STAT 239 HW 3
THEORETICAL QUESTIONS

1) $$Var\left(x^T \hat{\beta}_{RR}\right) = Var\left(\sum_{j=1}^{p} \mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T y\right)$$

$$\Leftrightarrow Var(x) = \sum_{j=1}^{p} Var\left(\mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T y\right)$$

$$\Leftrightarrow Var(x) = \sum_{j=1}^{p} \left(\frac{y^T \mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T \mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T y}{N} \right)$$

$$\Leftrightarrow Var(x) = \sum_{j=1}^{p} \frac{d_j^4}{(d_j^2 + \lambda)^2} \frac{y^T \mu_j \mu_j^T y}{N}$$

Since $X^T X$ is diagonal, it is a scaling matrix, therefore its eigenbasis is equal to its basis $\Rightarrow V = V^T = I \Rightarrow X^T X = D^2 \Rightarrow x_i^2 = d_i^2$ $\forall i \in \{1, \dots, p\}$

Also, $X^T X$ is full rank, therefore $r = s \Rightarrow U \in O(r)$

$\Rightarrow \mu_j \mu_j^T = I$.

$$\Leftrightarrow Var(x) = \sum_{j=1}^{p} \frac{d_j^4}{(d_j^2 + \lambda)^2} \frac{y^T y}{N}$$

$$\Leftrightarrow Var(x) = \sigma^2 \sum_{j=1}^{p} \frac{d_j^4}{(d_j^2 + \lambda)^2}$$

$$B(x) = E\left[\hat{F}(x)\right] - F(x)$$

$$\Leftrightarrow \quad B(x) = E\left[x^T \hat{\beta}_{RR}\right] - x^T \beta$$

$$\Leftrightarrow \quad B(x) = E\left[\sum_{j=1}^{p} u_j \frac{d_j^2}{d_j^2 + \lambda} u_j^T Y\right] - \sum_{i=1}^{p} x_i \beta_i$$

$$\Leftrightarrow \quad B(x) = \sum_{j=1}^{p} E\left[u_j \frac{d_j^2}{d_j^2 + \lambda} u_j^T Y\right] - \sum_{i=1}^{p} x_i \beta_i$$

$$\Leftrightarrow \quad B(x) = \sum_{j=1}^{p} x_i \beta_i \left(\frac{d_j^2}{d_j^2 + \lambda} - 1\right)$$

since $E[\varepsilon] = 0$, $E[Y] = F(x) = x^T \beta$.

2) ~~$M = (\beta - \hat{\beta}_{OLS})^T(\beta - \hat{\beta}_{OLS}) + \lambda \sum_{i=1}^{p} |\beta_i|$ is the quantity to be minimized.~~

Expressing the problem in Lagrangian form:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Ignoring β_0 and assuming orthogonality, we get

$$M = \frac{1}{2}(\beta - \hat{\beta}_{OLS})^T(\beta - \hat{\beta}_{OLS}) + \lambda \sum_{i=1}^{p} |\beta_i|$$

For M to be minimal, β_{LASSO} can't be of a different sign than $\hat{\beta}_{OLS}$ or else $\|\beta - \hat{\beta}_{OLS}\|_2$ would not be minimal. (*)

~~Case 1~~ Fix $i \in \{1, \dots p\}$

Case 1 : $\beta_{OLS\,i} > 0$

$$M' = \frac{1}{2}\beta_i^2 - \beta_i \hat{\beta}_{OLS\,i} + \frac{1}{2}\hat{\beta}_{OLS\,i}^2 + \lambda \beta_i \quad \text{where } M = \sum_{D=1}^{p} M'$$

$$\Leftrightarrow M' = \frac{1}{2}\beta_i^2 - \beta_i(\hat{\beta}_{OLS\,i} - \lambda) + \frac{1}{2}\hat{\beta}_{OLS\,i}^2$$

$$\frac{\partial M'}{\partial \beta_i} = \beta_i - (\hat{\beta}_{OLS\,i} - \lambda)$$

Setting the derivative to 0 (which can't be a maximum by the argument given in (*)) gives us

$$\beta_i = (\hat{\beta}_{OLS\,i} - \lambda)^+ = (|\hat{\beta}_{OLS\,i}| - \lambda)^+ \text{ since } \hat{\beta}_{OLS\,i} > 0.$$

and $\hat{\beta}_{OLS} - \lambda$

$\beta_i = \text{sign}(\beta_{OLS\,i})(|\hat{\beta}_{OLS\,i}| - \lambda)^+$ must be positive for M' to be as small as possible.

