Agenda: perceptrons as linear separators; the on-line learning setting and the need, at least in the perceptron-learning case, for restrictions on the oracle. We may also get to Rosenblatt’s perceptron learning algorithm.

I. Recall: perceptron functions

Given a weight vector \( \vec{w} \) in \( \mathbb{R}^n \), \( n \geq 1 \), and a threshold value \( T \), the function \( f_{\vec{w}, T} : \mathbb{R}^n \to \{+1, -1\} \) is given by

\[
f_{\vec{w}, T}(\vec{x}) = \begin{cases} 
+1, & \vec{w} \cdot \vec{x} \geq T \\
-1, & \text{otherwise}
\end{cases}
\]

If length \((\vec{w}) > 0\), then this can be rewritten as

\[
f_{\vec{w}, T}(\vec{x}) = \begin{cases} 
+1, & \text{proj}(\vec{x}, \vec{w}) \geq \frac{T}{\text{length}(\vec{w})} \\
-1, & \text{otherwise}
\end{cases}
\]

where we are defining the projection of \( \vec{x} \) onto \( \vec{w} \) to be the (signed) length from the origin to the point on \( \vec{w} \) that results from dropping a perpendicular from \( \vec{x} \) to \( \vec{w} \). (This is a signed distance because it can be negative if \( \vec{x} \) points “behind” \( \vec{w} \).)

II. Linear separation

Continuing with the above conditions, if we let \( \vec{w}' \) be the vector that points in the direction of \( \vec{w} \) but has length \( T/\text{length}(\vec{w}) \), we get this picture:

Hence, a perceptron function is a linear separator corresponding to a half-plane concept.
III. Conventions for the oracle’s sequence of labeled examples We denote the instances by \( \overrightarrow{x}(1), \overrightarrow{x}(2), \ldots, \overrightarrow{x}(i), \ldots \). Those that are given label +1 are known as positive examples; those that are given label -1 are known as negative examples.

IV. Identification in the limit (a success criterion for on-line learning) After seeing a finite number of examples,

- the learner always outputs the same hypothesis (\( \overrightarrow{w} \) and \( T \)) from then on, and
- this hypothesis correctly predicts the label of every subsequent example.