Fundamentals of Linear Algebra and Optimization

Hard Margin Support Vector Machine: Version I

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

May 7, 2020

Classification/Separation Problem

In this section we describe the following *classification problem*, or perhaps more accurately, *separation problem* (into two classes).

Classification/Separation Problem

In this section we describe the following *classification problem*, or perhaps more accurately, *separation problem* (into two classes).

Suppose we have two nonempty disjoint finite sets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$ in $\mathbb{R}^n$

Classification/Separation Problem

In this section we describe the following *classification problem*, or perhaps more accurately, *separation problem* (into two classes).

Suppose we have two nonempty disjoint finite sets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$ in $\mathbb{R}^n$

Our 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.

Classification/Separation Problem

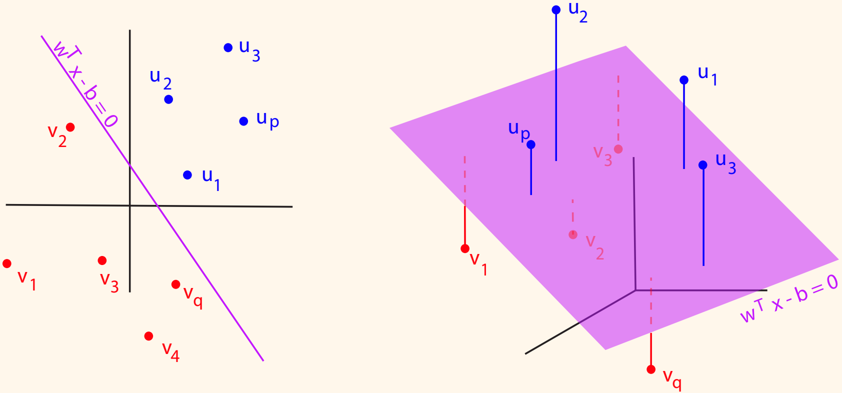

Figure 1: Two examples of the SVM separation problem. The left figure is SVM in $\mathbb{R}^2$, while the right figure is SVM in $\mathbb{R}^3$.

Classification/Separation Problem

Without loss of generality, we may assume that

$$
\begin{aligned}
w^\top u_i - b &> 0 && \text{for } i = 1, \ldots, p \\
w^\top v_j - b &< 0 && \text{for } j = 1, \ldots, q.
\end{aligned}
$$

Classification/Separation Problem

Of course, separating the blue and the red points may be impossible, as we will see in the next figure for four points where the line segments (u_1, u_2) and (v_1, v_2) intersect.

Classification/Separation Problem

Of course, separating the blue and the red points may be impossible, as we will see in the next figure for four points where the line segments (u_1, u_2) and (v_1, v_2) intersect.

If a hyperplane separating the two subsets of blue and red points exists, we say that they are *linearly separable*.

Example of An Inseparable Problem

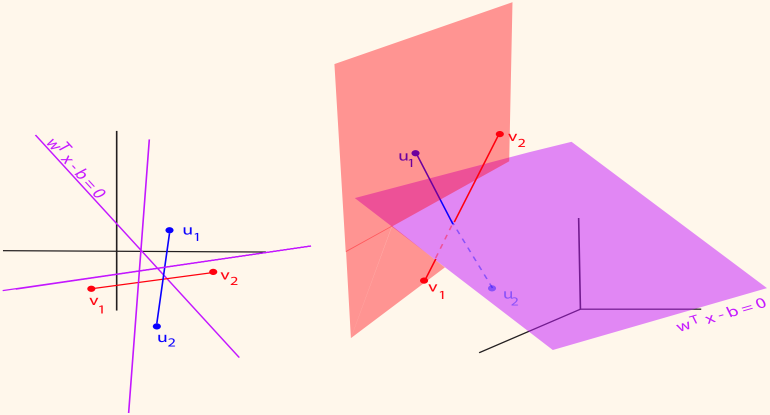

Figure 2: Two examples in which it is impossible to find purple hyperplanes which separate the red and blue points.

