Fundamentals of Linear Algebra and Optimization

Introduction to Soft Margin Support Vector Machines

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

May 6, 2020

SVM Separation Problem

In the previous module we considered the problem of separating two nonempty disjoint finite sets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$ in $\mathbb{R}^n$.

The goal is to find a hyperplane H of equation $w^\top x - b = 0$ (where $w \in \mathbb{R}^n$ is a nonzero vector and $b \in \mathbb{R}$), such that all the blue points u_i are in one of the two open half-spaces determined by H, and all the red points v_j are in the other open half-space determined by H; see Figure 1.

Two Examples of SVM Separation Problem

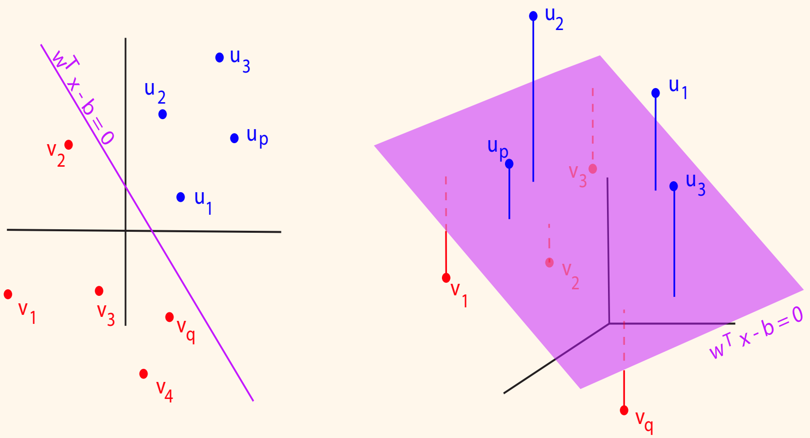


Figure 1: Two examples of the SVM separation problem.

SVM Soft Margin Problem

SVM picks a hyperplane which *maximizes the minimum distance* from these points to the hyperplane.

In this module we return to the problem of separating two disjoint sets of points, $\{u_i\}_{i=1}^{p}$ and $\{v_j\}_{j=1}^{q}$, but this time *we do not assume that these two sets are separable*.

To cope with nonseparability, we allow points to invade the safety zone around the separating hyperplane, and even points on the wrong side of the hyperplane. Such a method is called *soft margin support vector machine*.

Soft Margin Support Vector Machines

We discuss variations of this method, and in each case we present the dual.

It turns out that the soft margin SVM arising from Problem (SVM_{h1}) has some problem (a potential division by 0). Soft margin SVMs arising from Problem (SVM_{h2}) do not suffer from this problem.

Soft Margin Support Vector Machines

If the sets of points $\{u_1, \ldots, u_p\}$ and $\{v_1, \ldots, v_q\}$ are not linearly separable (with $u_i, v_j \in \mathbb{R}^n$), we can use a trick from linear programming which is to introduce nonnegative "slack variables" $\epsilon = (\epsilon_1, \ldots, \epsilon_p) \in \mathbb{R}^p$ and $\xi = (\xi_1, \ldots, \xi_q) \in \mathbb{R}^q$ to relax the "hard" constraints

$$\begin{aligned} w^\top u_i - b &\geq 1 & i &= 1, \ldots, p \\ -w^\top v_j + b &\geq 1 & j &= 1, \ldots, q \end{aligned}$$

of Problem (SVM_{h2}) to the "soft" constraints

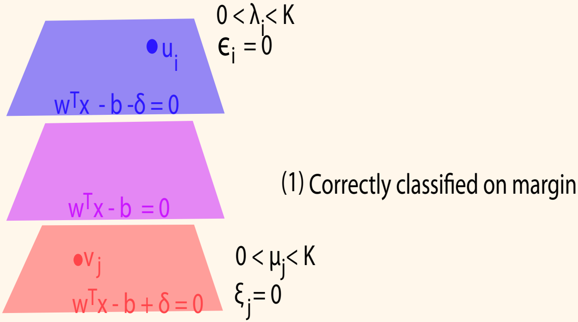
$$\begin{aligned} w^\top u_i - b &\geq 1 - \epsilon_i, & \epsilon_i &\geq 0 & i &= 1, \ldots, p \\ -w^\top v_j + b &\geq 1 - \xi_j, & \xi_j &\geq 0 & j &= 1, \ldots, q. \end{aligned}$$

