Assignment 3

1) given, $\hat{y}_i = x_i \hat{\beta}$, where

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i y_i\right) / \left(\sum_{i=1}^{n} x_i^2\right)$$

To show, $\hat{y}_i = \sum_{i=1}^{n} a_i y_i$

$$\rightarrow \quad \hat{y}_i = x_i \frac{\left(\sum_{i=1}^{n} x_i y_i\right)}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{y}_i = \left(\frac{x_i}{\sum_{i=1}^{n} x_i^2}\right)\left(\sum_{i=1}^{n} x_i y_i\right)$$

let ~~weight~~ $\frac{x_i}{\sum x_i^2} = w_i$ (for each data point)

$$\hat{y}_i = w_i \left(\sum_{i=1}^{n} x_i y_i\right)$$

$$\left[\hat{y}_i = \sum_{i=1}^{n} (w_i x_i) y_i = \sum_{i=1}^{n} a_i y_i\right]$$

$\rightarrow$ so $[a_i = w_i x_i]$, where $w_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}$

2) Given, $J(\beta) = \frac{1}{2}\|X\beta - Y\|_2^2 + \frac{\lambda}{2}\|\beta\|^2$

$\rightarrow J(\beta) = \frac{1}{2}(X\beta - Y)^T(X\beta - Y) + \frac{\lambda}{2}\beta^T\beta$

$J(\beta) = \frac{1}{2}(\beta^T X^T - Y^T)(X\beta - Y) + \frac{\lambda}{2}\beta^T\beta$

$J(\beta) = \frac{1}{2}[\beta^T X^T X\beta - \beta^T X^T Y - Y^T X\beta + Y^T Y] + \frac{\lambda}{2}\beta^T\beta$

$\therefore$ β is scalar $\beta = \beta^T$

$J(\beta) = \frac{1}{2}[X^T X\beta^2 - 2\beta X^T Y + Y^T Y] + \frac{\lambda}{2}\beta^2$

$\frac{\partial(J(\beta))}{\partial\beta} = X^T X\beta - X^T Y + \lambda\beta = 0$

$(X^T X + \lambda I)\beta = X^T Y$

$$\boxed{\beta = (X^T X + \lambda I)^{-1} X^T Y}$$

$\frac{\partial^2(J(\beta))}{\partial\beta} = X^T X + \lambda > 0$

So, $\beta = (X^T X + \lambda I)^{-1} X^T Y$ minimizes $J(\beta)$.

3) given a design matrix X

n-vector of labels $y = [y_1, \dots y_n]^T$

$\rightarrow X^T X = nI$

let x_{*i} denote the i^{th} column of X.

a) for L_1 - regularized least squares, we have

$$J(\beta) = \|X\beta - Y\|^2 + \lambda\|\beta\|$$

$$J(\beta) = (X\beta - Y)^T (X\beta - Y) + \lambda\|\beta\|$$

$$J(\beta) = (\beta^T X^T - Y^T)(X\beta - Y) + \lambda\|\beta\|$$

$\therefore \beta$ is scalar $\beta^T = \beta$

$$J(\beta) = \|\beta\|^2 X^T X - 2\|\beta\| X^T Y + \|Y\|^2 + \lambda\|\beta\|$$

$$J(\beta) = \|Y\|^2 + n\|\beta\|^2 + \|\beta\|(\lambda - 2X^T Y)$$

~~for~~ we can also write above equation in terms of summation of the calculation for each column.

$$\left[J(\beta) = \|Y\|^2 + \sum_{i=1}^{d} \|\beta_i\| (n\|\beta_i\| + \lambda - 2x_{*i} Y_i)\right]$$

$$\left[J(\beta) = \|Y\|^2 + \sum_{i=1}^{d} f(x_{*i}, \beta_i)\right]$$

where $f(x_{*i}, \beta_i) = \|\beta_i\|(n\|\beta_i\| + \lambda - 2x_{*i}Y)$

$$\boxed{f(x_{*i}, \beta_i) = n\beta_i^2 - 2Y^T x_i \beta_i + \lambda|\beta_i|}$$

b) we have

$$J(\beta) = ||Y||^2 + \sum_{i=1}^{d} (n\beta_i^2 - 2x_i^T Y \beta_i + \lambda|\beta_i|)$$

i) case $\beta_j > 0$

$$\frac{\partial(J(\beta_j))}{\partial \beta_j} = 2n\beta_i - 2Y^T x_{*i} + \lambda = 0$$

$$\left[\beta_i = \frac{1}{n}\left(Y^T x_{*i} - \frac{\lambda}{2}\right)\right]$$

since $Y^T x_{*i} - \frac{\lambda}{2}$ can be < 0. So, we will take

$$\left[\beta_i = \max\left\{\frac{1}{n}\left(Y^T x_{*i} - \frac{\lambda}{2}\right), 0\right\}\right]$$

ii) $\beta_i < 0$

by similar calculation

$$\beta_i = \frac{1}{n}\left(Y^T x_{*i} + \frac{\lambda}{2}\right)$$

and $\beta_i = \min\left\{\frac{1}{n}\left(Y^T x_{*i} + \frac{\lambda}{2}\right), 0\right\}$

iii) when $\beta_i = 0$, for other cases

So

$$\hat{\beta}_i^* = \begin{cases} \min\left\{\frac{1}{n}\left(Y^T x_{*i} + \frac{\lambda}{2}\right), 0\right\} & \text{if } \beta_i < 0 \\ \max\left\{\frac{1}{n}\left(Y^T x_{*i} - \frac{\lambda}{2}\right), 0\right\} & \text{if } \beta_i > 0 \\ 0, & \cancel{\text{if } \beta_i = 0} \text{ otherwise} \end{cases}$$

c) given $J_2(\beta) = \|X\beta - Y\|^2 + \lambda\|\beta\|^2$, $\lambda > 0$

- $J_2(\beta_i) = (X\beta_i - Y)^T(X\beta_i - Y) + \lambda\beta_i^2$

$J_2(\beta_i) = (\beta_i X^T - Y^T)(X\beta_i - Y) + \lambda\beta_i^2$

$J_2(\beta_i) = \beta_i^2 X^T X - 2\beta_i X^T Y - Y^T Y + \lambda\beta_i^2$

$$\frac{\partial(J_2(\beta_i^{\#}))}{\partial\beta_i^{\#}} = 2\beta_i^{\#} X^T X - 2X^T Y + 2\lambda\beta_i^{\#} = 0$$

$$\left[(X^T X + \lambda)\beta_i^{\#} = X^T Y\right]$$

→ ∴ $X^T Y$ will never going to be zero

and λ is > 0 , hence

there is no condition such that

$\beta_i^{\#} = 0$

∴ $\beta_i^{\#}$ will always have non-zero values

d) from our solution β^* will going to be sparse because it is possible for it to have its components 0 but it is not possible for $\beta^{\#}$.