Brief Introduction to the Voted Perceptron Algorithm

Recall the Perceptron Algorithm

\[ w = 0 \]
\[ \text{while some } (x_i, y_i) \text{ is misclassified:} \]
\[ w = w + y_i x_i \]

Non-separable data: will never converge.
How can this be fixed?
Dream: somehow find the separator that misclassifies the fewest points… but this is NP-hard (in fact, even NP-hard to approximately solve).
Fixing the Perceptron

Idea one: only go through the data once, or a fixed number of times

\[ w = 0 \]
\[ \text{for } k = 1 \text{ to } K \]
\[ \quad \text{for } i = 1 \text{ to } N \]
\[ \quad \text{if } (x_i, y_i) \text{ is misclassified:} \]
\[ \quad \quad w = w + y_i x_i \]

At least this stops!

Problem: the final \( w \) might not be good

\text{e.g., right before terminating, the algorithm might perform an update on a total outlier…}

Voted-Perceptron

Idea two: keep around intermediate hypotheses, and have them “vote” [Freund and Schapire, 1998]

\[ n = 1 \]
\[ w_1 = 0 \]
\[ c_1 = 0 \]
\[ \text{for } k = 1 \text{ to } K \]
\[ \quad \text{for } i = 1 \text{ to } N \]
\[ \quad \text{if } (x_i, y_i) \text{ is misclassified:} \]
\[ \quad \quad w_{n+1} = w_n + y_i x_i \]
\[ \quad \quad c_{n+1} = 1 \]
\[ \quad \quad n = n + 1 \]
\[ \text{else} \]
\[ \quad \quad c_n = c_n + 1 \]

At the end, a collection of linear separators \( w_0, w_1, w_2, \ldots \), along with survival times: \( c_n = \text{amount of time that } w_n \text{ survived.} \)
Voted-Perceptron, cont’d

Idea two: keep around intermediate hypotheses, and have them “vote” [Freund and Schapire, 1998]

At the end, a collection of linear separators \(w_0, w_1, w_2, \ldots\), along with survival times: \(c_n = \) amount of time that \(w_n\) survived.

This \(c_n\) is a good measure of the reliability of \(w_n\).

To classify a test point \(x\), use a weighted majority vote:

\[
\text{sgn} \left\{ \sum_{n=0}^{N} c_n \text{sgn}(w_n \cdot x) \right\}
\]

Voted-perceptron, cont’d

Problem: need to keep around a lot of \(w_n\) vectors, and it’s expensive to compute

\[
\text{sgn} \left\{ \sum_{n=0}^{N} c_n \text{sgn}(w_n \cdot x) \right\}
\]

Solution:

\[
\text{sgn} \left\{ \sum_{n=0}^{N} c_n (w_n \cdot x) \right\} = \text{sgn} \left\{ \left( \sum_{n=0}^{N} c_n w_n \right) \cdot x \right\}
\]

\[w_{\text{avg}}\]