CS 189 Spring 2021 — Introduction to Machine Learning — Jonathan Shewchuk — DIS1

In this discussion, we'll develop some intuition for the hard-margin Support Vector Machine (SVM) optimization problem,

$$\min_{w,\alpha} \; |w|^2 \;\text{ subject to }\; y_i(X_i \cdot w + \alpha) \geq 1, \;\; \forall i \in \{1, \ldots, m\}.$$

1 Hard-margin SVM: Decision Rule

A *decision rule* (or *classifier*) is a function $r : \mathbb{R}^d \to \pm 1$ that maps a feature vector (test point) to $+1$ ("in class") or -1 ("not in class"). The decision rule for hard-margin linear SVMs is

$$r(x) = \begin{cases} +1 & \text{if } w \cdot x + \alpha \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{1}$$

where $w \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ are the weights (parameters) of the SVM.

(a) Draw a figure depicting the line $\ell = \{u \mid u \cdot w + \alpha = 0\}$ with $w = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\alpha = -25$. Include in your figure the vector w, drawn relative to ℓ.

(b) ℓ can be thought of as the decision boundary for a binary classification problem. Indicate in your figure the region in which data points $x \in \mathbb{R}^2$ would be classified as 1. Do the same for data points that would be classified as -1.

2 Constraints in Hard-margin Support Vector Machines

A maximum margin classifier maximizes the distance from the decision boundary to both positive (+1) and negative (−1) training points. The gap between the decision boundary and the closest training point is called the margin. We could express the margin requirement by imposing the constraints

$$y_i(X_i \cdot w + \alpha) \geq c, \quad \forall i \in \{1, \ldots, n\}, \tag{2}$$

where c is the margin.

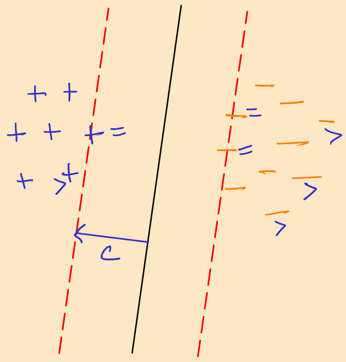
(a) What role does y_i play in Equation (2)?

(b) The margin $c > 0$ can be rescaled to be 1 without affecting the decision rule:

$$y_i(X_i \cdot w + \alpha) \geq 1, \quad \forall i \in \{1, \ldots, n\}. \tag{3}$$

Why can we rescale the margin to 1? Hint: Consider the decision rule $c(x \cdot w + \alpha) \geq 0$. What role does c play in classifying the point x? Rescale w to change the margin

(c) For which sample points i is $y_i(X_i \cdot w + \alpha) = 1$? What is the geometric interpretation and significance of these examples? "support vectors" margin maximization



$w^T x + \alpha = 0$

$x_i = (2,3)$

$w = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \alpha = 0$

$w' = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

$w^T x_i = 8$

$w'^T x_i = 16$

$c = 8$

$w^T x + \alpha > 0 \Rightarrow +$

$w^T x + \alpha < 0 \Rightarrow -$

$y_i + \Rightarrow w^T x_i + \alpha \geq 1$

$y_i - \Rightarrow w^T x_i + \alpha \leq -1$

$\max c$
$\Updownarrow$
$\min \|w\|^2$

$\forall i \quad y_i(w^T x_i + \alpha) \geq 1$

$w^T x_i + \alpha \geq 1$

$w^T x_i + \alpha \leq -1$

"Sequential Minimal Optimization"

