

Regression and SVM

October 26, 2015

Abstract

1 Logistic Regression

2 Support Vector Machines

Support Vector Machines are a machine-learning technique that is used to classify binary sets of data. We define θ as the normal to the decision hyperplane, and classify data points by $\text{sgn}(\theta^\top x + \theta_0)$. In order to train the classifier, we initially set up the minimization problem (known as **Hard-SVM**):

$$\max_{\theta, \theta_0} \frac{1}{\|\theta\|} \min_{1 \leq i \leq n} y^{(i)}(\theta^\top x^{(i)} + \theta_0) \quad (1)$$

However, we cannot satisfy this minimization if the training set is not linearly separable. Therefore, we add a “slack variable” to each constraint, which is a measure of the “wrongness” of the decision boundary when classifying that point. We call these slack variables ξ_i . We then desire to minimize

$$\min_{\theta, \theta_0, \xi} \frac{\lambda}{2} \|\theta\|^2 + \frac{1}{n} \sum_{i=1}^n \xi_i, \quad (2)$$

$$s.t. \quad y^{(i)}(\theta^\top x^{(i)} + \theta_0) \geq 1 - \xi_i, \quad (3)$$

$$\xi_i \geq 0, i \in n \quad (4)$$

However, we find it more computationally efficient to solve the dual form of the SVM:

$$\max_{\alpha \in \mathbb{R}^n} \left[\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^\top x^{(j)} \right] \quad (5)$$

$$s.t. \quad 0 \leq \alpha_i \leq C \quad (6)$$

$$\sum_i \alpha_i y^{(i)} = 0 \quad (7)$$

To provide some intuition: α_i is the weight of each training point on the final decision boundary. It is 0 for all $x^{(i)}$ that are farther from the decision boundary than the margin, so that only the “important” $x^{(i)}$ are “support vectors”. C is the maximum value of α , and determines the size of the margin. **ADD SOMETHING ABOUT WHICH WAY THAT WORKS HERE.** To return from α to the more familiar θ and θ_0 , we plug in:

$$\theta = \sum_{i=1}^n \alpha_i y^{(i)} x^{(i)} \quad (8)$$

$$\theta_0 = \frac{1}{\mathcal{M}} \left[\sum_{j \in \mathcal{M}} \left(y^{(j)} - \sum_{i \in \mathcal{S}} \alpha_i y^{(i)} (x^{(j)})^\top x^{(i)} \right) \right] \quad (9)$$

Where $\mathcal{M} = \{i : 0 < \alpha_i < C\}$ and $\mathcal{S} = \{i : 0 < \alpha_i\}$.