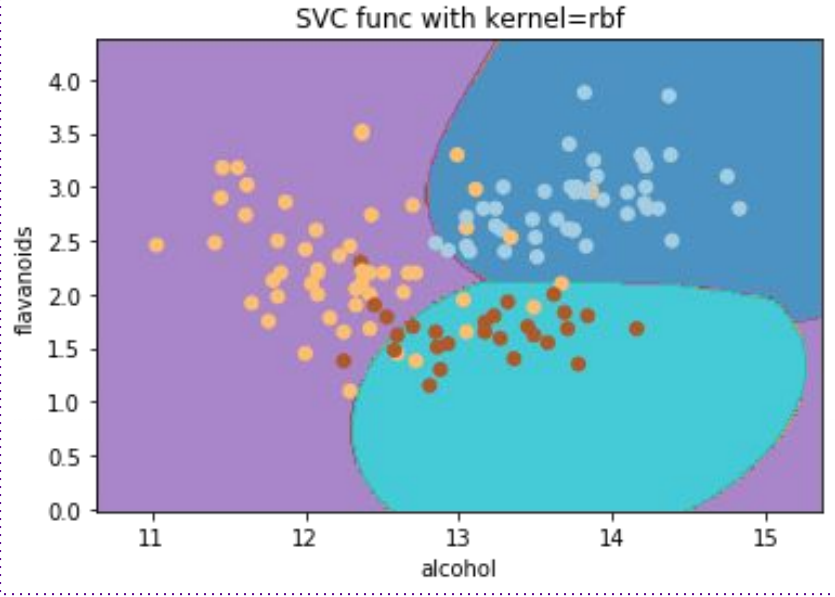
```
clf = SVC(kernel='rbf', gamma='auto')
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

- ➔ Skewed towards classifying samples as Type 1
- ➔ We can infer that it will have a high false positive score
- ➔ No data scaling nor any hyperparameter tuning

Confusion Matrix

		Predicted		
		0	1	2
Actual	0	3	14	0
	1	0	23	0
	2	0	13	1

	precision	recall	f1-score	support
0	1.00	0.18	0.30	17
1	0.46	1.00	0.63	23
2	1.00	0.07	0.13	14
accuracy			0.50	54
macro avg	0.82	0.42	0.35	54
weighted avg	0.77	0.50	0.40	54



Manually Tuned Model

جامعة نيويورك أبوظبي
NYU ABU DHABI

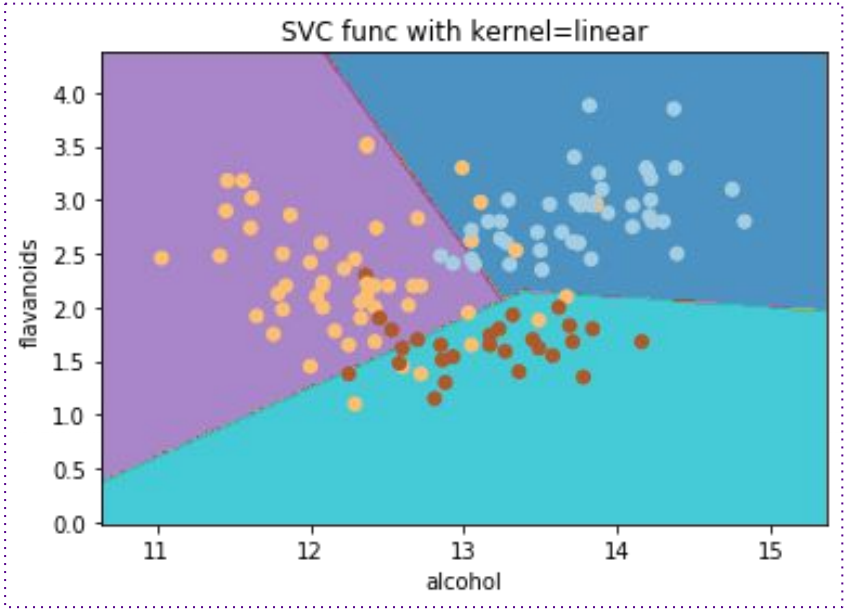
```
linear_svm_clf = Pipeline([( "scaler" ,
preprocessing.MinMaxScaler()) ,
( "svm_clf" , SVC(kernel='linear',
gamma='auto'))])
```

- ➔ Accurately classified all samples into categories
- ➔ We can infer that it will have a high true positive score
- ➔ Minmax data scaling but minimal hyperparameter tuning

Confusion Matrix

		Predicted		
		0	1	2
