Team Assignment –KNN Classification - 100 pts

Partition the data into training (70%) and test (30%) sets

a) Perform a k-NN classification with all predictors. Choose automatic k selection (range of k: 3 to 25). Make sure to check the box to normalize input data. Weight features by importance when computing distances. Perform cross-validation and no feature selection

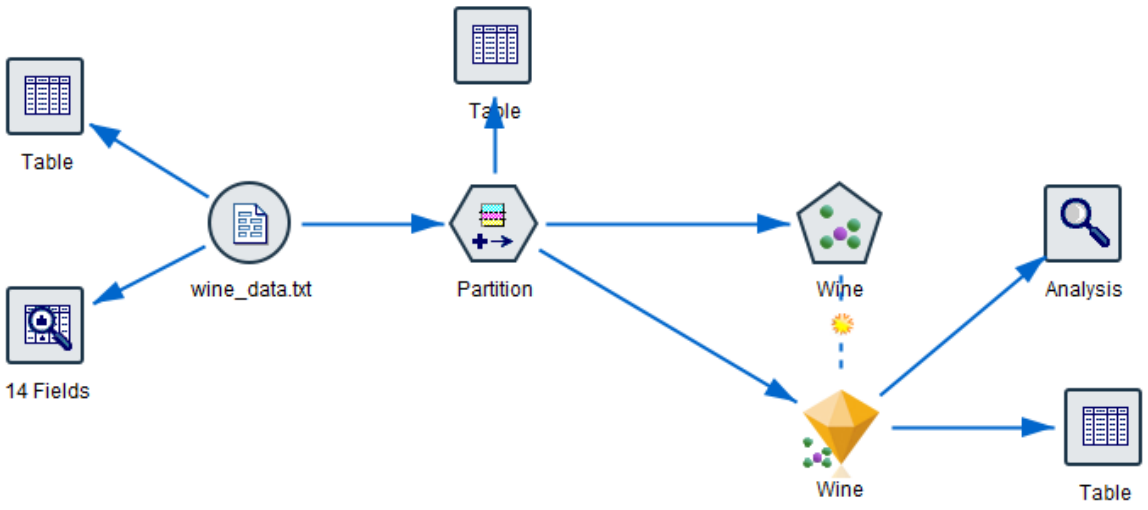
After importing the source file 'wine.txt', it is partitioned as 70% of data to training and 30% of data to testing the model.

A k-NN node is connected to this partitioned data. The node is adjusted to following settings:

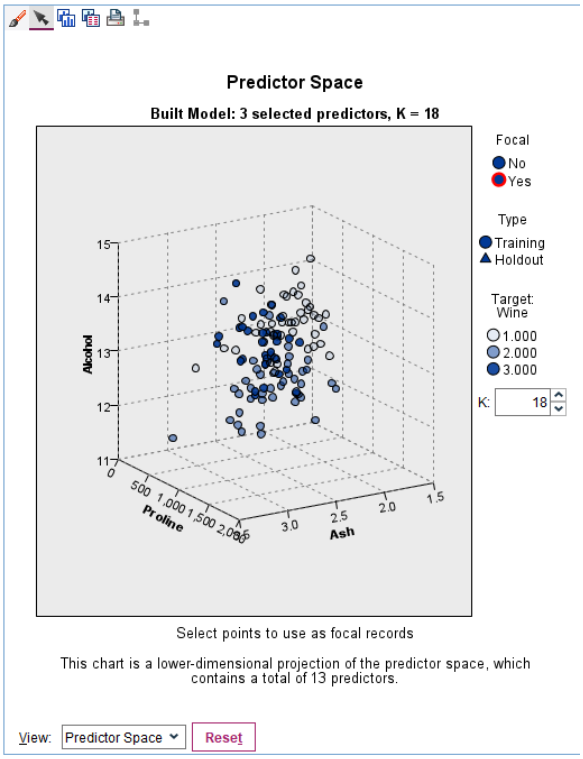
A target field is predicted on custom analysis as an objective.

Wine is selected as a target field with all other predictors as inputs.

The model is developed with reference to partitioned data with normalized input data. Automatic k selection is adjusted between 3 and 25 with Euclidean distance computation. The features are weighed by importance while computing the distance. Mean of nearest neighbor values is selected under predictions for target. No feature selection has been made with random cross-validation of 10 folds. Also, the data is analyzed to append all probabilities. All these selections are reflected in the attached stream.



A super node is generated when the model is executed with k as 18



When analysis node is attached to this super node, the following is displayed

Results for output field Wine

Comparing $KNN-Wine with Wine

'Partition'	1_Training		2_Testing	
Correct	117	96.69%	56	98.25%
Wrong	4	3.31%	1	1.75%
Total	121		57	

Coincidence Matrix for $KNN-Wine (rows show actuals)

'Partition' = 1_Training	1	2	3
1	43	0	0
2	2	43	2
3	0	0	31
'Partition' = 2_Testing	**1**	**2**	**3**
1	16	0	0
2	0	23	1
3	0	0	17

The above result shows the that accuracy for test data is 98.25% which is larger than training accuracy 96.69% (so no overfitting)

b) Let's assume you get a new record with these values. What is your prediction of the type of wine (1,2,3) that this data corresponds to?

Alcohol 13.79

Malic_acid 1.32

Ash 2.8

Alcalinity_of_ash 20.5

Magnesium 134

Total_phenols 2.75

Flavanoids 2.62

Nonflavanoid_phenols 0.53

Proanthocyanins 1.45

Color_intensity 5.5

Hue 1.35

OD280_OD315_of_diluted_wines 3.1

Proline 1217

The above data is stored in as a sample.txt file and scoring is done for prediction.

Prediction denotes under which type of wine the given combinations go. The result is

$KNN-Wine	$KNNP-Wine	$KNNP-1	$KNNP-2	$KNNP-3
1	0.905	0.905	0.048	0.048

The above table give the prediction as such that the given combinations denote the wine type 1 wit probability of 0.905. and so, the type 2 and type 3 probability being very small which is 0.048.