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.

   True

   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.

   True

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

   True

   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.

   False

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

   False

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

   True

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

   True

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?

   e^(-gamma*||x - center||) as x gets further from the 'center' point, the distance gets larger and we're looking at 1 divided by e raised to that large distance (because the exponent is negative) so the similarity gets much smaller

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?

   when sigma = 1/gamma gets large, the similarity of a single point reaches farther more training points have more effect on the boundary — this prevents overfitting, whereas when sigma is small, the decision boundary may be decided by only a few training observations.

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?

   When C gets larger (lambda gets smaller) we're using less regularization and allowing the models to put more weight on error in the objective function, meaning they may start to overfit the training data (more variance). When C gets smaller, we have more bias and are less likely to overfit.

List of Key Terms

- Support Vector Machine
- Support Vector
- Margin
- Kernel
- Radial Basis Function
- Regularization
- Parameter C
- RBF Parameter γ, σ