Perceptron and Online Perceptron

Daniel Hsu (COMS 4771)

Perceptron algorithm

The Perceptron algorithm is given as follows. The input to the algorithm is a collection $S$ of labeled examples from $\mathbb{R}^d \times \{-1,+1\}$.

- Begin with $\hat{w}_1 := 0 \in \mathbb{R}^d$.
- For $t = 1, 2, \ldots$:
  - If there is a labeled example in $S$ (call it $(x_t, y_t)$) such that $y_t \langle \hat{w}_t, x_t \rangle \leq 0$, then set $\hat{w}_{t+1} := \hat{w}_t + y_t x_t$.
  - Else, return $\hat{w}_t$.

**Theorem.** Let $S$ be a collection of labeled examples from $\mathbb{R}^d \times \{-1,+1\}$. Suppose there exists a vector $w_* \in \mathbb{R}^d$ such that

$$\min_{(x,y) \in S} y \langle w_*, x \rangle \geq 1.$$ 

Then Perceptron on input $S$ halts after at most $\|w_*\|_2^2 R^2$ loop iterations, where $R := \max_{(x,y) \in S} \|x\|_2$.

**Proof.** Suppose Perceptron does not exit the loop in the $t$-th iteration. Then there is a labeled example $(x_t, y_t) \in S$ such that

$$y_t \langle w_*, x_t \rangle \geq 1,$$

$$y_t \langle \hat{w}_t, x_t \rangle \leq 0.$$

We bound $\langle w_*, \hat{w}_{t+1} \rangle$ from above and below to deduce a bound on the number of loop iterations. First, we bound $\langle w_*, \hat{w}_t \rangle$ from below:

$$\langle w_*, \hat{w}_t \rangle = \langle w_*, \hat{w}_t \rangle + y_t \langle w_*, x_t \rangle \geq \langle w_*, \hat{w}_t \rangle + 1.$$

Since $\hat{w}_1 = 0$, we have

$$\langle w_*, \hat{w}_t \rangle \geq t.$$

We now bound $\langle w_*, \hat{w}_{t+1} \rangle$ from above. By Cauchy-Schwarz,

$$\langle w_*, \hat{w}_{t+1} \rangle \leq \|w_*\|_2 \|\hat{w}_{t+1}\|_2.$$

Also,

$$\|\hat{w}_{t+1}\|_2^2 = \|\hat{w}_t\|_2^2 + 2y_t \langle \hat{w}_t, x_t \rangle + y_t^2 \|x_t\|_2^2 \leq \|\hat{w}_t\|_2^2 + R^2.$$ 

Since $\|\hat{w}_1\|_2 = 0$, we have

$$\|\hat{w}_{t+1}\|_2 \leq tR^2,$$

so

$$\langle w_*, \hat{w}_{t+1} \rangle \leq \|w_*\|_2 R \sqrt{t}.$$ 

Combining the upper and lower bounds on $\langle w_*, \hat{w}_{t+1} \rangle$ shows that

$$t \leq \langle w_*, \hat{w}_{t+1} \rangle \leq \|w_*\|_2 R \sqrt{t},$$

which in turn implies the inequality $t \leq \|w_*\|_2^2 R^2$. \[\square\]
Online Perceptron algorithm

The Online Perceptron algorithm is given as follows. The input to the algorithm is a sequence 
\((x_1, y_1), (x_2, y_2), \ldots\) of labeled examples from \(\mathbb{R}^d \times \{-1, +1\}\).

- Begin with \(\hat{w}_1 := 0 \in \mathbb{R}^d\).
- For \(t = 1, 2, \ldots\):
  - If \(y_t \langle \hat{w}_t, x_t \rangle \leq 0\), then set \(\hat{w}_{t+1} := \hat{w}_t + y_t x_t\).
  - Else, \(\hat{w}_{t+1} := \hat{w}_t\).

We say that Online Perceptron makes a mistake in round \(t\) if \(y_t \langle \hat{w}_t, x_t \rangle \leq 0\).

**Theorem.** Let \((x_1, y_1), (x_2, y_2), \ldots\) be a sequence of labeled examples from \(\mathbb{R}^d \times \{-1, +1\}\) such that there exists a vector \(w_* \in \mathbb{R}^d\) satisfying

\[
\min_{t=1,2,\ldots} y_t \langle w_*, x_t \rangle \geq 1.
\]

Then Online Perceptron on input \((x_1, y_1), (x_2, y_2), \ldots\) makes at most \(\|w_*\|_2 R^2\) mistakes, where \(R := \max_{t=1,2,\ldots} \|x_t\|_2\).

**Proof.** Left as an exercise.