Problem Bank: SVM

November 27, 2019

1. **Moving support vectors and non-support vectors**

   State whether each of the following statements about maximal margin classifiers is true or false:

   (a) If you move the support vectors, the maximal margin hyperplane will change.

   (b) You can move the non-support vectors *anywhere* in the feature space, and the maximal margin hyperplane will not be affected.

2. **Separable vs. non-separable case**

   Imagine a situation where the two classes can be separated by a hyperplane.

   (a) In this situation, is it always the best option to use the separating hyperplane to construct your classifier? Why or why not?

   (b) What is the meaning of a 'soft'- vs. 'hard'- margin classifier?

3. **Robustness**

   Support vector classifiers are generally quite robust to observations that are far away from the hyperplane, which is different than many other classification methods. Why is this the case?

4. **Bias-variance tradeoff**

   (a) Conceptually, why does C affect the bias-variance tradeoff?

   (b) In terms of the optimization problem, what specifically does C control? Make sure to analyze the actual terms in the optimization problem.

5. **Non-linear boundaries + Kernels**

(a) Explain why enlarging the feature space can lead to non-linear classification boundaries.

(b) Why are kernels used in SVMs when enlarging the feature space? (Just explain in 1 sentence; no need to explain the math behind it.)

6. **SVMs with more than two classes**

Two common approaches are one-versus-one classification and one-versus-all classification. Comment on how these methods affect the:

(a) Computational cost

(b) Interpretability

7. **SVM vs. Logistic Regression**

   What are the loss functions used by SVMs and logistic regression? How do these compare? Consider looking at Figure 9.12 in ISL.

Ch. 9 SVM - Problem Bank Questions

November 17, 2019

1. SVM true and false. For each, explain your answer.
(a) T/F: In two dimensions, a hyperplane is simply a line.
(b) T/F: We can think of a p dimensional hyperplane as dividing the space into $p-1$ parts
(c) T/F: For data in 2d, there is always a unique hyperplane that can perfectly separate the data.
(d) T/F: When $y_i \in \{-1, 1\}$, then we have that a separating hyperplane satisfies

$$y_i(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) > 0.$$

(e) T/F: A classifier built on a separating hyperplane is a non-linear classifier.
(f) T/F: The maximal margin classifier is the hyperplane that has the farthest average distance to the training observations.
(g) T/F: Support vectors are the data points that are the farthest from the separating hyperplane.

2. Explain the bias-variance tradeoff for the parameter C in the SVM optimization problem. What happens when $C \to 0$? What happens when $C \to \infty$? Explain the relationship between C and the number of support vectors. Why does this intuitively make sense for a classifier?

3. We can write the equation of a separating hyperplane as:

$$\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p = 0.$$

Find an expression for the distance from any point to this hyperplane.

4. Explain in laymen's terms what the role of slack variables are and why we introduce them into the optimization problem for solving SVMs.

5. Kernel Methods: Calculate the Gram matrix for the following data points and kernel. The Gram matrix is defined as an N x N matrix, where entry i, j in the matrix is $K(X_i, X_j)$.

$$X = [1, 2, 3, 4, 5].$$

$$K(x_i, x_j) = \exp(-0.1(x_i - x_j)^2)$$

6.Multiple Classes

(a) What are two ways that we can handle more than two classes with SVMs?
(b) What are some pros and cons of each approach?