Title: Four plots for SVMs

The general form of the mathematics for a linear support vector machine are quite straightforward.

Support Vector Machines are Supervised tools.

SEE: https://rpubs.com/mzc/mlwr_svm_concrete (https://rpubs.com/mzc/mlwr_svm_concrete)

Therefore, given a set of labeled pairs of data:

$(x_i, y_i), \quad i = 1, \ldots, l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$

Such that $f(x_i) = \begin{Bmatrix} \geq 0; \; y_i = 1 \\ < 0; \; y_i = -1 \end{Bmatrix}$

where y is a set of two values which indicate a label, e.g. true or false.

min $\frac{1}{2} W^T W \quad + \quad C \sum_i^l = {}_1 \, \xi_i$

subject to $y_i(w^T \phi(x_i) \; + \; b) \geq (1 - \xi_i)$ where $\xi_i \geq 0,$

C is the penalty parameter, *cost*, of the error term, such that $C > 0$.

Furthermore, $K(x_i, x_j) \quad \equiv \quad \phi(x_i)^T \phi(x_j)$ is called the kernel function.

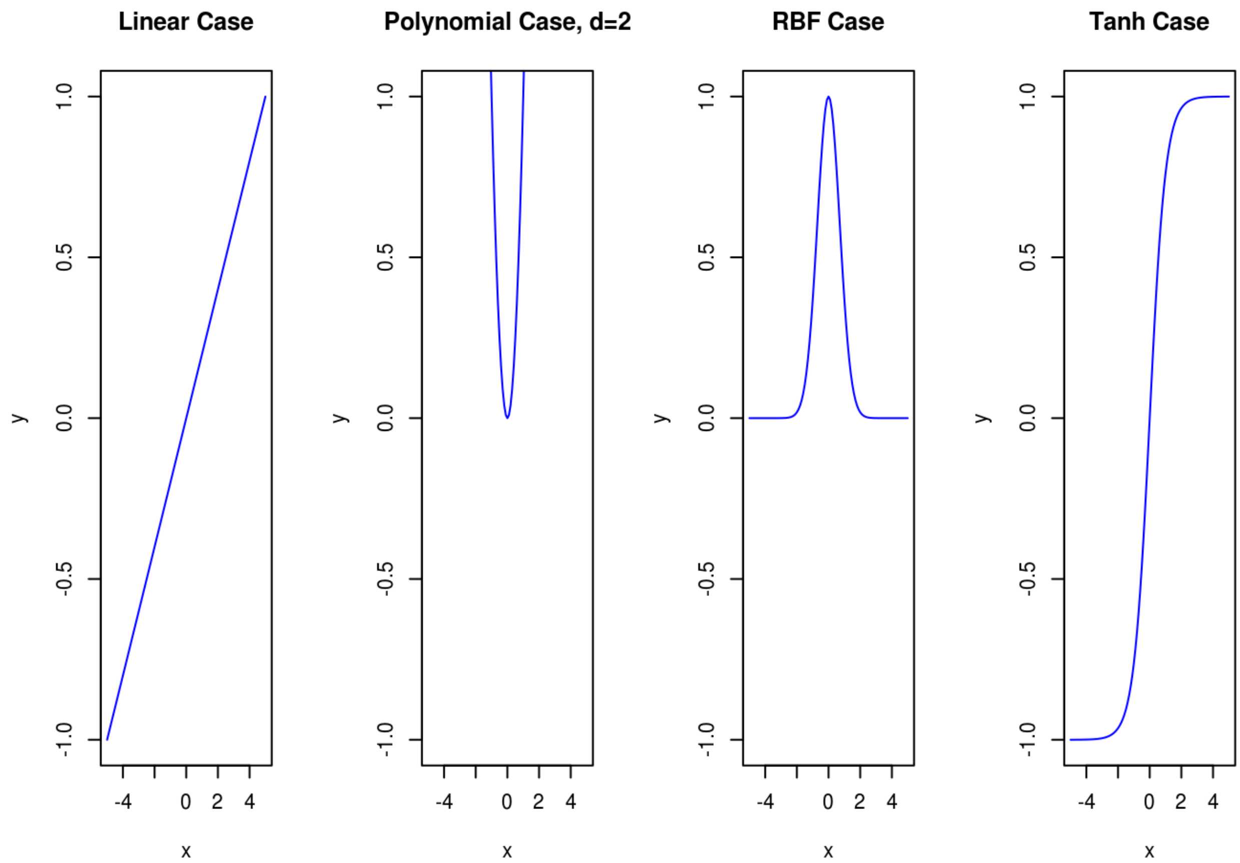
The 4 most common SVM formulae are:

1. Linear: $K(x_i, \; y_j) \quad = \quad < x, \; y >$
   - The linear kernel does not transform the data at all.
2. Polynomial: $K(x_i, x_j) \quad = \quad (\gamma \, x_i^T \, x_j \quad + \quad r)^d, \; \gamma > 0$
   - The polynomial kernel has a simple non-linear transform of the data.
3. Radial Basis Function (RBF): $K(x_i, x_j) \quad = \quad exp(-\gamma \parallel x_i^T - x_j \parallel^2), \; \gamma > 0$
   - The Guassian RBF kernel which performs well on many data and is a good default
4. Sigmoid: $K(x_i, x_j) \quad = \quad \tanh(\gamma \, x_i^T \, x_j \quad + \quad r), \; \gamma > 0$
   - The sigmoid kernel produces a SVM analagous to the activation function similar to a perceptron (https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Neuron/index.html) with a sigmoid activation function.[1]

- Such that $\gamma, \; r$ and d are kernel parameters.

There are no reliable rules for which kernel to use with any given data set.

Plots for 4 most common SVM formulae:



EOF

---

1. https://rpubs.com/mzc/mlwr_svm_concrete (https://rpubs.com/mzc/mlwr_svm_concrete)↩