Classification Problem: Class Labels

Write $m = p + q$. The reader should be aware that in machine learning the classification problem is usually defined as follows.

Classification Problem: Class Labels

Write $m = p + q$. The reader should be aware that in machine learning the classification problem is usually defined as follows.

We assign m so-called *class labels* $y_k = \pm 1$ to the data points in such a way that $y_i = +1$ for each blue point u_i, and $y_{p+j} = -1$ for each red point v_j, and we denote the m points by x_k, where $x_k = u_k$ for $k = 1, \ldots, p$ and $x_k = v_{k-p}$ for $k = p + 1, \ldots, p + q$.

Classification Problem: Training Data

Then the classification constraints can be written as

$$y_k(w^\top x_k - b) > 0 \quad \text{for } k = 1, \dots, m.$$

Classification Problem: Training Data

Then the classification constraints can be written as

$$y_k(w^\top x_k - b) > 0 \quad \text{for } k = 1, \ldots, m.$$

The set of pairs $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ is called a set of *training data* (or *training set*).

Choosing the Hyperplane

We will not use the above method, and we will stick to our two subsets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$.

Choosing the Hyperplane

We will not use the above method, and we will stick to our two subsets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$.

Since there are infinitely many hyperplanes separating the two subsets (if indeed the two subsets are linearly separable), we would like to come up with a "good" criterion for choosing such a hyperplane.

Hard Margin Support Vector Machine

The idea that was advocated by Vapnik is to consider the distances $d(u_i, H)$ and $d(v_j, H)$ from *all* the points to the hyperplane H, and to pick a hyperplane H that *maximizes the smallest of these distances*.

Hard Margin Support Vector Machine

The idea that was advocated by Vapnik is to consider the distances $d(u_i, H)$ and $d(v_j, H)$ from *all* the points to the hyperplane H, and to pick a hyperplane H that *maximizes the smallest of these distances*.

In machine learning this strategy is called finding a *maximal margin hyperplane*, or *hard margin support vector machine*.

Distance from Point to Hyperplane

Since the distance from a point x to the hyperplane H of equation $w^\top x - b = 0$ is

$$d(x, H) = \frac{|w^\top x - b|}{\|w\|},$$

(where $\|w\| = \sqrt{w^\top w}$ is the Euclidean norm of w), it is convenient to temporarily assume that $\|w\| = 1$, so that

$$d(x, H) = |w^\top x - b|.$$

See the following figure.

Distance from Point to Hyperplane

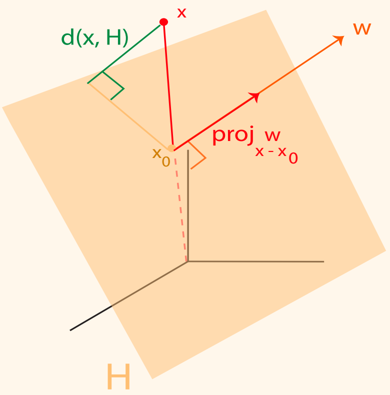

Figure 3: In $\mathbb{R}^3$, the distance from a point to the plane $w^\top x - b = 0$ is given by the projection onto the normal w.

Hard Margin Support Vector Machine

Then with our sign convention, we have

$$\begin{aligned} d(u_i, H) &= w^\top u_i - b & \quad i &= 1, \ldots, p \\ d(v_j, H) &= -w^\top v_j + b & \quad j &= 1, \ldots, q. \end{aligned}$$

Hard Margin Support Vector Machine

If we let

$$\delta = \min\{d(u_i, H),\ d(v_j, H) \mid 1 \leq i \leq p,\ 1 \leq j \leq q\},$$

then the hyperplane H should be chosen so that

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

and such that $\delta > 0$ is *maximal*.

Hard Margin Support Vector Machine

If we let

$$\delta = \min\{d(u_i, H),\ d(v_j, H) \mid 1 \leq i \leq p,\ 1 \leq j \leq q\},$$

then the hyperplane H should be chosen so that

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

and such that $\delta > 0$ is *maximal*.

