Statistical Learning **Scribed By: Tao Huang**
School of Information Science and Technology **ShanghaiTech University, China**
January 10, 2021

Outline *This note dedicates to conclude some important contents of Statistical Learning by Hang Li, augmented by some insights and self-understanding to them. The organization follows directly from Li, and I wish it can help you dive deeper in this interesting area. Let's start now!*

Contents

1 Introduction to Statistical Learning

2 Perceptron

2.1 Model

$$f(x) = sign(w \cdot x + b)$$

$$sign(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

2.2 Strategy

$$\min_{w,b} L(w,b) = -\sum_{x \in M} y_i (w \cdot x_i + b)$$

2.3 Algorithm

2.3.1 Primal Problem

$$\begin{aligned} w &\leftarrow w + \eta \nabla_w L(w,b) = w + \eta \sum_{x \in M} y_i x_i \\ b &\leftarrow b + \eta \nabla_b L(w,b) = b + \eta \sum_{x \in M} y_i \end{aligned}$$

2.3.2 Dual Problem

$$\begin{aligned} w &= \sum_{i=1}^{N} n_i \eta y_i x_i \triangleq \sum_{i=1}^{N} \alpha_i y_i x_i \\ b &= \sum_{i=1}^{N} \alpha_i y_i \end{aligned}$$

$$\begin{aligned} \alpha_i &\leftarrow \alpha_i + \eta \\ b &\leftarrow b + \eta y_i \end{aligned}$$

2.3.3 Convergence

Proof. (1)

$$\gamma = \min_i \{y_i(\hat{w}_{opt} \cdot \hat{x}_i)\}$$

(2)

$$\hat{w}_k \leftarrow \hat{w}_{k-1} + \eta y_i \hat{x}_i$$

$$\begin{aligned} \hat{w}_k \cdot \hat{w}_{opt} &\geq k\eta\gamma \\ \|\hat{w}_k\|^2 &\leq k\eta^2 R^2 \end{aligned}$$

□

3 K-Means

4 Naive Bayes

5 Decision Tree

6 LR & MEM

7 Support Vector Machine

8 Boosting

9 EM Algorithm

$$\theta^\star = \arg\max_\theta \log P(X_1, X_2, ..., X_n|\theta)$$

$$\theta^\star = \arg\max_\theta P(\theta|X_1, ..., X_n) = \arg\max_\theta \frac{P(X_1, ..., X_n, \theta)}{P(X_1, ..., X_n)}$$

$$
\begin{aligned}
L(\theta) &= \log P(Y|\theta) \\
&= \log \sum_Z P(Y, Z|\theta) \\
&= \log \left(\sum_Z P(Z|Y, \theta) P(Y|\theta) \right)
\end{aligned}
$$

$$
\begin{aligned}
L(\theta) - L(\theta^{(i)}) &= \log \sum_Z P(Y, Z|\theta) - \log P(Y|\theta^{(i)}) \\
&= \log \sum_Z P(Z|Y, \theta^{(i)}) \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)})} - \log P(Y|\theta^{(i)}) \\
&\geq \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)})} - \log P(Y|\theta^{(i)}) \\
&= \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)})} - \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y|\theta^{(i)}) \\
&= \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)}) P(Y|\theta^{(i)})}
\end{aligned}
$$

$$
B(\theta, \theta^{(i)}) = L(\theta^{(i)}) + \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)}) P(Y|\theta^{(i)})}
$$

$$
\begin{aligned}
\theta^{(i+1)} &= \arg\max_\theta B(\theta, \theta^{(i)}) \\
&= \arg\max_\theta \left(L(\theta^{(i)}) + \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)}) P(Y|\theta^{(i)})} \right) \\
&= \arg\max_\theta \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta) \\
&\triangleq \arg\max_\theta Q(\theta, \theta^{(i)})
\end{aligned}
$$

$$
\begin{aligned}
Q(\theta, \theta^{(i)}) &= E_Z \left[\log P(Y, Z|\theta) | Y, \theta^{(i)} \right] \\
&= \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta)
\end{aligned}
$$

10 Hidden Markov Model

References