CSE 446: Machine Learning **Lecture**

Proof of the Perceptron Mistake Lemma

Instructor: Sham Kakade

Our *data set* is: $\{(x_1, y_1), \ldots (x_N, y_N)\}$, the x_n's are real vectors and where $y_n \in \{-1, +1\}$. Suppose this data set is *linearly separable*. That is, there is a w_* such that (for all n in our dataset),

$$y_n = \text{sgn}(w_* \cdot x_n) \; . \tag{1}$$

Separability means that $y_n(w_* \cdot x_n) > 0$ for all points in our data set. Let us make the assumption that this quantity is lower bounded by 1 (which is true without loss of generality).

Assumption M. *Without loss of generality suppose $\|x_n\| \leq 1$. Suppose there exists a $w_* \in \mathbb{R}^d$ for which (1) holds. Further assume that (for all n in our dataset),*

$$y_n(w_* \cdot x_n) \geq 1 \; , \tag{2}$$

Note the choice of 1 is arbitrary, as w_* has an arbitrary scale.

Connection to the "geometric margin": Note that the above implies that (for all n in our dataset):

$$y_n(\frac{w_*}{\|w_*\|} \cdot x_n) \geq \frac{1}{\|w_*\|} \; .$$

In other words, the width of the strip separating the positives from the negatives is of size $\frac{2}{\|w_*\|}$. The margin is often defined this way (where we have implicitly used the scaling that $\|w_*\| = 1$ and that the margin is some positive value rather than 1). In particular in CIML, the "geometric margin" implicitly assumes the scaling where $\|w_*\| = 1$. Both lead the same convergence guarantee for the number of total mistakes made by the perceptron algorithm.

We are now ready to prove the mistake lemma.

Lemma 0.1. *(Perceptron Mistake Lemma) Suppose Assumption M holds. Define $m_t = 1$ if a mistake occurs at time t and 0 otherwise. By construction of the perceptron algorithm, the update rule is:*

$$w_{t+1} = w_t + m_t y_t x_t$$

(Note that using m_t provides a compact way to write the update rule, i.e. if $m_t = 0$, there is no update, and, if $m_t = 1$, the corresponds to the usual update $w_{t+1} = w_t + y_t x_t$.) We have that:

$$\|w_{t+1} - w_*\|^2 \leq \|w_t - w_*\|^2 - m_t.$$

Proof. Observe:

$$\begin{aligned} \|w_{t+1} - w_*\|^2 &= \|w_t + m_t y_t x_t - w_*\|^2 \\ &= \|w_t - w_*\|^2 + 2m_t y_t x_t (w_t - w_*) + m_t^2 y_t^2 \|x_t\|^2 \\ &= \|w_t - w_*\|^2 + 2m_t y_t x_t (w_t - w_*) + m_t \|x_t\|^2 \\ &\leq \|w_t - w_*\|^2 + 2m_t y_t x_t (w_t - w_*) + m_t \end{aligned}$$

The middle term can be bounded as follows:

$$m_t y_t x_t (w_t - w_*) = m_t y_t x_t w_t - m_t y_t x_t w_* \leq 0 - m_t < -m_t$$

where have used that $y_t x_t w_t < 0$ when there is a mistake (to bound the first term by 0) and we have used the margin assumption to bound the second term.

Putting this together

$$\begin{aligned} \|w_{t+1} - w_*\|^2 &\leq \|w_t - w_*\|^2 + 2m_t y_t x_t (w_t - w_*) + m_t \\ &\leq \|w_t - w_*\|^2 - 2m_t + m_t \\ &\leq \|w_t - w_*\|^2 - m_t \end{aligned}$$

which completes the proof. □