Support Vector Machine based Classification

Objective

Use the given data set to build SVM based classification model. Also, analyze the difference on outcome by varying different parameters of SVM generation.

Classification using SVM Node

The SPSS Modeler offers two type of modes for SVM model-

- Simple mode
- Expert mode

Partitioning the data-

1. Select a *"Partition Node"* from *"Field Ops"* tab in nodes palette and connect source node to this node.
2. Double click on partition node and go to settings tab.
3. Name the "Partition Field" as *Partition Field*. Select *"Train and test"* and specify "70:30" ratio for training and testing data splits.
4. Click *"Apply"* and *"OK"*. Now, a Partition will be visible on the modeler canvas.
5. [Optional] To view which records have been marked as "training record" and which as "testing record", add a *"Table node"*, *"Connect"* it to partition node and *"Run"* the stream.

Building SVM model in Simple mode-

1. Select a *"SVM Node"* from *"Classification"* subgroup under *"Modeling"* tab in nodes palette and connect *Training select* node to this node.
2. Go to *"Model"* tab and select *"Use partitioned data"* as *Output type.*
3. Go to *"Expert"* tab and select *Mode* as *"Simple"*. Go to *"Annotations"* tab and rename the model as *"SVM Simple"*.
4. Click *"Apply"* and *"OK"*. Now, a SVM node will be visible on the modeler canvas**.**
5. *"Run"* this stream from SVM node, a *"SVM Model Nugget"* will be created and placed on stream canvas.

Building SVM model in Expert mode-

1. Repeat steps 1 and 2 as explained above.
2. Go to *"Annotations"* tab and rename the model as *"SVM Expert"*.
3. Go to *"Expert"* tab and select *Mode* as *"Expert"*.
4. Change different parameters to increase the *accuracy* of model.

5. *"Run"* this stream from SVM node, a *"SVM Model Nugget"* will be created and placed on stream canvas.
6. [Optional]Connect a *Analysis* node to this nugget and see the partition wise classification error.

Table 1- Explanation of different expert mode parameters of SVM model

Parameter	Explanation
Stopping criteria	Determines when to stop the optimization algorithm. Values range from 1.0E–1 to 1.0E–6; default is 1.0E–3. Reducing the value results in a more accurate model, but the model will take longer to train.
Regularization parameter (C)	Controls the trade-off between maximizing the margin and minimizing the training error term. Value should normally be between 1 and 10 inclusive; default is 10. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting.
Regression precision (epsilon)	Used only if the measurement level of the target field is Continuous. Causes errors to be accepted provided that they are less than the value specified here. Increasing the value may result in faster modeling, but at the expense of accuracy.
Kernel type	Determines the type of kernel function used for the transformation. Different kernel types cause the separator to be calculated in different ways, so it is advisable to experiment with the various options. Default is RBF (Radial Basis Function).
RBF gamma	Enabled only if the kernel type is set to RBF. Value should normally be between 3/k and 6/k, where k is the number of input fields. For example, if there are 12 input fields, values between 0.25 and 0.5 would be worth trying. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting.
Gamma	Enabled only if the kernel type is set to Polynomial or Sigmoid. Increasing the value improves the classification accuracy (or reduces the regression error) for the training data, but this can also lead to overfitting.
Bias	Enabled only if the kernel type is set to Polynomial or Sigmoid. Sets the coef0 value in the kernel function. The default value 0 is suitable in most cases.
Degree	Enabled only if Kernel type is set to Polynomial. Controls the complexity (dimension) of the mapping space. Normally you would not use a value greater than 10.

Comparing performance-

6. Select a *"Merge"* node from *"Record Ops"* tab in nodes palette.
7. Connect *Simple mode* and *Expert mode* model nuggets to this *Merge* node.

8. Go to *Merge* tab of merge node and select *"Keys"* as Merge Method. Select all possible keys except predicted target field and its probability field (Fields starting with $S-sign).
9. Connect an *Analysis* node to the merge node. *Run* the stream
10. Carefully analyze the results i.e. performance of individual models and their comparison with original value of target field which was given in the dataset.
11. Vary the parameters on *Expert* node to generate better models.

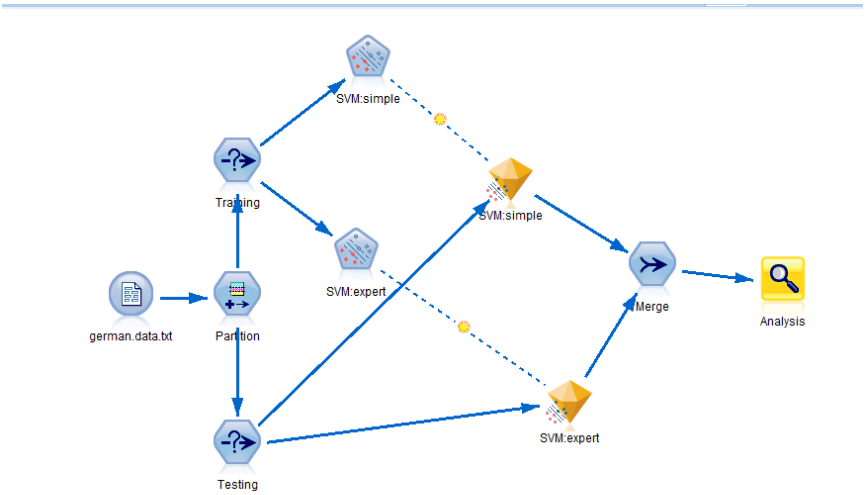

Fig 1- Stream for comparing SVM models