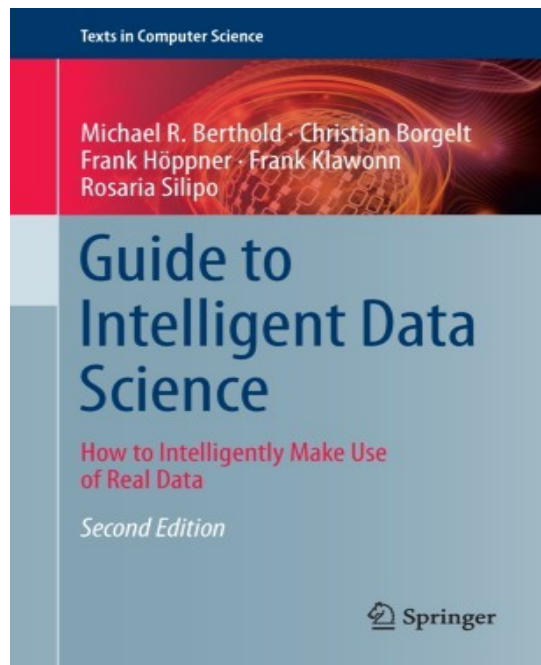




Big Data ACUA6002 Tutorial 8

Support Vector Machines

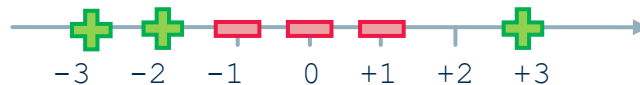
Exercise 1

1. SVM

– Given is a one dimensional data set in R .

Data points for $+1$ are at $\{-3, -2, 3\}$ and data points for -1 are at $\{-1, 0, 1\}$

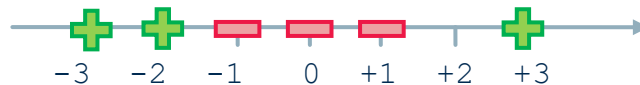
- Can this data set in its current feature space be separated using a linear separator? Why/why not?
- Define a kernel function to transform the data points in a space where the transformed data points are linearly separable



1. SVM

Can this data set in its current feature space be separated using a linear separator? Why/why not?

- From the plot of the data is clear that the two classes cannot be separate using a single straight line



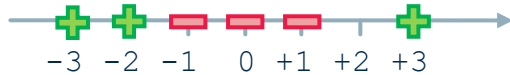
1. SVM

Define a kernel function to transform the data points in a space where the transformed data points are linearly separable

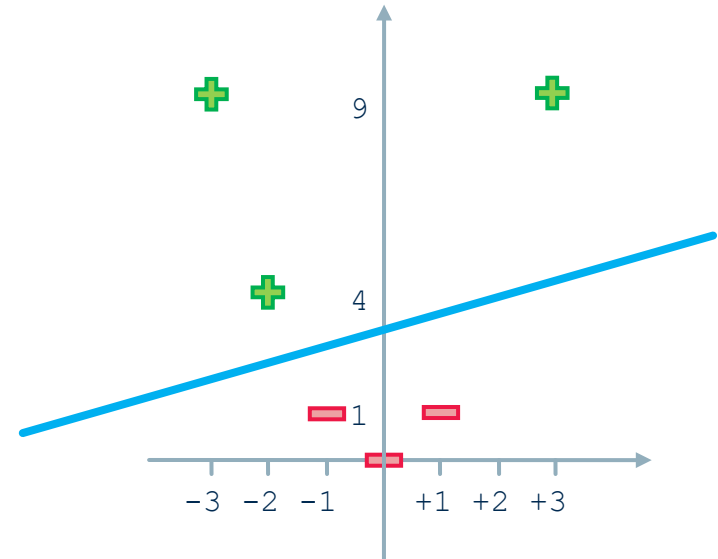
- **Recap**
- Discriminant function in projected space:
- And, using a kernel K :

1. SVM

Define a kernel function to transform the data points in a space where the transformed data points are linearly separable



$$\phi(x) = (x, x^2)$$



Exercise 2

2. SVM

- *Given is the following two dimensional data set () together with its classification*
- *Apply the Primal Perceptron Algorithm by Rosenblatt. Give the definition of the resulting hyperplane. Paint the data set as well as the hyperplane into a diagram*

2. SVM

Apply the Primal Perceptron Algorithm by Rosenblatt.