libsvm

3 Hard-margin SVM: Objective

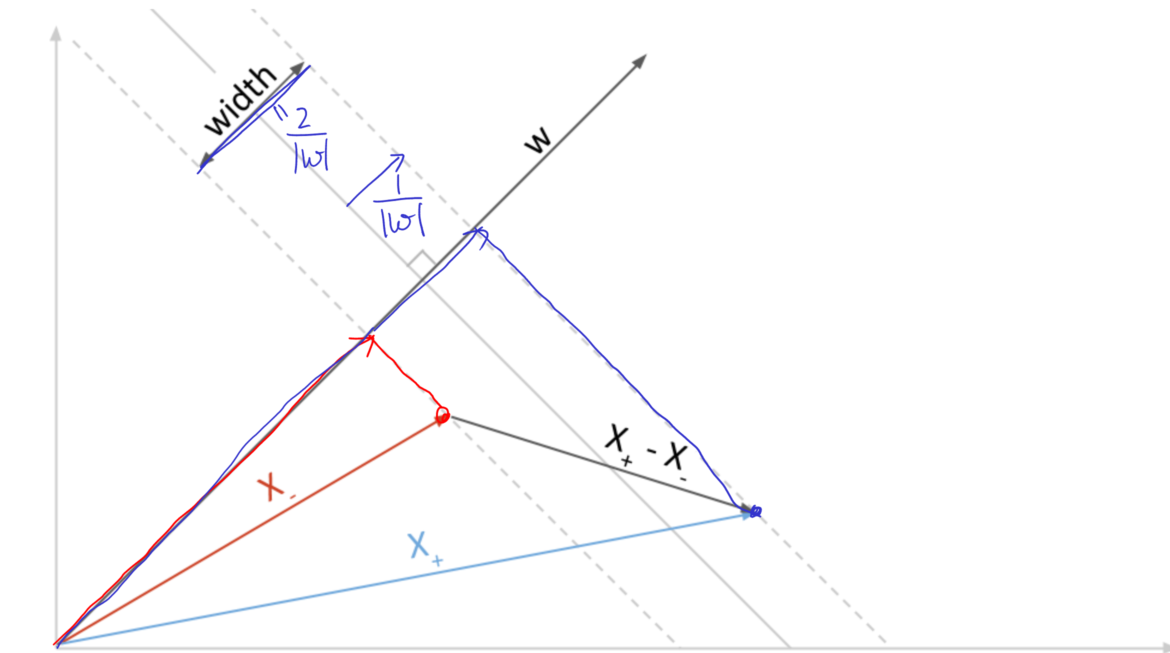


Figure 1: Diagram depicting X_+, X_-, w, and the width of the margins.

The constraints we obtained in the previous problem restrict the possible decision boundaries to those which separate the data with some margin that depends on w and b. We want the maximum possible margin. We'll need an objective we can optimize to obtain a maximum margin in terms of w and b. To obtain this objective, we rewrite Equation 2 as

$u, v \quad \mathrm{Proj}_v u = \frac{u^T v}{\|v\|_2}$

$$y_i X_i \cdot w \geq 1 - y_i \alpha, \quad i = 1, ..., m. \tag{4}$$

$|w| = \|w\|_2$

Let X_- and X_+ be negative and positive examples **on the margins**, as depicted in Figure 1. The **width** is the distance from the negative margin to the decision boundary plus the distance from the decision boundary to the positive margin, as shown in Figure 1. We can compute the width in terms of w as follows.

$X_+^T w = 1 - \alpha, \quad -X_-^T w = 1 + \alpha$

$X_-^T w = -1 - \alpha$

(a) Write down Equation 4 for X_- and X_+. Divide through by $|w|$ to obtain a scalar projection of X_- and X_+ onto $\frac{w}{|w|}$. Do the same for X_+.

$\frac{X_+^T w}{|w|} = \frac{1-\alpha}{|w|}, \quad \frac{X_-^T w}{|w|} = -\frac{1+\alpha}{|w|}$

(b) You now have two vectors pointing in the same direction, both on the margins. Compute the width using these two vectors to obtain $\frac{2}{|w|}$. $\frac{2}{|w|}$

(c) Explain in words why we want to maximize $\frac{2}{|w|}$. bc it's the size of the margin

(d) Show that $\max_{w,b} \frac{2}{|w|}$ can be rewritten as $\min_{w,b} \frac{1}{2}|w|^2$.

$|w| = \sqrt{w_1^2 + w_2^2 + \dots}$

$|w|^2 = w_1^2 + w_2^2 + \dots$

$\max \to \min \frac{|w|}{2} \to \min \frac{|w|^2}{2}$

$\frac{1}{x}$ monotone decreasing

x^2 monotone increasing

$f(x) = x$

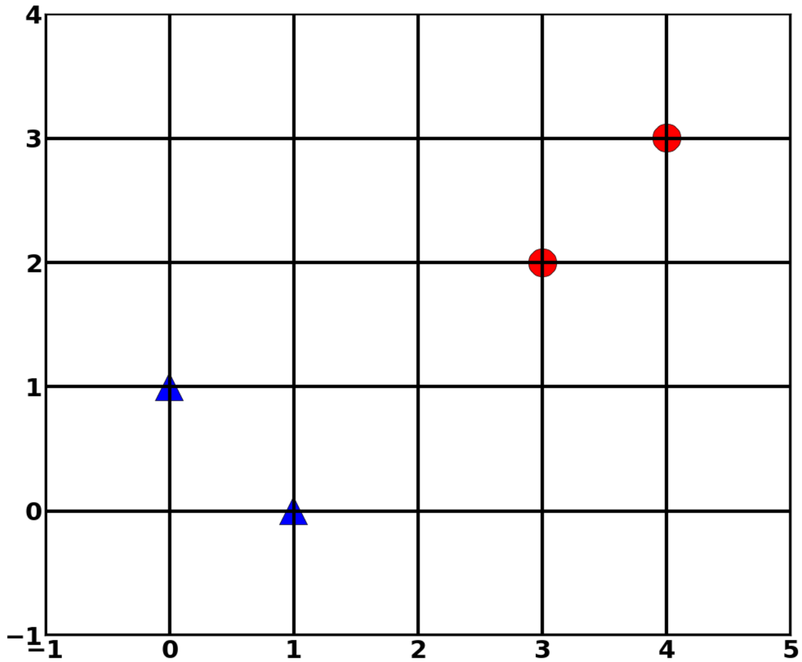
$x \leq 0$, x^2 is mon. dec.

$x \geq 0$, x^2 is mon. inc.

$\frac{df}{dx} \geq 0$

4 A Hard-Margin SVM Hyperplane Exercise

You are given the following sample points (triangle = $+1$, circle = -1).



Find (by hand) the equation of the hyperplane $w^T x + \alpha = 0$ that a hard-margin SVM classifier would learn. Draw the decision boundary and its margins.