Actual	0	17	0	0
	1	0	23	0
	2	0	0	14

	precision	recall	f1-score	support
0	1.00	1.00	1.00	17
1	1.00	1.00	1.00	23
2	1.00	1.00	1.00	14
accuracy			1.00	54
macro avg	1.00	1.00	1.00	54
weighted avg	1.00	1.00	1.00	54



K-Cross Validation

جامعة نيويورك أبوظبي
NYU ABU DHABI

```
SVC with kernel=rbf Result
k=5 Cross Validation Mean: 0.4277
k=5 Cross Validation Variance: 0.0009
```

```
SVC with kernel=linear Result
k=5 Cross Validation Mean: 0.9663
k=5 Cross Validation Variance: 0.0005
```

```
SVC with kernel=linear GridSearch Result
k=5 Cross Validation Mean: 0.9944
k=5 Cross Validation Variance: 0.0001
```

- ➔ The baseline model has very little practical use with relatively high variance
- ➔ The manually tuned model has surprisingly high accuracy with a middle-ground variance
- ➔ The gridsearch tuned model has the best accuracy with and the lowest variance

GridSearch Tuned Model

جامعة نيويورك أبوظبي
NYU ABU DHABI

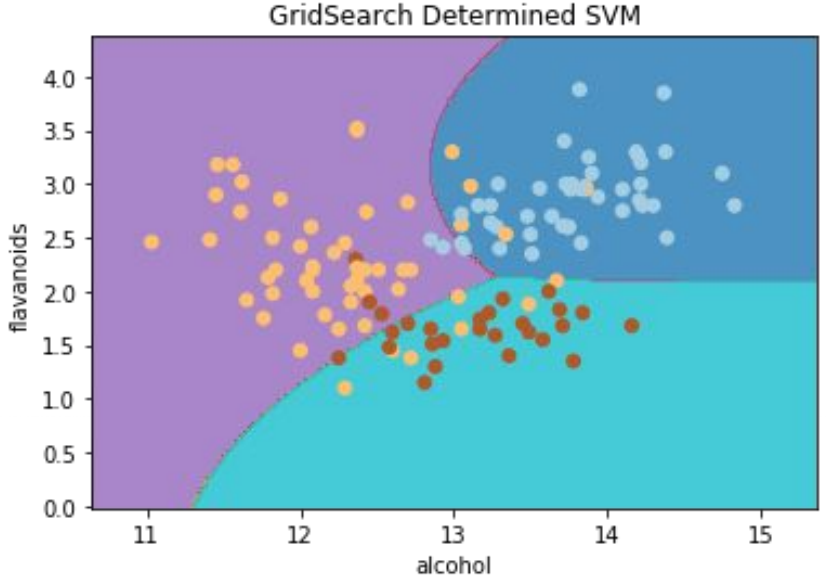
```
