ML HW2 : Regression and Gradient Descent

Xinyue Chen (xc1305)

Problem 1:

1. $J(w) = \frac{1}{2\cdot 4}\sum_{i=1}^{4}\left[h_w(x^{(i)}) - y^{(i)}\right]^2$

$$= \frac{1}{8}\left[(w_0+1600w_1+1770w_2+3w_3-330)^2+(w_0+2400w_1+2740w_2+3w_3-369)^2+(w_0+1416w_1+1634w_2+2w_3-232)^2+(w_0+3000w_1+3412w_2+4w_3-540)^2\right]$$

2. $w = (w_0, w_1, w_2, w_3) = (0,0,0,0)$

$$\widetilde{w_0} = w_0 - 0.1\times\frac{1}{4}\left[(0-330)\cdot 1+(0-369)\cdot 1+(0-232)\cdot 1+(0-540)\cdot 1\right] = 36.775$$

$$\widetilde{w_1} = w_1 - 0.1\times\frac{1}{4}\left[(0-330)\cdot 1600+(0-369)\cdot 2400+(0-232)\cdot 1416+(0-540)\cdot 3000\right] = 84052.8$$

$$\widetilde{w_2} = w_2 - 0.1\times\frac{1}{4}\left[(0-330)\cdot 1770+(0-369)\cdot 2740+(0-232)\cdot 1634+(0-540)\cdot 3412\right] = 95418.2$$

$$\widetilde{w_3} = w_3 - 0.1\times\frac{1}{4}\left[(0-330)\cdot 3+(0-369)\cdot 3+(0-232)\cdot 2+(0-540)\cdot 4\right] = 118.025$$

$$\widetilde{w} = (\widetilde{w_1}, \widetilde{w_2}, \widetilde{w_3}, \widetilde{w_4}) = (36.775, 84052.8, 95418.2, 118.025)$$

Problem 2:

A. (a). $\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$

$$\sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_i - \frac{\sum_{j=1}^{m} x_j}{m}\right)^2}$$ (population standard deviation)

B. (a). $J(w) = J(w_0, w_1, w_2) = \frac{1}{2m}\sum_{i=1}^{m}\left(w_0 + w_1x_1^{(i)} + w_2x_2^{(i)} - y^{(i)}\right)^2$

Problem 3:

1. $\frac{\partial J(w)}{\partial w_j} = \frac{1}{m}\left[\sum_{i=1}^{m}(f_w(x^{(i)})-y^{(i)})x_j^{(i)} + \lambda\sum_{j=1}^{n} w_j\right]$

$$\widetilde{w_j} = w_j - \alpha\cdot\frac{1}{m}\left[\sum_{i=1}^{m}(f_w(x^{(i)})-y^{(i)})x_j^{(i)} + \lambda\sum_{j=1}^{n} w_j\right]$$

2. (a). $f_{reg}(w) = \lambda\sum_{j=1}^{n} w_j^2$,

$\forall w, w'$, $\forall c\in[0,1]$,

$$f_{reg}(cw+(1-c)w') = \lambda\sum_{j=1}^{n}\left[cw_j+(1-c)w_j'\right]^2 = \lambda\sum_{j=1}^{n}\left[c^2w_j^2+2c(1-c)w_jw_j'+(1-c)^2w_j'^2\right]$$

$$c f_{reg}(w)+(1-c)f_{reg}(w') = c\lambda\sum_{j=1}^{n}w_j^2+(1-c)\lambda\sum_{j=1}^{n}w_j'^2 = \lambda\sum_{j=1}^{n}\left[cw_j^2+(1-c)w_j'^2\right]$$

since $c f_{reg}(w)+(1-c)f_{reg}(w') - f_{reg}(cw+(1-c)w') = \lambda\sum_{j=1}^{n}\left[(c-c^2)w_j^2+(1-c-(1-c)^2)w_j'^2-2c(1-c)w_jw_j'\right]$

$$= \lambda\sum_{j=1}^{n}\left[c(1-c)w_j^2+(1-c)cw_j'^2-c(1-c)w_j\right]$$

$$= \lambda\sum_{j=1}^{n}\left[c(1-c)\cdot(w_j^2+w_j'^2-2w_jw_j')\right]$$

$$= \lambda c(1-c) \sum_{j=1}^{n} (w_j - w_j')^2 \geq 0^{*}$$

$*$ because $\lambda > 0$, $c \in [0,1] \Rightarrow c \geq 0$, $1-c \geq 0$;

and $\sum_{j=1}^{n} (w_j - w_j')^2 \geq 0$ since it is a sum of squares, $w_j, w_j' \in \mathbb{R}$

Thus $f_{reg}(cw + (1-c)w') \leq c f_{reg}(w) + (1-c) f_{reg}(w')$ for $\forall\, w, w'$, $\forall\, c \in [0,1]$,

i.e., $f_{reg}(w)$ is a convex function.

(b) Since, f_1, f_2 are both convex functions, for $\forall\, x, x'$, $\forall c \in [0,1]$,

$$\begin{cases} f_1(cx + (1-c)x') \leq c f_1(x) + (1-c) f_1(x') \\ f_2(cx + (1-c)x') \leq c f_2(x) + (1-c) f_2(x') \end{cases}$$

$$\begin{aligned} (f_1+f_2)(cx + (1-c)x') &= f_1(cx+(1-c)x') + f_2(cx+(1-c)x') \\ &\leq c f_1(x) + (1-c) f_1(x') + c f_2(x) + (1-c) f_2(x') \\ &= c[f_1(x) + f_2(x)] + (1-c)[f_1(x') + f_2(x')] \\ &= c(f_1+f_2)(x) + (1-c)(f_1+f_2)(x') \end{aligned}$$

By the definition of a convex function, $(f_1+f_2)(x)$ is convex.

(c) Since $J'(w) = \frac{1}{2m}\left[\ [f_w(x^{(i)}) - y^{(i)}]^2\right.$ and $f_{reg}(w) = \lambda \sum_{j=1}^{n} w_j^2$ are both convex, the new loss function $J(w) = J'(w) + f_{reg}(w)$, (using L2 regularization), $J(w)$ is also convex.