- **Recap**
- Learning Rule:

IF THEN

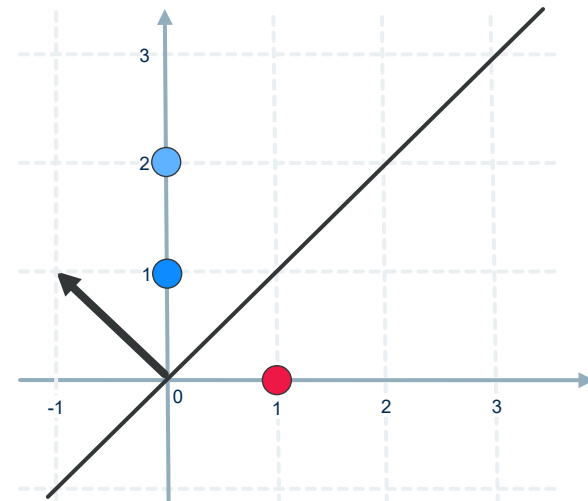
2. SVM

Apply the Primal Perceptron Algorithm by Rosenblatt.

- Initialization:
- First Iteration:
 - no change
- Second iteration:
 - No changes (Algorithm terminates)

Solution:

- Weight vector
- Offset



Exercise 3

3. SVM

- *Given are two training data in as in the following figure. What is the separating hyperplane? Define the weight vector and b in a way that the distance between the two hyperplanes is maximal.*



3. SVM

What is the separating hyperplane?

- Define “margins” of individual training examples:
- The margin of a hyperplane (with respect to a training set):



- Examples: and
- Decision function:
- Solution hyperplane needs to fulfil:
 - as well
 - Margin: with ()
 - Maximize
 - Maximize with