gridsearch_svm = Pipeline([("scaler", preprocessing.MinMaxScaler()), ("svm_clf", SVC(kernel='rbf', gamma=1, probability=True, C=3, decision_function_shape='ovr'))])
```

- ➔ Accurately classified all samples into categories
- ➔ We can infer that it will have a high true positive score
- ➔ Minmax data scaling and C, gamma, and kernel parameter tuning

Confusion Matrix

		Predicted		
		0	1	2
Actual	0	20	0	0
	1	0	23	0
	2	0	1	10

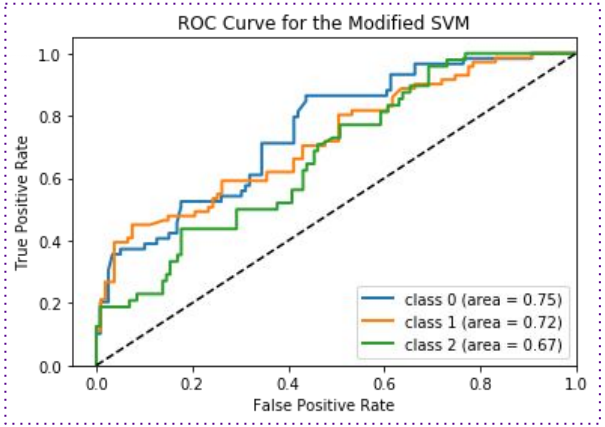
	precision	recall	f1-score	support
0	1.00	1.00	1.00	20
1	0.96	1.00	0.98	23
2	1.00	0.91	0.95	11
accuracy			0.98	54
macro avg	0.99	0.97	0.98	54
weighted avg	0.98	0.98	0.98	54



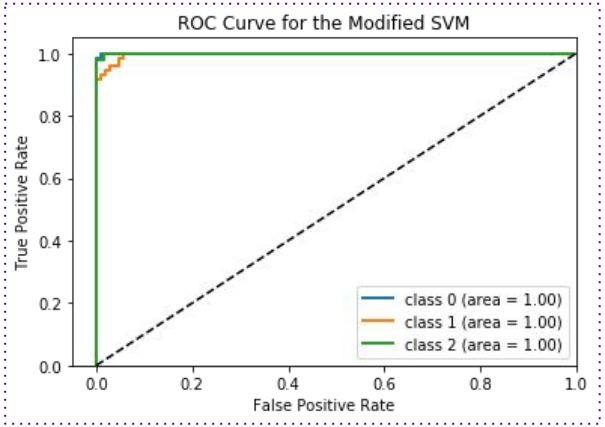
ROC Curves

جامعة نيويورك أبوظبي
NYU ABU DHABI

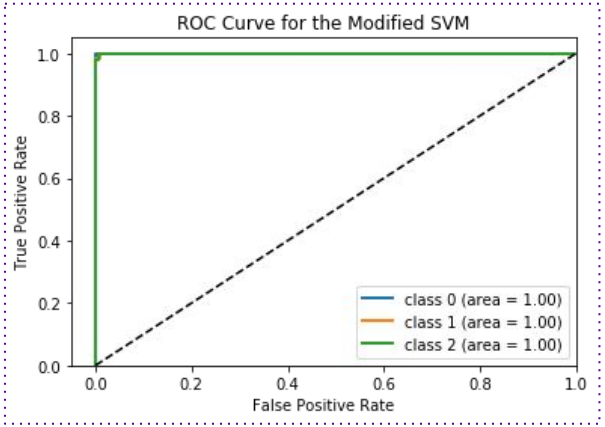
Baseline Model



Manually Tuned Model



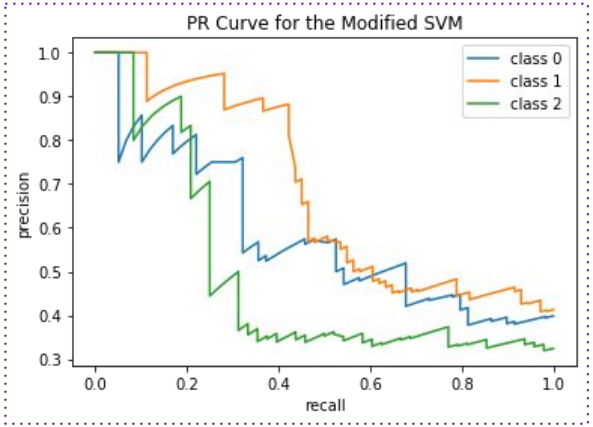
GridSearch Tuned Model



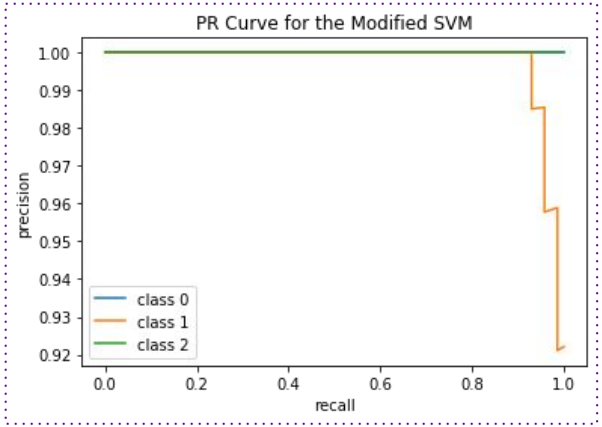
PR Curve

جامعة نيويورك أبوظبي
NYU ABU DHABI

Baseline Model



Manually Tuned Model



GridSearch Tuned Model