1.(1) $\nabla_w J(w) = 2X^T(Xw-y)$

$H_w J(w) = \frac{\partial \nabla_w J(w)}{\partial w^T} = 2X^TX$

牛顿法最优解满足 $\nabla_w J(w)=0$

即：$X^TXw = X^Ty$

$w = (X^TX)^{-1}X^Ty$

而 $w - H^{-1}\nabla_w J(w)$

$= w-(2X^TX)^{-1}2X^T(Xw-y)$

$= w-(X^TX)^{-1}(X^TX)w + (X^TX)^{-1}X^Ty$

$= (X^TX)^{-1}X^Ty$，为最优解

(2) 将 $x^{(i)}, y^{(i)}$ 组合成矩阵 $X \in R^{n\times d}$, $Y \in R^{n\times p}$,

有 $J(\theta) = \frac{1}{2}\|X\theta - Y\|_F^2$

$= \frac{1}{2} tr((X\theta-Y)^T(X\theta-Y))$

$\nabla_\theta J(\theta) = X^TX\theta - X^TY$,

令 $\nabla_\theta J(\theta)=0$,

解得 $\theta = (X^TX)^{-1}X^TY$

2. (1) $\|Xw-y\|_2^2+\lambda\|w\|_1$

$J_1(w) = (Xw-y)^T(Xw-y)+\lambda\|w\|_1$

$= w^TX^TXw - 2y^TXw + y^Ty + \lambda\|w\|_1$

$= n\|w\|_2^2 + \|y\|_2^2 - 2y^TXw + \lambda\|w\|_1$, $= \|y\|_2^2 + \sum_{i=1}^{d}(nw_i^2 + \lambda|w_i| - 2y^TX_{*i}w_i)$

(2) $\frac{\partial J_1(w)}{\partial w_i} = \begin{cases} 2nw_i + \lambda - 2y^TX_{*i} & (w_i>0) \\ 2nw_i - \lambda - 2y^TX_{*i} & (w_i<0) \end{cases}$

$\frac{\partial^2 J_1(w)}{\partial w_i^2} = 2n$ $w_i>0$ 或 $w_i<0$

若存在最优解，当仅当偏导在 $\begin{cases} w_i>0 & 恒正 \\ w_i<0 & 恒负 \end{cases}$

有 $\forall w_i>0$, $w_i > \frac{2y^TX_{*i}-\lambda}{2n}$

$\forall w_i<0$, $w_i < \frac{2y^TX_{*i}+\lambda}{2n}$

即：$\begin{cases} 2y^TX_{*i}-\lambda \le 0 \\ 2y^TX_{*i}+\lambda \ge 0 \end{cases}$

即条件 $-\frac{\lambda}{2} \le y^TX_{*i} \le \frac{\lambda}{2}$ 即可

(3) $J_2(w)$

$= n\|w\|_2^2 + \lambda\|w\|_2^2 + \|y\|_2^2 - 2y^TXw$

$= \|y\|_2^2 + \sum_{i=1}^{d}((n+\lambda)w_i^2 - 2y^TX_{*i}w_i)$

$\frac{\partial J_2(w)}{\partial w_i} = 2(n+\lambda)w_i - 2y^TX_{*i}$ $=0$ 时，$w_i = \frac{y^TX_{*i}}{n+\lambda}$

$\frac{\partial^2 J_2(w)}{\partial w_i^2} = 2(n+\lambda) > 0$ 恒正.

充要条件为 $y^TX_{*i}=0$

(4) L_1 更易：

L_1: $-\frac{\lambda}{2} \le y^TX_{*i} \le \frac{\lambda}{2}$ 存在 $w_i=0$

L_2: 仅 $y^TX_{*i}=0$ 才存在 $w_i=0$

前者更易，范围更大

3. $\nabla_\theta h_\theta(x) = h_\theta(x)(1-h_\theta(x))$

因此有：$\nabla \log h_\theta(x^{(i)}) = (1-h_\theta(x^{(i)}))x^{(i)}$

$\nabla \log(1-h_\theta(x^{(i)})) = -h_\theta(x^{(i)})x^{(i)}$

∴有 $\nabla l(\theta) = \lambda\theta + \sum_{i=1}^{n} w^{(i)}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$

有 $H l(\theta) = \lambda I + \sum_{i=1}^{n} w^{(i)} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x^{(i)}x^{(i)T}$

~~1: $\theta \leftarrow \theta$~~

~~2: while True:~~

~~$w^{(i)} \leftarrow \exp\left(-\frac{\|x-x^{(i)}\|^2}{2\tau^2}\right)$~~

~~3:~~

伪代码：

函数 train(X, y, θ_0)

1: $\theta \leftarrow \theta_0$;

2: for i in 1--n

3: $w^{(i)} \leftarrow \exp\left(-\frac{\|x-x^{(i)}\|^2}{2\tau^2}\right)$;

4: while True:

5: $\nabla l(\theta) = \lambda\theta + \sum_{i=1}^{n} w^{(i)}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$;

6: $H(\theta) = \lambda I + \sum_{i=1}^{n} w^{(i)} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x^{(i)}x^{(i)T}$;

7: if $(\|\nabla l(\theta)\|_2 < \varepsilon)$

8: break;

9: $\theta = \theta - H(\theta)^{-1}\nabla l(\theta)$;

10: return θ;

函数 test(X, y, θ)

1. sum ← 0

2. for i in 1--n

3. result_yi $= \theta(x^{(i)})$

4. if(result_yi == $y^{(i)}$)

5. sum += 1

6. acc = sum/n

7. return acc

main:

1. X_train, y_train, X_val, y_val = getdata()

2. $\theta \leftarrow$ train(X_train, y_train, θ)

3. acc ← test(X_val, y_val, θ)

扫描全能王 创建