Exercise 4

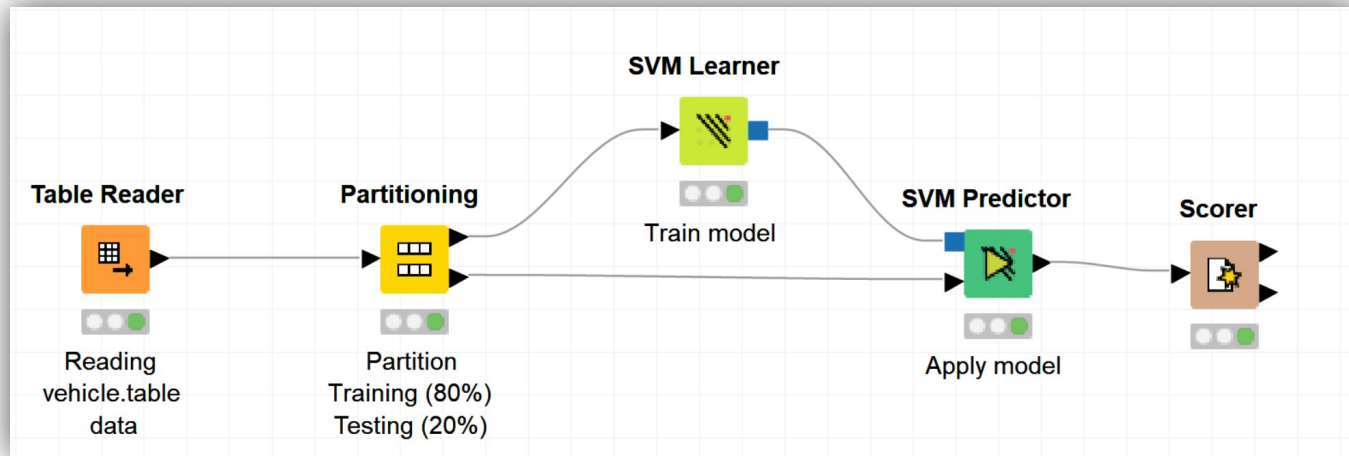
Practice with KNIME

SVM in KNIME

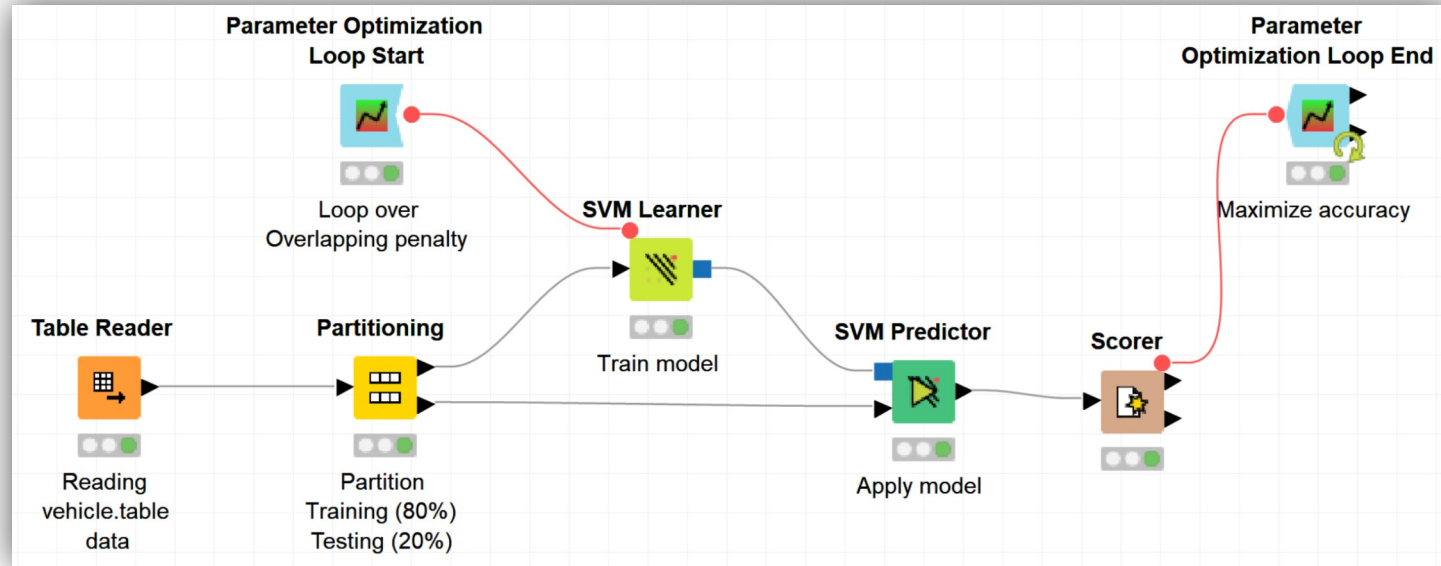
Train a classifier to predict the vehicle type. The attributes are numerical, and have been normalized to [0,1] scale. Use a support vector machine (SVM) as a classification model. Explore an optimal parameter for the SVM setting.

1. Read the data (*vehicle.table*) with a *Table Reader* node. This dataset contains 2D attributes of silhouettes created by different types of vehicles. The last column is the target column, describing the 4 types of vehicles (bus, Opel, Saab, and van)
2. Split the data with a *Partitioning* node into the training (80%) and testing (20%) data sets.
 - Use stratified sampling to preserve the target distribution on Col18.
3. Train an SVM classifier with the *SVM Learner* node. Use the RBF kernel with $\sigma=0.5$.
4. Use the trained model to predict the vehicle type with the *SVM Predictor* node
5. Assess the classification outcome by examining the confusion matrix with the *Scorer* node.
6. Determine the optimal margin hardness by varying the Overlapping penalty in the SVM learner node with a *Parameter Optimization Loop*.
 - For the *Parameter Optimization Loop Start*, create a parameter “c” ranging from 1 to 25, with the step size of 2.
 - Use the parameter c to control “c_parameter” in the *SVM Learner* flow variable.
 - Connect the flow variable output from the *Scorer* node to the *Parameter Optimization Loop End*. Maximize the Accuracy.

Train a classifier to predict the vehicle type. The attributes are numerical, and have been normalized to $[0,1]$ scale. Use a support vector machine (SVM) as a classification model.



Determine the optimal margin hardness by varying the Overlapping penalty in the SVM learner node with a Parameter Optimization Loop.



Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F. and Silipo, R. (2020). *Guide to Intelligent Data Science: How to Intelligently Make Use of Real Data*. 2nd ed. [online] www.springer.com. Springer International Publishing.

Available at:

<https://www.springer.com/gp/book/9783030455736> [Accessed 17 Jun. 2021].

Questions?