Cost Functions for Half-space Models

Chris Cornwell

November 3, 2025

Outline

Outline

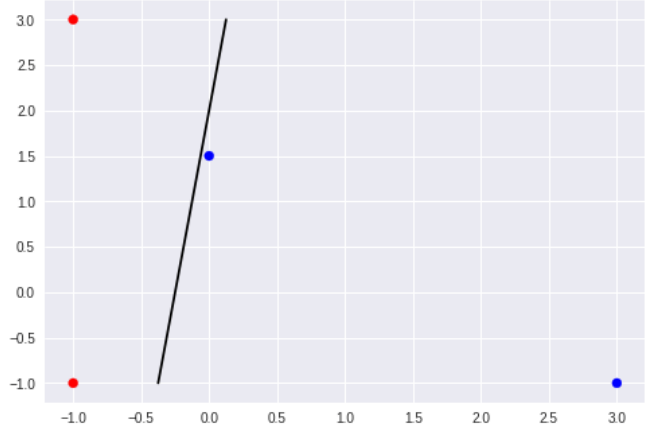
Outline

Returning to the "Simple Example" for Perceptron algorithm

Example in $\mathbb{R}^2$, with $P = 4$ points.

$$\mathrm{x}: \begin{bmatrix} -1 & 3 \\ -1 & -1 \\ 3 & -1 \\ 0 & 1.5 \end{bmatrix} \qquad \mathrm{y}: \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

Final $\tilde{\mathbf{w}} = \begin{bmatrix} 1 \\ 4 \\ -0.5 \end{bmatrix}$; hyperplane in $\mathbb{R}^2$ (a line) shown below.



What if this data were part of a sample, which has noise? What about the point near the hyperplane?

Getting a Buffer – a “margin”

Say that $S = \{\mathbf{x}_i, y_i\}_{i=1}^{P}$ (with $y_i = \pm 1$) is linearly separable. Instead of simply trying to find a hyperplane that successfully separates the data, try to find one that has some distance from points on each side.

Getting a Buffer – a "margin"

Say that $S = \{\mathbf{x}_i, y_i\}_{i=1}^{P}$ (with $y_i = \pm 1$) is linearly separable. Instead of simply trying to find a hyperplane that successfully separates the data, try to find one that has some distance from points on each side.
Can express this as wanting no points between two parallel hyperplanes – between the set of $\mathbf{x}$ where $\tilde{\mathbf{x}}^\top \tilde{\mathbf{w}} = 1$ and where $\tilde{\mathbf{x}}^\top \tilde{\mathbf{w}} = -1$.

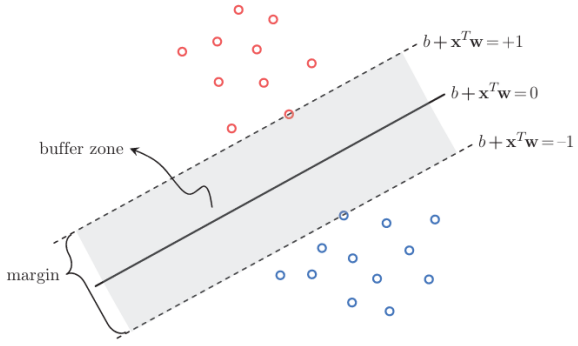


Figure: Taken from Figure 4.4 in textbook, p. 79. (Note: they used blue for the -1 label.)

This is equivalent to wanting $y_i(\tilde{\mathbf{x}}_i^\top \tilde{\mathbf{w}}) \geq 1$ for all $i = 1, \ldots, P$.
Recall, S being linearly separable implies that there must be a $\tilde{\mathbf{w}}$ that achieves this.

Outline

Outline