In this discussion, we'll develop some intuition for the 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 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 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 $w = \left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right]$ 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 SVM: Constraints

We train SVMs by maximizing the distance of the decision boundary from both positive (1) and negative (-1) examples. The gap between the decision boundary and the closest positive and negative examples is called the margin. We can express the margin requirement by imposing the constraints

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

where c is taken to be the maximum margin.

(a) What role does y_i play in Equation 3?

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

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

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

(c) For which examples i is $y_i(X_i \cdot w + \alpha) = 1$? What is the geometric interpretation and significance of these examples?

3 SVM: Objective

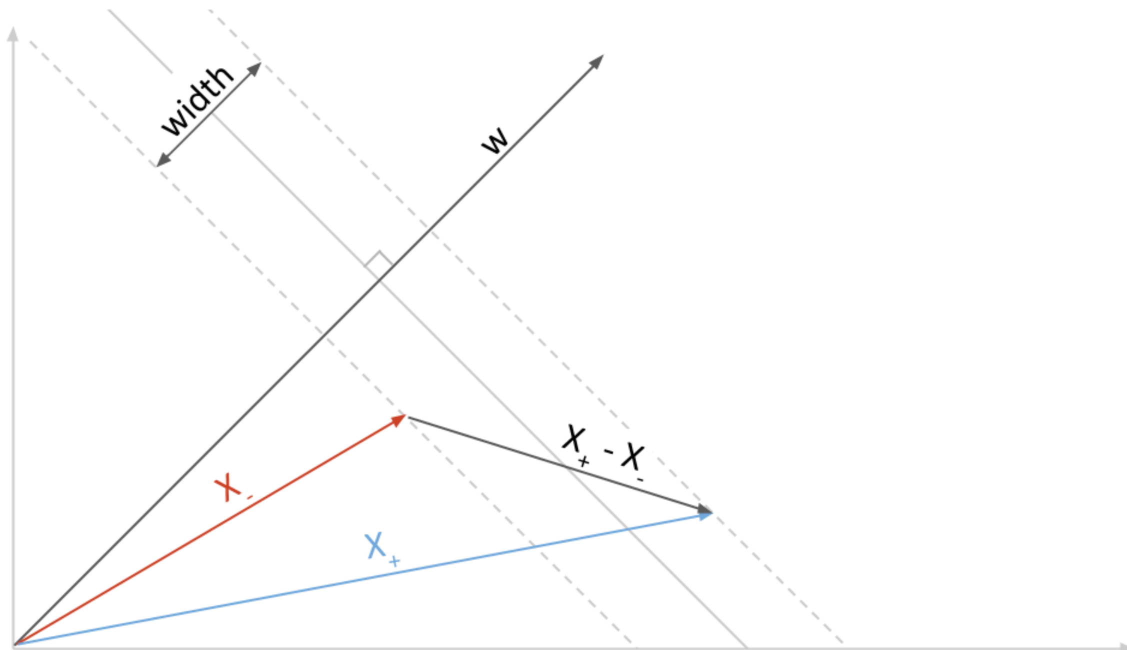


Figure 1: Diagram depicting X_+, X_-, w, and the width of the margins.

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 3 as

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

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.

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

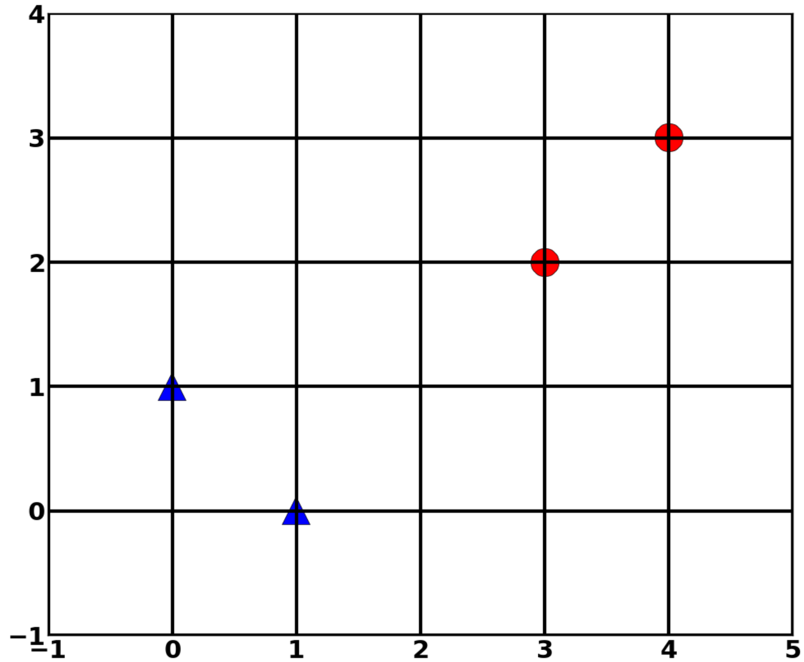
(b) You now have two vectors pointing in the same direction, both on the margins. Compute the width using these two vectors to obtain $\frac{2}{|w|}$.

(c) Explain in words why we want to maximize $\frac{2}{|w|}$.

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

4 SVM: Hyperplane Exercise

(a) You are presented with the following set of data (triangle = +1, circle = -1):



Find the equation (by hand) of the hyperplane $w^T x + b = 0$ that would be used by an SVM classifier.