Solution to homework 3

October 31, 2016

Maximum Likelihood Estimation

Exercise 1

Suppose i.i.d variable $X_i \sim Poisson(\lambda), i = 1, 2, \dots, n$, we have

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n} X_i}{n}$$

In this case, $X_i \in \{0, 1, 2, 3, 4\}, n = 480$. Thus

$$\hat{\lambda}_{MLE} = \frac{40 \times 0}{480} + \frac{281 \times 1}{480} + \frac{134 \times 2}{480} + \frac{19 \times 3}{480} + \frac{6 \times 4}{480} \approx 1.3125$$

The expected number of students arriving the Dining Hall per minute is 1.3125.

Perceptron

Excercise 2

1. $\boldsymbol{x}(t)$ is misclassified by $\boldsymbol{w}(t)$, we have

$$\boldsymbol{w}^T(t)\boldsymbol{x}(t) < 0, \qquad when \quad y(t) = +1$$

$$\boldsymbol{w}^T(t)\boldsymbol{x}(t) > 0, \qquad when \quad y(t) = -1$$

Thus $y(t)\boldsymbol{w}^T(t)\boldsymbol{x}(t) \leq 0$.

2.

$$\begin{aligned}
y(t)\boldsymbol{w}^T(t+1)\boldsymbol{x}(t) &= y(t)[\boldsymbol{w}^T(t) + \boldsymbol{x}(t)y(t)]^T\boldsymbol{x}(t) \\
&= y(t)\boldsymbol{w}^T(t)\boldsymbol{x}(t) + y(t)^2\boldsymbol{x}^T(t)\boldsymbol{x}(t) \\
&> y(t)\boldsymbol{w}^T(t)\boldsymbol{x}(t)
\end{aligned}$$

3. If the data set is linear separable, there must exits a perfect $\boldsymbol{w}_f$, for all $(\boldsymbol{x}_i, y_i)$, $y_i = sign(\boldsymbol{w}_f^T\boldsymbol{x}_i)$. Set initial $\boldsymbol{w}(0) = \boldsymbol{0}$ After T times update, we have:

$$\begin{aligned}
\boldsymbol{w}_f^T\boldsymbol{w}(T) &= \boldsymbol{w}_f^T(\boldsymbol{w}(T-1) + y_i\boldsymbol{x}) \\
&\geq \boldsymbol{w}_f^T\boldsymbol{w}(T-1) + min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n}) \\
&\geq \dots \geq \boldsymbol{w}_f^T\boldsymbol{w}(0) + T * min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n}) \\
&= T * min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n})
\end{aligned}$$

$$\begin{aligned}\|\boldsymbol{w}(T+1)\|^2 &= \|\boldsymbol{w}(T) + y_{n(T)}\boldsymbol{x}_{n(T)}\|^2 \\ &\|\boldsymbol{w}(T)\|^2 + 2y_{n(T)}\boldsymbol{w}_t^T\boldsymbol{x}_{n(T)} + \|y_{n(T)}\boldsymbol{x}_{n(T)}\| \\ &\leq |\boldsymbol{w}(T)\|^2 + 0 + \|y_{n(T)}\boldsymbol{x}_{n(T)}\|^2 \\ &\leq |\boldsymbol{w}(T)\|^2 + max_n\|y_n\boldsymbol{x}_n\|^2\end{aligned}$$

$$\begin{aligned}\|\boldsymbol{w}(T)\|^2 &\leq \|\boldsymbol{w}(T-1)\|^2 + max_n\|\boldsymbol{x}_n\|^2 \\ &\leq \|\boldsymbol{w}_0\|^2 + T * max_n\|\boldsymbol{x}_n\|^2 \\ &\leq T * max_n\|\boldsymbol{x}_n\|^2\end{aligned}$$

Thus we have

$$\begin{aligned}cos(\theta) &= \frac{\boldsymbol{w}_f^T\boldsymbol{w}(T)}{\|\boldsymbol{w}_f\|\|\boldsymbol{w}(T)\|} \\ &\geq \frac{T * min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n})}{\|\boldsymbol{w}_f\|\|\boldsymbol{w}(T)\|} \\ &\geq \frac{T * min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n})}{\sqrt{T} * max_n\|\boldsymbol{x}_n\|} \\ &= \sqrt{T} * \frac{min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n})}{\|\boldsymbol{w}_f\| * max_n\|\boldsymbol{x}_n\|}\end{aligned}$$

where $\frac{min_n(y_n\boldsymbol{w}_f^T\boldsymbol{x_n})}{\|\boldsymbol{w}_f\| * max_n\|\boldsymbol{x}_n\|}$ is a constant, hence the angle between vector $\boldsymbol{w}_f$ and $\boldsymbol{w}(t)$ is decreasing with increasing T. That is, $\boldsymbol{w}(T)$ is approaching the perfect $\boldsymbol{w}_f$, we are on the "right direction".

Hoeffding Inequality

Excercise 3

1. Since $\nu \leq 0.1$ and $n = 10$, we have $k \leq 1$, *i.e* $k = 0, 1$. According to Binomial distribution

$$P(k = 0, 1|\mu = 0.9) = 0.1^{10} + C_{10}^1 \times 0.9 \times 0.1^9 = 9.1 \times 10^{-9}$$

2. According to Hoeffding Inequality

$$\begin{aligned}P(\nu \leq 0.1|\mu = 0.9) &= P(|\nu - \mu| \geq 0.8|\mu = 0.9) \\ &\leq 2e^{-2\times 0.8^2 \times 10} \\ &\approx 5.5215 \times 10^{-6}\end{aligned}$$

Learning feasibility

Excercise 4

1. No.
2. yes.
3. $P[S\ is\ better|p = 0.9] = P[S\ chooses\ h_1|p = 0.9] = \sum_{i=13}^{25} C_{25}^i 0.9^i 0.1^{25-i}$
$\approx 0.9999998379165848$
4. $p < 0.5$