The distance δ is called the *margin* associated with the hyperplane H.

Formulating the Separation Problem SVM_{h1}

This is indeed one way of formulating the two-class separation problem as an optimization problem with a linear objective function $J(\delta, w, b) = \delta$, and affine and quadratic constraints (SVM_{h1}):

$$
\begin{aligned}
&\text{maximize} \quad \delta \\
&\text{subject to} \\
&\qquad\qquad w^\top u_i - b \geq \delta \qquad\qquad i = 1, \ldots, p \\
&\qquad\qquad - w^\top v_j + b \geq \delta \qquad\qquad j = 1, \ldots, q \\
&\qquad\qquad \|w\| \leq 1.
\end{aligned}
$$

Formulating the Separation Problem SVM_{h1}

This is indeed one way of formulating the two-class separation problem as an optimization problem with a linear objective function $J(\delta, w, b) = \delta$, and affine and quadratic constraints (SVM_{h1}):

$$
\begin{aligned}
&\text{maximize} \quad \delta \\
&\text{subject to} \\
&\qquad w^\top u_i - b \geq \delta \qquad i = 1, \ldots, p \\
&\qquad -w^\top v_j + b \geq \delta \qquad j = 1, \ldots, q \\
&\qquad \|w\| \leq 1.
\end{aligned}
$$

This problem has an optimal solution $\delta > 0$ iff the two subsets are linearly separable.

The Optimal Solution for SVM_{h1}

We used the constraint $\|w\| \leq 1$ rather than $\|w\| = 1$ because the former is qualified, whereas the latter is not. But if (w, b, δ) is an optimal solution, then $\|w\| = 1$, as shown in the following proposition.

The Optimal Solution for SVM_{h1}

We used the constraint $\|w\| \leq 1$ rather than $\|w\| = 1$ because the former is qualified, whereas the latter is not. But if (w, b, δ) is an optimal solution, then $\|w\| = 1$, as shown in the following proposition.

Proposition. If (w, b, δ) is an optimal solution of Problem (SVM_{h1}), so in particular $\delta > 0$, then we must have $\|w\| = 1$.

The Optimal Solution for SVM_{h1}

Vapnik proved that if the two subsets are linearly separable, then Problem (SVM_{h1}) has a *unique* optimal solution.

The Optimal Solution for SVM_{h1}

Vapnik proved that if the two subsets are linearly separable, then Problem (SVM_{h1}) has a *unique* optimal solution.

Theorem. If two disjoint subsets of p *blue* points $\{u_i\}_{i=1}^{p}$ and q *red* points $\{v_j\}_{j=1}^{q}$ are linearly separable, then Problem (SVM_{h1}) has a *unique* optimal solution consisting of a hyperplane of equation $w^\top x - b = 0$ separating the two subsets with maximum margin δ.

The Optimal Solution for SVM_{h1}

Vapnik proved that if the two subsets are linearly separable, then Problem (SVM_{h1}) has a *unique* optimal solution.

Theorem. If two disjoint subsets of p *blue* points $\{u_i\}_{i=1}^p$ and q *red* points $\{v_j\}_{j=1}^q$ are linearly separable, then Problem (SVM_{h1}) has a *unique* optimal solution consisting of a hyperplane of equation $w^\top x - b = 0$ separating the two subsets with maximum margin δ.

Furthermore, if we define $c_1(w)$ and $c_2(w)$ by

$$c_1(w) = \min_{1 \leq i \leq p} w^\top u_i$$
$$c_2(w) = \max_{1 \leq j \leq q} w^\top v_j,$$

The Optimal Solution SVM_{h1}

then w is the unique maximum of the function

$$\rho(w) = \frac{c_1(w) - c_2(w)}{2}$$

over the convex subset U of $\mathbb{R}^n$ given by the inequalities

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

and

$$b = \frac{c_1(w) + c_2(w)}{2}.$$

Reformulating the Classification Problem

We can proceed with the formulation (SVM_{h1}) but there is a way to reformulate the problem so that the constraints are all *affine*, which might be preferable since they will be *automatically qualified*.