ECE595 HW4

Ruijie Song

Mar.18.2021

Ex1.

ECE 595 HW4

Ex 1

i) $h_\theta(x) = \frac{1}{1+\exp(-\theta^T x)} = \frac{1}{1+\exp(-(\underline{w}^T\underline{x}+w_0))}$

$$h(x) = \begin{cases} 1 & g(x) > 0 \quad \cancel{x \to \infty} \\ 0 & g(x) < 0 \quad \cancel{x \to -\infty} \\ \text{either} & g(x) = 0 \end{cases}$$

For $h(x) = 1$ or 0, $g(x)$ has to $\to \infty$ or $-\infty$

if $x = 0$, $w^T x + w_0 \to \infty \Rightarrow |w_0| \to \infty$

if $x \neq 0$, $w^T x \leq \frac{1}{2}(w^T w + x^T x)$ since: $(\underline{w}+\underline{x})^2 \geq 0$.

$w^T x \to \infty$ to satify $\theta^T x \to \infty$

$\therefore \|w\|_2 \to \infty$

ii) when $\|w\|_2 \to \infty$ & $|w_0| \to \infty$, $\Rightarrow (h_\theta(x_n) - y_n) \to 0$

Then $\|\theta^{(k+1)} - \theta^{(k)}\|_2 \Rightarrow 0$

But GD needs ∞ steps to ~~make~~ make $\|\theta^{(k+1)} - \theta^{(k)}\|_2 = 0$

iii) $\|w\|_2 \leq c_1$, $|w_0| < c_2$ for $c_1, c_2 > 0$

One way: $\|\theta^{(k+1)} - \theta^{(k)}\|_2 \leq$ Const.

Another way: Change α_k

iv) No. Linear regression.

Ex.2

Ex 2.

$$J(\theta) = -\frac{1}{N}\sum_{n=1}^{N}\{y_n \log h_\theta(x_n) + (1-y_n)\log(1-h_\theta(x_n))\}$$

$$h_\theta(x) = \frac{1}{1+\exp\{-\theta^T x\}} \qquad J = \sum_{n=1}^{N} -\frac{1}{N}\left(y_n \log h_\theta(x_n) + (1-y_n)\log(1-h_\theta(x_n))\right)$$

$y_n \log h_\theta(x_n)$ ~~covex~~ convex

$$\nabla_\theta[-\log(1-h_\theta(x_n))] = \nabla_\theta\left[-\log\left(1-\frac{1}{1+\exp(-\theta^T x)}\right)\right]$$

$$= \nabla_\theta\left[-\log\left(\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right)\right] = \nabla_\theta\left(-\log(e^{-\theta^T x}) + \log(1+e^{-\theta^T x})\right)$$

$$= \nabla_\theta\left(+\theta^T x + \log(1+e^{-\theta^T x})\right) = +x + \nabla_\theta \log(1+e^{-\theta^T x})$$

$$= +x + \left(\frac{-e^{-\theta^T x}}{1+e^{-\theta^T x}}\right)x = +h_\theta(x)x$$

Hessian: $\nabla_\theta^2[-(1-y_n)\log(1-h_\theta(x_n))] = \nabla_\theta[+h_\theta(x)x]\cdot(1-y_n)$

$$= +\nabla_\theta\left[\left(\frac{+1}{1+e^{-\theta^T x}}\right)x\right] = +\left(\frac{1}{(1+e^{-\theta^T x})^2}\right)(-e^{-\theta^T x})xx^T$$

$$= +\left(\frac{1}{1+e^{-\theta^T x}}\right)\left(1-\frac{1}{1+e^{-\theta^T x}}\right)xx^T = +h_\theta(x)[1-h_\theta(x)]xx^T \cdot (1-y_n)$$

For $v \in R^d$

$\cancel{v^T \nabla_\theta^2[-\log(1-h_\theta(x))]v = (h_\theta(x)[1-h_\theta(x)])\|v^T x\|^2 \geq 0}$

$\therefore$ ~~$(1-y_n)\log(1-h$~~ $\qquad (1-y_n) \geq 0$

$$v^T \nabla_\theta^2[-\log(1-h_\theta(x_n))] = h_\theta(x)[1-h_\theta(x)])\|v^T x\|^2 \geq 0$$

$\therefore$ Hessian is positive semi-definite.

$-(1-y_n)\log(1-h_\theta(x_n))$ is convex in θ.

