Support Vector Machines

Exercises

1. Mark each statement as *True/False.* If a statement is False, explain why.

   a. Support Vector Machines are generally classified as black-box approaches that make good predictions but provide little insight into underlying variable relationships.

   b. The linear kernel in an SVM really means that no kernel is used - the resulting model is merely a separating hyperplane in the input space.

   c. Using a kernel in an SVM is a way to introduce non-linearity to the decision boundary.

   d. Standardization of your data is not important for a SVM - it makes no difference if all of your variables are on vastly different scales.

   e. The most commonly used kernel for SVMs is the polynomial kernel.

   f. One problem with using kernel functions is the computational complexity - they don't work well in situations with many observations.

   g. For text classification, a linear kernel typically performs best.

2. Describe the radial basis function (Gaussian similarity) mathematically. Explain how to interpret the formula in terms of the 'center' of the function. Is the function value larger or smaller for points that are farther away from the center?

3. Describe the effect of the radial basis function parameter σ on the overall model created by an SVM. In particular, how does the parameter interact with the bias-variance tradeoff?

   Text

4. Describe the effect of the cost parameter C on the overall model created by an SVM. In particular, how does the parameter interact with the bias-variance tradeoff?

List of Key Terms

Support Vector Machine

Support Vector

Margin

Kernel

Radial Basis Function

Regularization

Parameter C

RBF Parameter γ, σ