Lerning from Data

Maciej Leks

15.02.2016

Excercise 1.3:

(a)

$$y_t \mathbf{w}_t^T \mathbf{x}_t \quad < \quad 0$$

since $sign(y_t) \neq sign(\mathbf{w}_t^T \mathbf{x}_t)$

(b)

$$\begin{aligned}
y_t \mathbf{w}_{t+1}^T \mathbf{x}_t &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t \mathbf{w}_{t+1}^T \mathbf{x}_t &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t (\mathbf{w}_t + y_t \mathbf{x}_t)^T \mathbf{x}_t &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t (\mathbf{w}_t^T + y_t \mathbf{x}_t^T) \mathbf{x}_t &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t (\mathbf{w}_t^T \mathbf{x}_t + y_t \mathbf{x}_t^T \mathbf{x}_t) &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t \mathbf{w}_t^T \mathbf{x}_t + y_t^2 \mathbf{x}_t^T \mathbf{x}_t &> y_t \mathbf{w}_t^T \mathbf{x}_t \\
y_t^2 \mathbf{x}_t^T \mathbf{x}_t &> 0 \\
\mathbf{x}_t^T \mathbf{x}_t &> 0 \\
\mathbf{x}_t \cdot \mathbf{x}_t &> 0 \\
\|\mathbf{x}_t\|^2 &> 0
\end{aligned}$$

only if $\mathbf{x}_t \neq \mathbf{0}$ (by the definition $x_t \cdot x_t \geq 0$) but in our case the zeroth coordinate$x_0$of every $\mathbf{x}_t$is always fixed at $+1$.

See my question: `http://book.caltech.edu/bookforum/showthread.php?t=4658`

(c) Move from $\mathbf{w}_t$to $\mathbf{w}_{t+1}$is a move is a move "in the right direction" since

$$\begin{aligned}
\mathbf{w}_{t+1} &= \mathbf{w}_t + y_t \mathbf{x}_t \\
\mathbf{w}_{t+1}^T &= \mathbf{w}_t^T + y_t \mathbf{x}_t^{\mathbf{T}} \\
\mathbf{w}_{t+1}^T \mathbf{x}_t &= \mathbf{w}_t^T \mathbf{x}_t + y_t \mathbf{x}_t^T \mathbf{x}_t
\end{aligned}$$

- If x_t was correctly classified, then the algorithm does not apply the update rule, so nothing changes.

- If x_t was incorrectly classified as negative, then $y_t = 1$. It follows that the new dot product increased by x(t) x(t) (which is positive). The boundary moved in the right direction as far as $x_t$is concerned, therefore.
- Conversely, if x(t) was incorrectly classified as positive, then $y_t = -1$. It follows that the new dot product decreased by $x_t \cdot x_t$ (which is positive). The boundary moved in the right direction as far as x_t is concerned, therefore.

Source of explanation: `http://stackoverflow.com/questions/34477827/intuition-for-perceptron-`

Excercise 1.11

We must recall that $E_{out} = \mathbb{P}[h(x) \neq f(x)]$.

(a) Question: Can S produce a hypothesis that is guaranteed to perform better than random on any point outside D?
Answer: No, because random number getting from Bernoulli distribution will give $E_{out} \approx 0.5$ and there is now way to get better than random using Sin deterministic sense.

(b) Question: Assume for the rest of the exercise that all the examples i $\mathcal{D}$ have $y_n = +1$. Is it possible that the hypothesis that C produces turns out to be better than the hypothesis that S produces?
Answer: tba

(c) Question: If $p = 0.9$, what is the probability that S will produce a better hypothesis than C.
Answer: tba

(d)

Question: Is there any value of p for which it is more likely than not that C will produce a better hypothesis than S? Answer:

$$\mathbb{P}[E_{out}(C(D)) < E_{out}(S(D))] = \mathbb{P}[\mathbb{P}[f(x) \neq h_2(x)] < \mathbb{P}[f(x) \neq h_1(x)]] = \mathbb{P}[\mathbb{P}[f(x) \neq -1] < \mathbb{P}[f(x) \neq +1]] =$$