Ex3.

a.

3.

a) $$J(\underline{\theta}) = \frac{-1}{N}\sum_{n=1}^{N}\{y_n \log h_{\underline{\theta}}(\underline{x}_n) + (1-y_n)\log(1-h_{\underline{\theta}}(\underline{x}_n))\}$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\{y_n \log h_{\underline{\theta}}(\underline{x}_n) + \log(1-h_{\underline{\theta}}(\underline{x}_n)) - y_n \log(1-h_{\underline{\theta}}(\underline{x}_n))\}$$

$$= \frac{-1}{N}\sum^{N}\{y_n \log\left(\frac{h_{\underline{\theta}}(\underline{x}_n)}{1-h_{\underline{\theta}}(\underline{x}_n)}\right) + \log(1-h_{\underline{\theta}}(\underline{x}_n))\}$$

$$= \frac{-1}{N}\sum\{y_n \underline{\theta}^T\underline{x}_n + \log(1-h_{\underline{\theta}}(\underline{x}_n))\}$$

$$= \frac{-1}{N}\sum\{y_n \underline{\theta}^T\underline{x}_n + \log\left(1-\frac{1}{1+e^{-\underline{\theta}^T\underline{x}_n}}\right)\}$$

$$= \frac{-1}{N}\sum\{y_n \underline{\theta}^T\underline{x}_n + \log\left(\frac{e^{-\underline{\theta}^T\underline{x}_n}}{1+e^{-\underline{\theta}^T\underline{x}_n}}\right)\}$$

$$= \frac{-1}{N}\sum\{y_n \underline{\theta}^T\underline{x}_n + \log\left(\frac{1}{1+e^{\underline{\theta}^T\underline{x}_n}}\right)\}$$

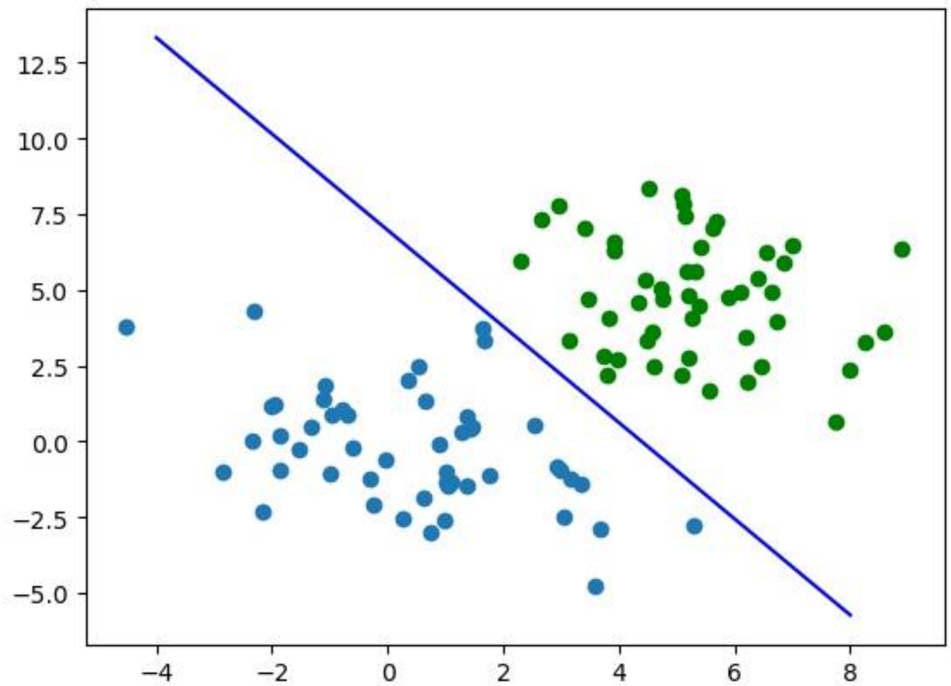
$$= \frac{-1}{N}\sum\{y_n \underline{\theta}^T\underline{x}_n - \log(1+e^{\underline{\theta}^T\underline{x}_n})\}$$

$$= \frac{-1}{N}\cdot\{(\sum y_n\underline{x}_n)^T\underline{\theta} - \sum\log(1+e^{\underline{\theta}^T\underline{x}_n})\}$$

