EE240: Pattern Recognition and Machine Learning
Homework 2

Yuan-Pu Hsu
#862057597

May 5, 2018

1 H2.1

1.1 a

$$f(\boldsymbol{w}) = \log(1 + exp(-\boldsymbol{w}^T\boldsymbol{x}_i))$$

We take the derivative of $f(w)$

$$\frac{\partial f(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{-\boldsymbol{x} \cdot exp(-\boldsymbol{w}^T\boldsymbol{x}_i)}{1 + exp(-\boldsymbol{w}^T\boldsymbol{x}_i)}$$

And the second derivative of $f(w)$

$$\frac{\partial^2 f(\boldsymbol{w})}{\partial \boldsymbol{w}^2} = \frac{(-\boldsymbol{x}_i \cdot exp(-\boldsymbol{w}^T\boldsymbol{x}_i))(-\boldsymbol{x}_i \cdot (1 + exp(-\boldsymbol{w}^T\boldsymbol{x}_i))) - \boldsymbol{x}_i^2 \cdot exp(-\boldsymbol{w}^T\boldsymbol{x}_i)^2}{(1 + exp(-\boldsymbol{w}^T\boldsymbol{x}_i))^2}$$

$$= \frac{\boldsymbol{x}_i^2 \cdot exp(-\boldsymbol{w}^T\boldsymbol{x}_i)}{(1 + exp(-\boldsymbol{w}^T\boldsymbol{x}_i))^2} \geq 0$$

While the the second gradient of $f(w)$ is always greater or equal to 0, $f(w)$ is a convex.

1.2 b,c

Present in h2_1.ipynb

2 H2.2

Present in h2_2.ipynb

3 H2.3

3.1 a

Present in h2_3.ipynb

3.2 b

$$\phi(\boldsymbol{x}_1) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix}, \phi(\boldsymbol{x}_2) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}, \phi(\boldsymbol{x}_3) = \begin{bmatrix} 1 \\ \sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}, \phi(\boldsymbol{x}_4) = \begin{bmatrix} 1 \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \\ \sqrt{2} \end{bmatrix}$$

For $G(i,j) = y_i y_j \phi^T(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j)$, we'll have

$$G = \begin{bmatrix} 9 & -1 & -1 & 1 \\ -1 & 9 & 1 & -1 \\ -1 & 1 & 9 & -1 \\ 1 & -1 & -1 & 9 \end{bmatrix}$$

In order to maximize $L(\alpha)$, take $\dfrac{\partial L(\alpha)}{\partial \alpha} = G\alpha + \mathbf{1} = 0$

and then get: $\alpha = \begin{bmatrix} \frac{1}{8} \\ \frac{1}{8} \\ \frac{1}{8} \\ \frac{1}{8} \end{bmatrix}$

3.3 c

$y_1 = 1, y_2 = -1, y_3 = -1, y_4 = 1$

Take the α from the previous problem, $\displaystyle\sum_{i=1}^{4} \alpha_i y_i = \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 = 0$

and $\alpha_i \geq 0$

3.4 d

Present in h2_3.ipynb

4 H2.4

Present in h2_4.ipynb