Homework 2

Lidian Lin

January 2020

1 XOR Problem

$$J(\theta) = \frac{1}{4}\sum_{x \in X}(f^*(x) - f(x;\theta))^2 \tag{1}$$

$$May\ x = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{2}$$

$$Then\ f^*(x) = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix} \tag{3}$$

$$\begin{aligned} f(x;\theta) &= x^T w + b \\ &= \begin{bmatrix} b & w_2 + b & w1 + b & w_1 + w_2 + b \end{bmatrix} \end{aligned} \tag{4}$$

$$\begin{aligned} \therefore J(\theta) &= \frac{1}{4}\sum_{x \in X}(f^*(x) - f(x;\theta))^2 \\ &= \frac{1}{4}(b^2 + (w_2 + b - 1)^2 + (w_1 + b - 1)^2 + (w_1 + w_2 + b)^2) \\ &= \frac{1}{4}(4b^2 + 2w_1^2 + 2w_2^2 + 2w_1w_2 + 4w_1b + 4w_2b - 4b + 2 - 2w_1 - 2w_2) \end{aligned} \tag{5}$$

$$Let\ \nabla J(\theta) = 0 We\ have:$$

$$\begin{aligned} \nabla_b J(w,b) &= 0 \\ \nabla_{w_1} J(w,b) &= 0 \\ \nabla_{w_2} J(w,b) &= 0 \end{aligned} \tag{6}$$

$$Which\ means$$

$$\begin{cases} w_1 + w_2 + 2b - 1 = 0 \\ 2w_1 + w_2 + 2b - 1 = 0 \\ 2w_2 + w_1 + 2b - 1 = 0 \end{cases} \tag{7}$$

$$Solving\ these\ equations,\ we\ get$$

$$\begin{cases} w_1 = 0 \\ w_2 = 0 \\ b = \dfrac{1}{2} \end{cases} \tag{8}$$

$$In\ conclusion, w_1 = 0, w_2 = 0, and\ b = 0.5\ will\ minimize\ the\ function\ J(w, b)$$

2 L2-regularized Linear Regression via Stochastic Gradient Descent

Code written in Python.

3 Regularization to encourage symmetry

$$Let\ S = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{9}$$

$$\begin{aligned} \frac{\alpha}{2} w^T S w &= \frac{\alpha}{2} \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \frac{\alpha}{2} \begin{bmatrix} w_1 - w_2 & -w_1 + w_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \frac{\alpha}{2} \begin{bmatrix} (w_1 - w_2) w1 & (-w_1 + w_2) w2 \end{bmatrix} \\ &= \frac{\alpha}{2} (w_1 - w_2)^2 \end{aligned} \tag{10}$$

To minimize the regularization, w_1 will be equal to w_2. At that situation, the matrix is symmetric.

4 Recursive state estimation in Hidden Markov Models

$$\because P(a|b, c) \propto P(b|a, c) P(a|c) \tag{11}$$

$$\therefore P(x_t|y_1, ..., y_t) \propto P(y_t|x_t, y_1, ..., y_{t-1}) P(x_t|y_1, ..., y_{t-1}) \tag{12}$$

$$\because P(y_t|x_t, y_1, ..., y_{t-1}) = P(y_t|x_t) \tag{13}$$

$$\therefore P(x_t|y_1, ..., y_t) \propto P(y_t|x_t) P(x_t|y_1, ..., y_{t-1}) \tag{14}$$

$$\because P(x_t|y_1,...,y_{t-1}) = \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}|y_1,...,y_{t-1}) \tag{15}$$

$$\therefore P(x_t|y_1,...,y_t) \propto P(y_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}|y_1,...,y_{t-1}) \tag{16}$$

5 Linear-Gaussian Prediction Model

$$\because P(y \mid x, w, \sigma^2) = \mathcal{N}(y; x^T w, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y - x^T w)^2}{2\sigma^2}) \tag{17}$$

$$\therefore P(D \mid w, \sigma^2) = \prod_{i=1}^{n} P(y^{(i)} \mid x^{(i)}, w) \tag{18}$$

$$\begin{aligned}
\therefore \log P(D \mid w, \sigma^2) &= log \prod_{i=1}^{n} P(y^{(i)} \mid x^{(i)}, w) \\
&= \sum_{i=1}^{n} \log P(y^{(i)} \mid x^{(i)}, w) \\
&= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y^{(i)} - x^{(i)T}w)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \sum_{i=1}^{n} \frac{(y^{(i)} - x^{(i)T}w)^2}{2\sigma^2}
\end{aligned} \tag{19}$$

$$\nabla_w \log P(D \mid w, \sigma^2) = \sum_{i=1}^{n} \frac{x^{(i)T}(y^{(i)} - x^{(i)T}w)}{\sigma^2} \tag{20}$$

$$Let\ \nabla_w \log P(D \mid w, \sigma^2) = 0 \tag{21}$$

$$Then\ \sum_{i=1}^{n} \frac{x^{(i)}(y^{(i)} - x^{(i)T}w)}{\sigma^2} = 0 \tag{22}$$

$$Then\ \sum_{i=1}^{n} x^{(i)}y^{(i)} = \sum_{i=1}^{n} x^{(i)}x^{(i)T}w \tag{23}$$

$$\therefore w = (\sum_{i=1}^{n} x^{(i)}x^{(i)T})^{-1}(\sum_{i=1}^{n} x^{(i)}y^{(i)}) \tag{24}$$

$$\nabla_{\sigma^2} \log P(D \mid w, \sigma^2) = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(y^{(i)} - x^{(i)T}w)^2}{2(\sigma^2)^2} \tag{25}$$

$$Let\ \nabla_{\sigma^2} \log P(D \mid w, \sigma^2) = 0 \tag{26}$$

$$Then \ -\frac{n}{2\sigma^2}+\sum_{i=1}^{n}\frac{(y^{(i)}-x^{(i)T}w)^2}{2(\sigma^2)^2}=0 \tag{27}$$

$$Then \ -\frac{n}{2}+\sum_{i=1}^{n}\frac{(y^{(i)}-x^{(i)T}w)^2}{2\sigma^2}=0 \tag{28}$$

$$\therefore \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(y^{(i)}-x^{(i)T}w)^2=\frac{1}{n}\sum_{i=1}^{n}(x^{(i)T}w-y^{(i)})^2 \tag{29}$$

$$\begin{gathered} In \ conclusion \ , w=(\sum_{i=1}^{n}x^{(i)}x^{(i)T})^{-1}(\sum_{i=1}^{n}x^{(i)}y^{(i)}) \\ \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x^{(i)T}w-y^{(i)})^2 \end{gathered} \tag{30}$$