Case 2 : $\hat{\beta}_{ols\,i} < 0$

$$M' = \frac{1}{2}\beta_i^2 - \beta_i(\hat{\beta}_{ols\,i} + \lambda) + \frac{1}{2}\hat{\beta}_{ols\,i}^2$$

$$\frac{\partial M'}{\partial \beta_i} = 0 \iff \beta_i - (\hat{\beta}_{ols\,i} + \lambda) = 0$$

$$\iff \beta_i = (\hat{\beta}_{ols\,i} + \lambda)^+$$

$$\iff \beta_i = (-|\hat{\beta}_{ols\,i}| + \lambda)^+ \quad \text{since } \hat{\beta}_{ols\,i} < 0$$

$$\iff \beta_i = -(|\hat{\beta}_{ols\,i}| - \lambda)^+$$

$$\iff \beta_i = \text{sign}(\hat{\beta}_{ols\,i})(|\hat{\beta}_{ols\,i}| - \lambda)^+$$

3) a) Using L_2 loss, $L_2(y, \hat{F}(x)) = (y - \hat{F}(x))^2$

$$F = \underset{F}{\operatorname{argmin}} \left\{ L(1, F) \times p + L(-1, F)(1-p) \right\}$$

$$= \underset{F}{\operatorname{argmin}} \left\{ (1-F)^2 \times p + (-1-F)^2 (1-p) \right\}$$

taking the derivative with respect to F:

$$-2(1-F) \times p \quad -2(-1-F)(1-p) = 0$$

$$\Leftrightarrow (-2 + 2F) \times p + (2 + 2F) - p \times (2 + 2F) = 0$$

$$\Leftrightarrow p(-2 + 2F - (2 + 2F)) + 2 + 2F = 0$$

$$\Leftrightarrow -4p + 2 + 2F = 0$$

$$\Leftrightarrow F = 2p - 1$$

Replacing p by its estimate, given a training set

$$\hat{F} = 2\hat{p} - 1 = 2\left(\frac{1 + a y_i}{2}\right) - 1 = a y_i, \; y_i \in T$$

b) $R = \frac{1}{N} \sum_{i=1}^{N} E\left[(y_i^* - \hat{F}(x_i))^2\right]$

$$R = \frac{1}{N} \sum_{i=1}^{N} \left[E_{T,Y^*}\left[(y_i - F(x_i))^2\right] + \left(E_{T,Y^*}\left[\hat{F}(x_i)\right] - F(x_i)\right)^2 + E_{T,Y}\left[\left(\hat{F}(x_i) - E_{T,Y}\hat{F}(x_i)\right)^2\right] \right]$$

$$E_{T,Y^*}\left[(y_i - f(x_i))^2\right] = E_{T,Y^*}\left[y_i^2 - 2y_i(2p_i - 1) + (2p_i - 1)^2\right]$$

$$= E_{T,Y}\left[y_i^2\right] - 2(2p_i - 1)E[y_i] + (2p_i - 1)^2$$

$$= 0 \quad - 2(2p_i - 1)(2p_i - 1) + (2p_i - 1)^2$$

$$= -(2p_i - 1)^2$$

$$\left(E_{T,Y}\left[\hat{F}(x_i)\right] - F(x_i)\right)^2 = \left(E_{T,Y}\left[a\, y_i\right] - (2p_i - 1)\right)^2$$

$$= \left(a(2p_i - 1) - (2p_i - 1)\right)^2$$

$$= \left((2p_i - 1)(a - 1)\right)^2$$

$$E_{T,Y}\left[\left(\hat{F}(x_i) - E_{T,Y}\left[\hat{F}(x_i)\right]\right)^2\right] = -a^2(2p_i - 1)^2$$

$$\Rightarrow R = \frac{1}{M}\sum_{i=1}^{M}\left(-(2p_i - 1)^2 + (a-1)^2(2p_i - 1)^2 - a^2(2p_i - 1)^2\right)$$

$$R = \frac{1}{M}\sum_{i=1}^{M}\left(-2a(2p_i - 1)^2\right)$$