

## 1 SVMs: Step by Step

A *decision rule* (or *classifier*) is a function  $r : \mathbb{R}^d \rightarrow \pm 1$  that maps a feature vector (test point) to +1 or -1. The decision rule for SVMs is

$$r(x) = \begin{cases} +1 & \text{if } \mathbf{w}^\top \mathbf{x} + b \geq 0, \\ -1 & \text{otherwise,} \end{cases} \quad (1)$$

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

- (a) **Draw a figure depicting the line  $\ell = \{\mathbf{u} \mid \mathbf{u}^\top \mathbf{w} + b = 0\}$  with  $\mathbf{w} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$  and  $b = -25$ .** Include in your figure the vector  $w$ , drawn relative to  $\ell$ . Indicate in your figure the region in which data points  $x \in \mathbb{R}^2$  would be classified as 1 vs -1.

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.

Recall from homework that the *signed distance* of a point from a hyperplane defined by  $H = \{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$  is equal to:

$$\text{SignedDistance}(\mathbf{x}, H) = \frac{\mathbf{w}^\top \mathbf{x} + b}{\|\mathbf{w}\|_2}$$

If we are given  $c$  as the maximum margin width from the decision boundary to the nearest point and  $\mathbf{w}$  as a *unit vector* defining the decision boundary hyperplane, then we can express the margin requirement (all points must lie on the correct side of the boundary) with the constraints:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq c, \quad \forall i \in \{1, \dots, n\}, \quad (2)$$

(b) **What role does  $y_i$  play in Equation 2?**

(c) The maximum margin width  $c > 0$  is something we're trying to find, so we can't plug it into our constraints right away. However, one thing to realize is that the constraint can be rescaled to 1 without affecting the decision rule:

$$y_i(\mathbf{w}'^\top \mathbf{x} + b') \geq 1, \quad \forall i \in \{1, \dots, n\}. \quad (3)$$

**What is  $\mathbf{w}', b'$  in terms of  $\mathbf{w}, b$ ? Why can we rescale right hand side to 1?**

(d) **For which examples is  $y_i(\mathbf{w}^\top \mathbf{x}_i + b) = 1$ ? What is the geometric interpretation and significance of these examples?**

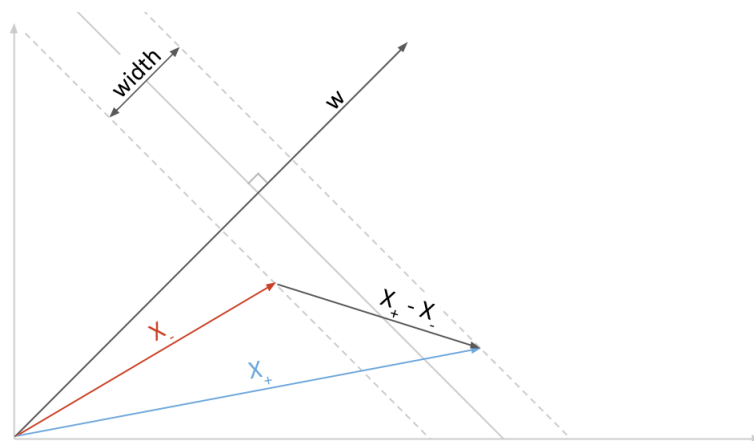


Figure 1: Diagram depicting  $\mathbf{x}_+$ ,  $\mathbf{x}_-$ ,  $\mathbf{w}$ , and the width of the margins.

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  $\mathbf{w}$  and  $b$ . We want the maximum possible margin. We'll need an objective we can optimize to obtain a maximum margin in terms of  $\mathbf{w}$  and  $b$ . To obtain this objective, we rewrite Equation 3 as

$$y_i \mathbf{w}^\top \mathbf{x}_i \geq 1 - y_i b, \quad i \in \{1, \dots, n\} \quad (4)$$

Let  $\mathbf{x}_-$  and  $\mathbf{x}_+$  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 find the margin width by computing the magnitude of the projection of  $\mathbf{x}_+ - \mathbf{x}_-$  onto the  $\mathbf{w}$  direction. Recall the formula for the projection of one vector  $\mathbf{u}$  onto another vector  $\mathbf{v}$ :

$$\text{proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{v}\|_2} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|_2}$$

(e) **Write out the projection**  $\text{proj}_{\mathbf{w}}(\mathbf{x}_+ - \mathbf{x}_-) = \text{proj}_{\mathbf{w}} \mathbf{x}_+ - \text{proj}_{\mathbf{w}} \mathbf{x}_-$ .

Draw this projection as a vector in Figure 1. What is its direction and magnitude?

(f) **Write down Equation 4 for  $\mathbf{x}_+$  and  $\mathbf{x}_-$ .**

Use it to substitute a value for  $\mathbf{x}_+^\top \mathbf{w}$  and  $\mathbf{x}_-^\top \mathbf{w}$  into the projection you found in the previous part.

(g) **Now, take the magnitude of this projection to calculate the margin width in terms of  $\mathbf{w}$ .**

(h) We want to maximize the margin width. Using the answer from the previous part, **show how the SVM objective can be written as  $\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2$ .**

## 2 SVM with custom margins

In the lecture, we covered the soft-margin SVM. The objective to be optimized over the training set  $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$  is

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (5)$$

$$s.t. \quad y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1 - \xi_i \quad \forall i \quad (6)$$

$$\xi_i \geq 0 \quad \forall i \quad (7)$$

In this problem, we are interested in a modified version of the soft-margin SVM where we have a custom margin for each of the  $n$  data points. In the standard soft-margin SVM, we pay a penalty of  $\xi_i$  for each of the data point. In practice, we might not want to treat each training point equally – for example, we might know that some data points are more important than the others.

We formally define the following optimization problem:

$$\min_{\mathbf{w}, b, \{\xi_i\}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \phi_i \xi_i \quad (8)$$

$$s.t. \quad y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1 - \xi_i \quad \forall i \quad (9)$$

$$\xi_i \geq 0 \quad \forall i \quad (10)$$

Note that the only difference is that we have a custom weighting factor  $\phi_i > 0$  for each of the slack variables  $\xi_i$  in the objective function. These  $\phi_i$  are some constants given by the prior knowledge, and thus they can be treated as known constants in the optimization problem. Intuitively, this formulation weights each of the violations ( $\xi_i$ ) differently according to the prior knowledge ( $\phi_i$ ).

- (a) For the standard soft-margin SVM, we have shown that the constrained optimization problem is equal to the following unconstrained optimization problem:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i - b)) \quad (11)$$

What's the corresponding unconstrained optimization problem for the SVM with custom margins?

(b) As seen in lecture, the dual form of the standard soft-margin SVM is:

$$\max_{\alpha} \quad \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T \mathbf{Q} \alpha \quad (12)$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (13)$$

$$0 \leq \alpha_i \leq C \quad i = 1, \dots, n \quad (14)$$

where  $\mathbf{Q} = (\text{diag } \mathbf{y}) \mathbf{X} \mathbf{X}^T (\text{diag } \mathbf{y})$ .

What's the dual form of the SVM with custom margins? To start, we provide you the Lagrangian, which is given by

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i (\mathbf{w}^T \mathbf{x}_i - b) + \sum_{i=1}^n (C \phi_i - \alpha_i - \beta_i) \xi_i \quad (15)$$