STAT 239 Homework 3

Tarek Tohme

May 25th, 2018

1 Theoretical Questions

Question 1 Since $X^T X$ is diagonal, it is a scaling matrix, therefore its eigenbasis is equal to its basis, hence

$$V^T = V = I \implies X^T X = D^2 \implies x_i^2 = d_i^2 \quad \forall i \in 1, ..., p$$

Also, $X^T X$ is full rank, therefore

$$r = s \implies U \in O(r) \implies \mu_j \mu_j^T = I$$

$$\begin{aligned}
Var(x) &= Var\left(\sum_{j=1}^{p} \mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T Y\right) \\
&= \sum_{j=1}^{p} Var\left(\mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T Y\right) \\
&= \sum_{j=1}^{p} \left(\frac{Y^T \mu_j \frac{d_j^2}{d_j^2+\lambda} \mu_j^T \mu_j \frac{d_j^2}{d_j^2+\lambda} \mu_j^T Y}{N}\right) \\
&= \sum_{j=1}^{p} \frac{d_j^4}{\left(d_j^2 + \lambda\right)^2} \frac{Y^T \mu_j \mu_j^T Y}{N} \\
&= \sum_{j=1}^{p} \frac{d_j^4}{\left(d_j^2 + \lambda\right)^2} \frac{Y^T Y}{N} \\
&= \sigma \sum_{j=1}^{p} \frac{d_j^4}{\left(d_j^2 + \lambda\right)^2}
\end{aligned}$$

$$\begin{aligned}
B(x) &= \mathbf{E}\left[\hat{F}(x)\right] - \hat{F}(x) \\
&= \mathbf{E}\left[X^T \hat{\beta}_{RR}\right] - X^T \hat{\beta_{RR}} \\
&= \mathbf{E}\left[\sum_{j=1}^{p} \mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T Y\right] - \sum_{j=1}^{p} x_i \beta_i \\
&= \sum_{j=1}^{p} \mathbf{E}\left[\mu_j \frac{d_j^2}{d_j^2 + \lambda} \mu_j^T Y\right] - \sum_{j=1}^{p} x_i \beta_i \\
&= \sum_{j=1}^{p} x_i \beta_i \left(\frac{d_j^2}{d_j^2 + \lambda} - 1\right).
\end{aligned}$$

since $\mathbf{E}[\epsilon] = 0, \mathbf{E}[Y] = F(x) = X^T\beta$

Question 2 Expressing the problem in Lagrangian form,

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

Ignoring β_0 and assuming orthogonality, we get

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

For M to be minimal, β_{Lasso} can't be of a different sign than $\hat{\beta}_{OLS}$ or else $\left\| \beta - \hat{\beta}_{OLS} \right\|_2$ would not be minimal. $(*)$
Fix $i \in 1, ..., p$
Case1 : $\beta_{OLSi} > 0$

$$\begin{aligned}
M' &= \frac{1}{2}\beta_i^2 - \beta_i \hat{\beta}_{OLSi} + \frac{1}{2}\hat{\beta}_{OLSi}^2 + \lambda\beta_i \qquad\qquad Where \quad M = \sum_{i=1}^{p} M' \\
&= \frac{1}{2}\beta_i^2 - \beta_i \left(\hat{\beta}_{OLSi} - \lambda \right) + \frac{1}{2}\hat{\beta}_{OLSi}^2 \\
\frac{\partial M'}{\partial \beta_i} &= \beta_i - \left(\hat{\beta}_{OLSi} - \lambda \right)
\end{aligned}$$

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

$$\begin{aligned}
\beta_i &= \left(\hat{\beta}_{OLSi} - \lambda \right)^+ \\
&= \left(|\hat{\beta}_{OLSi}| - \lambda \right)^+ \\
&= \text{sign}\left(\hat{\beta}_{OLSi} \right) \left(|\hat{\beta}_{OLSi}| - \lambda \right)^+
\end{aligned}$$

Since $\hat{\beta}_{OLSi} > 0$ and $\hat{\beta}_{OLSi} - \lambda$ must be positive for M' to be as small as possible.

Case2 : $\beta_{OLSi} < 0$

$$\begin{aligned}
M' &= \frac{1}{2}\beta_i^2 - \beta_i \left(\hat{\beta}_{OLSi} + \lambda \right) + \frac{1}{2}\hat{\beta}_{OLSi}^2 \\
\frac{\partial M'}{\partial \beta_i} &= 0 \implies \beta_i - \left(\hat{\beta}_{OLSi} + \lambda \right) = 0 \\
\beta_i &= \left(\hat{\beta}_{OLSi} + \lambda \right)^+ \\
&= \left(-|\hat{\beta}_{OLSi}| + \lambda \right)^+ \quad since \quad \hat{\beta}_{OLSi} < 0 \\
&= -\left(|\hat{\beta}_{OLSi}| - \lambda \right)^+ \\
&= \text{sign}\left(\hat{\beta}_{OLSi} \right) \left(|\hat{\beta}_{OLSi}| - \lambda \right)^+
\end{aligned}$$

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

$$\begin{aligned} F &= \arg\min_F \{L(1, F) * p + L(-1, F)(1-p)\} \\ &= \arg\min_F \left\{(1-F)^2 * p + (-1-F)^2(1-p)\right\} \end{aligned}$$

Setting the derivative with respect to F to zero,

$$\begin{aligned} -2(1-F) * p - 2(-1-F)(1-p) &= 0 \\ (-2+2F) * p + (2+2F) - p * (2+2F) &= 0 \\ p * (-2+2F-(2+2F)) + 2 + 2F &= 0 \\ -4p + 2 + 2F &= 0 \\ F &= 2p - 1 \end{aligned}$$

Replacing p by its estimate

$$\hat{F} = 2\hat{p} - 1 = 2 * \frac{1 + ay_i}{2} - 1 = ay_i$$

b)

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

$$\begin{aligned} \mathbf{E}_{T,Y^*}\left[(y_i - F(x_i))^2\right] &= \mathbf{E}_{T,Y^*}\left[y_i^2 - 2y_i(2p_i - 1) + (2p_i - 1)^2\right] \\ &= \mathbf{E}_{T,Y^*}\left[y_i^2\right] - 2(2p_i - 1)\mathbf{E}_{T,Y^*}\left[y_i\right] + (2p_i - 1)^2 \\ &= 0 - 2(2p_i - 1)(2p_i - 1) + (2p_i - 1)^2 \\ &= -(2p_i - 1)^2 \end{aligned}$$

$$\begin{aligned} \left(\mathbf{E}_{T,Y^*}\left[\hat{F}(x_i)\right] - F(x_i)\right)^2 &= \left(\mathbf{E}_{T,Y^*}\left[ay_i\right] - (2p_i - 1)\right)^2 \\ &= \left(a(2p_i - 1) - (2p_i - 1)\right)^2 \\ &= \left((2p_i - 1)(a - 1)\right)^2 \end{aligned}$$

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

$$\begin{aligned} R &= \frac{1}{n}\sum_{i=1}^{p}\left(-(2p_i - 1)^2 + \left((2p_i - 1)(a - 1)\right)^2 - a^2(2p_i - 1)^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{p}\left(-2a(2p_i - 1)^2\right) \end{aligned}$$