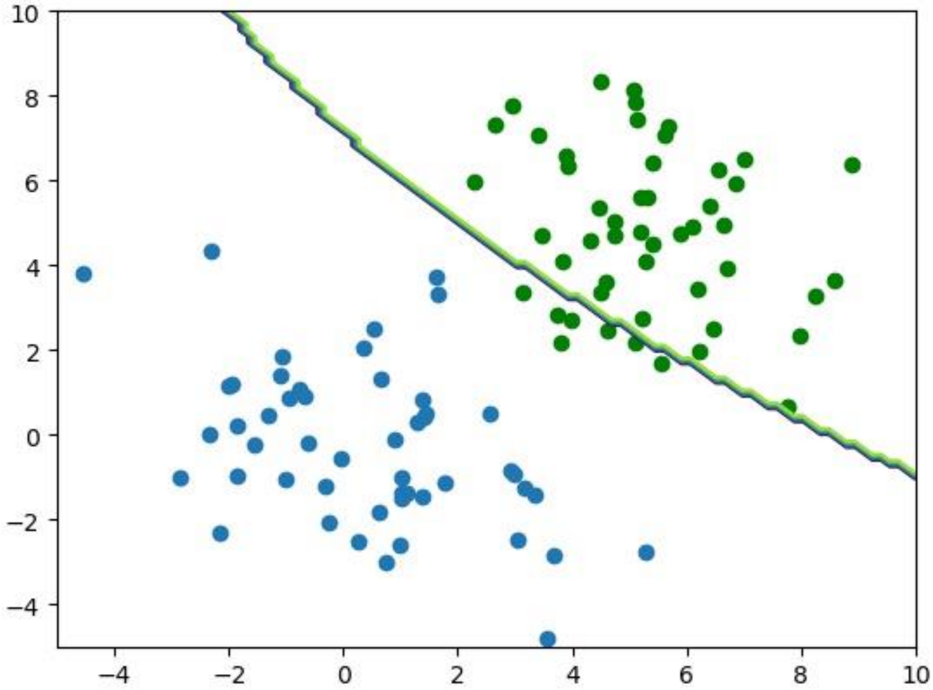
b.

theta = [2.38, 1.50, -10.44]

c.



d.



Ex4.

a.

[[1.00000000e+00 5.78667187e-13 6.19586231e-20 7.94511151e-25
4.31787909e-17]
[5.78667187e-13 1.00000000e+00 2.66618165e-12 2.63683326e-31
1.07334027e-11]
[6.19586231e-20 2.66618165e-12 1.00000000e+00 3.36906580e-42
1.49290933e-18]
[7.94511151e-25 2.63683326e-31 3.36906580e-42 1.00000000e+00
1.20563730e-05]
[4.31787909e-17 1.07334027e-11 1.49290933e-18 1.20563730e-05
1.00000000e+00]]

b.

4 b) $\theta = \sum_{n=1}^{N} \alpha_n x_n$, $\theta^T x = \sum \alpha_n \langle x_n, x \rangle$

$$J(\theta) = \frac{-1}{N} \left\{ \left(\sum y_n x_n\right)^T \left(\sum \alpha_m x_m\right) - \sum \log\left(1 + e^{x_n^T (\sum \alpha_m x_m)}\right) \right\} + \lambda \|\sum \alpha_n x_n\|^2$$

$$= \frac{-1}{N} \left\{ \sum_{n=1}^{N} y_n \left(\sum \alpha_m \underbrace{x_n^T x_m}_{K(x_n, x_m)} \right) - \sum \log\left(1 + e^{\sum \alpha_m \underbrace{(x_n^T x_m)}_{K(x_n, x_m)}}\right) \right\} + \lambda \|\sum \alpha_n x_n\|^2$$

$\therefore \sum \alpha_m K(x_n, x_m)$

$= K_n^T \alpha$

$$= \frac{-1}{N} \left\{ \sum y_n K_n^T \alpha - \sum \log\left(1 + e^{K_n^T \alpha}\right) \right\} + \lambda \alpha^T X^T X \alpha$$

$$= \frac{-1}{N} \left\{ y^T K \alpha - 1^T \left(\log\left(1 + e^{K\alpha}\right)\right) \right\} + \lambda \alpha^T K \alpha$$

$$= \frac{-1}{N} \left\{ y^T K \alpha - 1^T \log\left(e^0 + e^{K\alpha}\right) \right\} + \lambda \alpha^T K \alpha$$

c.

[-0.52, -0.60]

d.