HW3

Q1. Logistic Regression

$$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad y_1 = 1$$

$$x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad y_2 = -1$$

$$E_{in}(w) = \sum_{n=1}^{N} \log(\hat{P}r(y_n | x_n)) + \lambda \|w\|_2^2, \quad \lambda > 0$$

(a) Find w^* (Assume $\lambda = 0$ & $w_0 = 0$)

$$E_{in}(w) = \log(Pr(y_1 | x_1)) + \log(Pr(y_2 | x_2))$$

$$= \log\left(1 + e^{-[w_1 \; w_2]\begin{bmatrix} 1 \\ 1 \end{bmatrix}}\right) + \log\left(1 + e^{[w_1 \; w_2]\begin{bmatrix} 1 \\ 0 \end{bmatrix}}\right)$$

$$= \log\left(1 + e^{-(w_1 + w_2)}\right) + \log\left(1 + e^{w_1}\right)$$

$w_1 + w_2 = \infty$ $\qquad$ $w_1 \longrightarrow -\infty$

$-\infty + w_2 = \infty$ $\qquad$ This Term Became $\log(1) = 0$

$w_2 = 2\infty$

$$w^* = \lim_{k \to \infty} \begin{bmatrix} -k \\ 2k \end{bmatrix}$$

(b) Assume $\lambda \rightarrow \infty$, so $\|w\|_2^2 \ll 1$

$\log(1+e^{-y_n w^T x_n}) \approx \log(2) - 0.5\, y_n w^T x_n$

$E_{in}(w) = \log(1+e^{-(w_1+w_2)}) + \log(1+e^{w_1}) + \lambda(w_1^2+w_2^2)$

$\approx \log(2) - \frac{1}{2}(w_1+w_2) + \log(2) + \frac{1}{2} w_1 +$
$\lambda(w_1^2+w_2^2)$

$\frac{\partial E_{in}(w)}{\partial w_1} = -\frac{1}{2} + \frac{1}{2} + 2\lambda w_1 = 0 \longrightarrow$ To Get Minimum

$\frac{\partial E_{in}(w)}{\partial w_2} = -\frac{1}{2} + 2\lambda w_2 = 0$

$2\lambda w_2 = \frac{1}{2}$

$w_2 = \frac{1}{4\lambda}$

Q2, Multi Class Linear Classification

$x = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}, \; y \in \{1, 2, 3\}$

$\nabla_{\underline{w(1)}} \left\{ -\log \frac{e^{\underline{[w^T(1)\hat{x}_i]}}}{\sum_{j=1}^{3} e^{\underline{[w^T(j)\hat{x}_i]}}} \right\}$

$= \nabla_{\underline{w(1)}} \left[-\log e^{\underline{w^T(1)\hat{x}_i}} \right] + \nabla_{\underline{w(1)}} \left\{ \sum_{j=1}^{3} e^{\underline{w^T(j)\hat{x}_i}} \right\}$

Only $j=1$ $\quad -\hat{x}_i \quad\quad$ 3 Gradients

$$\frac{\nabla_{w(1)}}{}\left\{\sum_{j=1}^{3} e^{\frac{w^T(j)\,\hat{x}_1}{}}\right\}$$

$j = 1$

$$\frac{1}{\sum_{j=1}^{3} e^{\frac{[w^T(j)\,\hat{x}_1]}{}}}\,\hat{x}_1\, e^{\frac{w^T(1)\hat{x}_1}{}}$$

Plug in w_1, w_2, w_3, $\hat{x}_1$ and get final answer

Review Of Taylor Series

$$f(w) \approx f(v) + \nabla f(v)^T (w - v)$$

Close to v

If $w = v$, $f(w) = f(v)$

$$\log(1+e^{-yw^Tx}) \approx \log(2) - \frac{1}{2} y w^T x$$

$$\lambda \longrightarrow \infty$$

$$\|w\|_2^2 << 1$$

$$v = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

for $v = \underline{0}$

$$f(w) \approx \overbrace{f(0)}^{\log(2)} - \overbrace{\nabla f(0)^T}^{\frac{1}{2}(-yx)^T}(w - 0)$$

$$\nabla \log(1+e^{-yw^Tx}) = \frac{1}{1+e^{-yw^Tx}} (-yx) e^{-yw^Tx}$$

Example:

$$f(x) = x_1 + e^{x_2 - x_1}, \quad x = (x_1, x_2)$$

find the taylor approximation near the point $z = (1, 2)$

$$f(x) \approx \underbrace{f(z)}_{1+e^1} + \nabla f(z)^T (x - z) = 1 + e^1 + (1 - e)(x_1 - 1) + (e)(x_2 - 2)$$

$$\nabla f(z) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \Big|_z = \left(1 - e^{x_2 - x_1}\right) \Big|_z = 1 - e \\ \frac{\partial}{\partial x_2} f(x) \Big|_z = \left(e^{x_2 - x_1}\right) \Big|_z = e \end{bmatrix}$$

Softmax Regression (More General Case Of Sigmoid)

Hidden Layer $\left(\overset{\text{Layer First}}{100}, \overset{\text{Layer Second}}{200}\right)$

Adam Optimizer

Alpha, L2 Regularization