Soft Margin Support Vector Machines

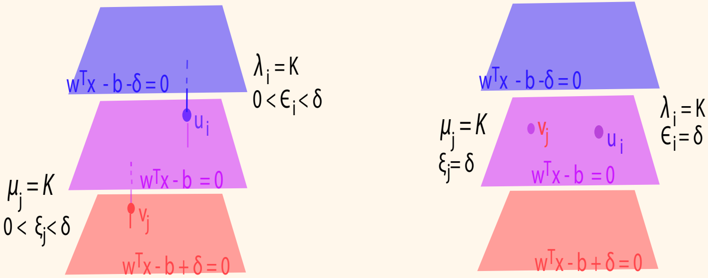
In this case there is no constraint on w, and we minimize $(1/2)w^\top w$. The margin is $\delta = 1/\|w\|$.

If $\epsilon_i > 0$, the point u_i may be misclassified, in the sense that it can belong to the margin (the slab), or even to the wrong half-space classifying the negative (red) points. See Figures (2) and (3) in the following illustrations.

Correctly Classified Points for Soft SVM

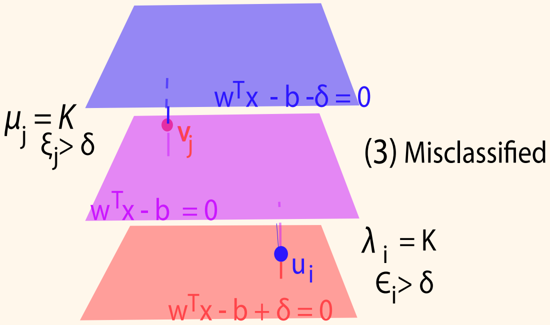


Correctly Classified Points for Soft SVM



(2) Correctly classified in slab

Misclassified Points for Soft SVM



Point Classifications for Soft Margin SVM

Figure (1) illustrates the case of u_i contained in the margin and occurs when $\epsilon_i = 0$. Figure (1) also illustrates the case of v_j contained in the margin when $\xi_j = 0$. The left illustration of Figure (2) is when u_i is inside the margin yet still on the correct side of the separating hyperplane $w^\top x - b = 0$. Similarly, v_j is inside the margin on the correct side of the separating hyperplane. The right illustration depicts u_i and v_j on the separating hyperplane. Figure (3) illustrates a misclassification of u_i and v_j.

Point Classifications for Soft Margin SVM

Similarly, if $\xi_j > 0$, the point v_j may be misclassified, in the sense that it can belong to the margin (the slab), or even to the wrong half-space classifying the positive (blue) points.

We can think of ϵ_i as a measure of how much the constraint $w^\top u_i - b \geq 1$ is violated, and similarly of ξ_j as a measure of how much the constraint $-w^\top v_j + b \geq 1$ is violated.

Soft Margin SVM Hyperplane Terminology

The hyperplane $H_{w,b}$ of equation

$$w^\top x = b$$

is called the *separating hyperplane*.

The hyperplane $H_{w,b+1}$ of equation

$$w^\top x = b + 1$$

is called the *blue margin hyperplane* and the hyperplane $H_{w,b-1}$ of equation

$$w^\top x = b - 1$$

is called the *red margin hyperplane*.

Soft Margin Support Vector Machines

If $\epsilon = 0$ and $\xi = 0$, then we recover the original constraints. By making ϵ and ξ large enough, these constraints can always be satisfied.

Ideally we would like to find a separating hyperplane that *minimizes the number of misclassified points*, which means that the variables ϵ_i and ξ_j should be as small as possible, but there is a trade-off in maximizing the margin (the thickness of the slab), and minimizing the number of misclassified points.

Soft Margin Support Vector Machines

This is reflected in the choice of the objective function, and there are several options, depending on whether we minimize a linear function of the variables ϵ_i and ξ_j, or a quadratic functions of these variables, or whether we include the term $(1/2)b^2$ in the objective function.

These methods are known as *support vector classification* algorithms (for short *SVC algorithms*).

Soft Margin Support Vector Machines

A more flexible problem is obtained by using the margin $\delta = \eta / \|w\|$, where η is some positive constant that we wish to maximize.

To do so, we add a term $-K_m \eta$ to the objective function $(1/2)w^\top w$, as well as the "regularizing term"

$$K_s\Big(\sum_{i=1}^{p} \epsilon_i + \sum_{j=1}^{q} \xi_j\Big)$$

whose purpose is to make ϵ and ξ sparse, where $K_m > 0$ (m refers to margin) and $K_s > 0$ (s refers to sparse) are fixed constants that can be adjusted to determine the influence of η and the regularizing term.