Fundamentals of Linear Algebra and Optimization

Classification of the Data Points in Terms of nu

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

May 6, 2020

Classification of Data Points for $(\mathrm{SVM}_{s2'})$

For a finer classification of the points it turns out to be convenient to consider the ratio

$$\nu = \frac{K_m}{(p+q)K_s}.$$

First note that in order for the constraints to be satisfied, *some* relationship between K_s and K_m must hold. In addition to the constraints

$$0 \leq \lambda_i \leq K_s, \quad 0 \leq \mu_j \leq K_s,$$

Classification of Data Points for $(\mathrm{SVM}_{s2'})$

we also have the constraints

$$\sum_{i=1}^{p} \lambda_i = \sum_{j=1}^{q} \mu_j$$

$$\sum_{i=1}^{p} \lambda_i + \sum_{j=1}^{q} \mu_j \geq K_m,$$

which imply that

$$\sum_{i=1}^{p} \lambda_i \geq \frac{K_m}{2} \quad \text{and} \quad \sum_{j=1}^{q} \mu_j \geq \frac{K_m}{2}. \qquad (\dagger)$$

Relationship Between K_s and K_m

Since λ, μ are all nonnegative, if $\lambda_i = K_s$ for all i and if $\mu_j = K_s$ for all j, then

$$\frac{K_m}{2} \leq \sum_{i=1}^{p} \lambda_i \leq pK_s \quad \text{and} \quad \frac{K_m}{2} \leq \sum_{j=1}^{q} \mu_j \leq qK_s,$$

so these constraints are not satisfied unless $K_m \leq \min\{2pK_s, 2qK_s\}$, so we assume that $K_m \leq \min\{2pK_s, 2qK_s\}$.

Definition of ν for $(\mathrm{SVM}_{s2'})$

The equations in $(\dagger)$ also imply that there is *some* i_0 such that $\lambda_{i_0} > 0$ and *some* j_0 such that $\mu_{j_0} > 0$, and so $p_m \geq 1$ and $q_m \geq 1$.

For a finer classification of the points we find it convenient to define $\nu > 0$ such that

$$\nu = \frac{K_m}{(p+q)K_s},$$

so that the objective function $J(w, \epsilon, \xi, b, \eta)$ is given by

$$J(w, \epsilon, \xi, b, \eta) = \frac{1}{2} w^{\top} w + (p+q)K_s \left(-\nu\eta + \frac{1}{p+q} \begin{pmatrix} \epsilon^{\top} & \xi^{\top} \end{pmatrix} \mathbf{1}_{p+q} \right).$$

Normalization of ν for $(\mathrm{SVM}_{s2'})$

Observe that the condition $K_m \leq \min\{2pK_s, 2qK_s\}$ is equivalent to

$$\nu \leq \min\left\{\frac{2p}{p+q}, \frac{2q}{p+q}\right\} \leq 1.$$

Since we obtain an equivalent problem by rescaling by a common positive factor, theoretically it is convenient to normalize K_s as

$$K_s = \frac{1}{p+q},$$

in which case $K_m = \nu$.

This method is called the *ν-support vector machine*.

Classification of Data Points for $(\mathrm{SVM}_{s2'})$

Actually, to program the method, it may be more convenient assume that K_s is arbitrary. This helps in avoiding λ_i and μ_j to become to small when $p + q$ is relatively large.

The equations $(\dagger)$ and the box inequalities

$$0 \leq \lambda_i \leq K_s, \quad 0 \leq \mu_j \leq K_s$$

also imply the following facts:

Classification of Data Points for $(\mathrm{SVM}_{s2'})$

Proposition. If Problem $(\mathrm{SVM}_{s2'})$ has an optimal solution with $w \neq 0$ and $\eta > 0$, then the following facts hold:

(1) Let p_f be the number of points u_i such that $\lambda_i = K_s$, and let q_f the number of points v_j such that $\mu_j = K_s$. Then $p_f, q_f \leq \nu(p+q)/2$.

(2) Let p_m be the number of points u_i such that $\lambda_i > 0$, and let q_m the number of points v_j such that $\mu_j > 0$. Then $p_m, q_m \geq \nu(p+q)/2$. We have $p_m \geq 1$ and $q_m \geq 1$.

(3) If $p_f \geq 1$ or $q_f \geq 1$, then $\nu \geq 2/(p+q)$.

Condition for Separablity of Data Points

Observe that $p_f = q_f = 0$ means that there are no points in the open slab containing the separating hyperplane, namely, the points u_i and the points v_j are separable.

So if the points u_i and the points v_j are not separable, then we must pick ν such that $2/(p+q) \leq \nu \leq \min\{2p/(p+q), 2q/(p+q)\}$ for the method to succeed. Otherwise, the method is trying to produce a solution where $w = 0$ and $\eta = 0$, and it does not converge (γ is nonzero).

Upper and Lower Bounds for ν of $(\mathrm{SVM}_{s2'})$

Actually, above Proposition yields more accurate bounds on ν for the method to converge, namely

$$\max\left\{\frac{2p_f}{p+q}, \frac{2q_f}{p+q}\right\} \leq \nu \leq \min\left\{\frac{2p_m}{p+q}, \frac{2q_m}{p+q}\right\}.$$

By a previous remark, $p_f \leq p_m$ and $q_f \leq q_m$, the first inequality being strict if there is some i such that $0 < \lambda_i < K$, and the second inequality being strict if there is some j such that $0 < \mu_j < K$. This will be the case under the **Standard Margin Hypothesis**.

Value of ν Controls Width of Slab

Observe that a small value of ν keeps p_f and q_f small, which is achieved if the δ-slab is narrow (to avoid having points on the wrong sides of the margin hyperplanes).

A large value of ν allows p_m and q_m to be fairly large, which is achieved if the δ-slab is wide.

Thus the smaller ν is, the narrower the δ-slab is, and the larger ν is, the wider the δ-slab is.