$$\frac{\partial J}{\partial \theta} = 2X^T X \theta - 2X$$

Problem 6

$$J(w) = \sum (y_i - w^T x_i)^2$$

$$\underset{n\times d}{X} = \begin{bmatrix} (x_1)^T \\ (x_2)^T \\ \vdots \\ (x_n)^T \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \underset{d\times 1}{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

$$J(w) = (Y - XW)^T (Y - XW)$$

$$= \left((XW)^T - Y^T \right) (XW - Y)$$

$$= (XW)^T (XW) - (XW)^T Y - Y^T (XW) + Y^T Y$$

$$= (XW)^T (XW) - 2(X\theta)^T Y + Y^T Y$$

J with respect to w is quadratic so it's convex function and has a global minimum which is found by derivation

→ Problem 6 (contine)

$$\frac{\partial J}{\partial W} = 2X^TXW - 2X^Ty = 0$$

$$\rightarrow \hat{W} = (X^TX)^{-1}X^Ty$$

first of all, the problem is high computational complexity due to calculate inverse matrix by size $(n+1)\times(n+1)$

second problem happend when the matrix $A = X^TX$ isn't invertible

for solve this problems, we can use approximate methos for calculating inverse matrix like psudu-inverse of matrix

also we can use proper regularization parameter to turn A matrix into invertible matrix

ج	پ	چ	س	د	ی	ش
۳	۲	۱				
۱۰	۹	۸	۷	۶	۵	۴
۱۷	۱۶	۱۵	۱۴	۱۳	۱۲	۱۱

۹۹/۰۵/۳

c) if one set of data is linear combination of two other ~~sets~~ data sets, ~~it means~~ we can not use general formula for Normal equation because ~~the~~ $A = X^T X$ is not invertible

it means one feature is depond on two other features

for solving ~~the~~ this problem we shoull agin do feature selection and select features that independent of each other

۹۹/۰۵/۱

D)

$$J(W) = \frac{1}{n} \sum_{t=1}^{n} \left[y^{(t)} - w x^{(t)} \right]^2 + \lambda \|W\|^2$$

$$\frac{\partial}{\partial w_i} (J(w)) = \frac{1}{n} \sum_{t=1}^{n} \left[(y^{(t)} - (x^{(t)})^T w \right] (x_i^{(t)})$$

$+ \lambda \theta_i$

setting the partial to zero:

$$\frac{1}{n} \sum_{t=1}^{n} x_i^{(t)} (x^t)^T \overline{w} + \lambda w_i = \frac{1}{n} \sum_{t=1}^{n} x_i^{(t)} y^{(t)}$$

thus, callapsing the equations for all partials:

$$\frac{1}{n} \sum_{t=1}^{n} x^{(t)} (x^t)^T w + \lambda w = \frac{1}{n} \sum_{t=1}^{n} x^t y^t$$

so in the matrix form we have

مرداد چهارشنبه

Jul. 2020

۱ ذی الحجه

۹۹/۰۴/۳۱

D) $\cancel{\frac{1}{n} x^T x \ldots}$

$$\frac{1}{n} x^T x w = \frac{1}{n} x^T y$$

$$\rightarrow \left[\frac{1}{n} x^T x + \lambda I\right] w = \frac{1}{n} x^T y$$

$$\rightarrow \boxed{\hat{w} = \left[\frac{1}{n} x^T x + \lambda I\right]^{-1} \cdot \frac{1}{n} x^T y}$$

E) $J_F(w) = \sum_{i=1}^{n} F_i (y_i - w^T x)^2$

$$= (Y - Xw)^T F (Y - Xw)$$

from problem 7 we know.

if $y_i = w_0 + w_1 x^i + \epsilon_i$

so $var(\epsilon_i) = x_i^2 \sigma^2$

we can transform this into a regular least squares problem by taking

$$y_i' = \frac{y_i}{x_i} \qquad x_i' = \frac{1}{x_i} \qquad \epsilon_i' = \frac{\epsilon_i}{x_i}$$

Then the model is

$$y_i' = \beta_1 + \beta_0 x_i' + \epsilon_i'$$

where $var(\epsilon_i') = \sigma^2$

The residual sum of squares for the transformed model is

$$S_1(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i' - \beta_1 - \beta_0 x_i')^2$$

$$= \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \beta_1 - \frac{\beta_0}{x_i} \right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{1}{x_i} \right)^2 (y_i - \beta_0 - \beta_1 x_i)^2$$

This is the weighted residual sum of squares with $\boxed{w_i = \frac{1}{x_i^2}}$

Fri	Thu	Wed	Tue	Mon	Sun	Sat
26	25	24	23	22	21	
3	2	1	30	29	28	27
10	9	8	7	6	5	4

۹۹/۰۴/۲۹

Hence the weighted least square solutions is the same as the regular least square solution of the transformed model.

in general we have

$$y = X\beta + \epsilon \qquad \text{where} \qquad Var(\epsilon) = W^{-1}\sigma^2$$

~~Let~~ The we have $Var(W^{\frac{1}{2}}\epsilon) = \sigma^2 I_n$

because $W^{\frac{1}{2}}$ be a diagnal matrix with diagnal entiries equal to $\sqrt{w_i}$

Hence we have

$$y' = W^{\frac{1}{2}}y \qquad X' = W^{\frac{1}{2}}X \qquad \epsilon' = W^{\frac{1}{2}}\epsilon.$$

$$\rightarrow y' = X'\beta + \epsilon'$$

Using the results from regular least squres we the get the solution

$$\cancel{\hat{w} = \left(2(X')^T X'\right)^{-1} 2 (X')^T y'}$$

$$\cancel{= (X^T W X)^{-1} X^T W}$$

$$\rightarrow \cancel{\hat{w} = (X^T P X)^{-1}}$$

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y$$

and with first notation we have

$$\boxed{\hat{w} = (X^T F X)^{-1} (X^T F Y)}$$