E6690 Stats Learning HW2

Yihang Ding, yd2459, yd2459@columbia.edu

P1

a) $$RR = RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$
$$= \sum_{i=1}^{2} \left(y_i - \sum_{j=1}^{2} \beta_j x_{i,j}\right)^2 + \lambda \sum_{j=1}^{2} \beta_j^2$$
$$= \left(y_1 - \sum_1^2 \beta_j x_{1j}\right)^2 + \left(y_2 - \sum_1^2 \beta_j x_{2j}\right)^2 + \lambda\beta_1^2 + \lambda\beta_2^2$$
$$= (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda\beta_1^2 + \lambda\beta_2^2$$

b) with
$$\begin{cases} \dfrac{\partial RR}{\partial \beta_1} = 2\beta_1 x_1^2 - 2(y_1 - \beta_2 x_{12}) + 2\beta_1 x_{21}^2 - 2(y_2 - \beta_2 x_{22}) + 2\lambda\beta_1 = 0 \\ \dfrac{\partial RR}{\partial \beta_2} = 2\beta_2 x_1^2 - 2(y_1 - \beta_1 x_{11}) + 2\beta_2 x_{22}^2 - 2(y_2 - \beta_1 x_{21}) + 2\lambda\beta_2 = 0. \\ x_{11} = x_{12} = x_1, x_{21} = x_{22} = x_2 \end{cases}$$

get
$$\begin{cases} \hat\beta_1 = \dfrac{x_1 y_1 + x_2 y_2 - \hat\beta_2(x_1^2 + x_2^2)}{\lambda + x_1^2 + x_2^2} \\ \hat\beta_2 = \dfrac{x_1 y_1 + x_2 y_2 - \hat\beta_1(x_1^2 + x_2^2)}{\lambda + x_1^2 + x_2^2} \end{cases} \quad \text{s.t. } \hat\beta_1 = \hat\beta_2$$

c). $$Lasso = \min\left(RSS + \lambda \sum_{j=1}^{p} |\beta_j|\right)$$
$$= \min\left((y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda|\beta_1| + \lambda|\beta_2|\right)$$

d). For Lasso Regression, $\min_\beta RSS$ subject to $\sum_{j=1}^{p} |\beta_j| = |\beta_1| + |\beta_2| \le s$.
$$(y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2$$
$$= (y_1 - \beta_1 x_{11} - \beta_2 x_{11})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{21})^2$$
$$= y_1^2 - 2y_1(\beta_1 + \beta_2)x_{11} + (\beta_1 + \beta_2)^2 x_{11}^2 + y_2^2 - 2y_2(\beta_1 + \beta_2)x_{21} + (\beta_1 + \beta_2)^2 x_{21}^2$$
$$= y_1^2 + y_2^2 + (\beta_1 + \beta_2)^2(x_{11}^2 + x_{21}^2)$$
$$= 2(y_1 - (\beta_1 + \beta_2)x_{11})^2.$$

The solution is $\beta_1 + \beta_2 = \frac{y_1}{x_{11}}$, which is parallel to edge of $\hat\beta_1 + \hat\beta_2 \le s$
So the edges on $\hat\beta_1 + \hat\beta_2 = s$ are solutions, the general form is,
$$\begin{cases} \hat\beta_1 \ge 0, \hat\beta_2 \ge 0 : \hat\beta_1 + \hat\beta_2 = s \\ \hat\beta_1 \le 0, \hat\beta_2 \le 0 : \hat\beta_1 + \hat\beta_2 = -s \end{cases}$$

P3

a) $L(\theta|\beta) = P(\beta|\theta) = \cancel{\prod_{i=1}^{n} P(\beta_i|\theta)} \; P(y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j) | \theta) = P(\epsilon_1, \epsilon_2, \ldots \epsilon_n | \theta)$

$$= \prod_{i=1}^{n} P(\epsilon|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \cancel{+\epsilon_i})|^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \cdot |y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i)|^2}$$

b). ~~P(θ|β) =~~

$f(\beta|X,Y) \propto P(Y|X,\beta)\, P(\beta)$

$$= L(\theta|\beta)\, P(\beta)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \cdot |y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \cancel{+\epsilon_i})|^2} \cdot \frac{1}{2b} e^{-|\beta|/b}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \frac{1}{2b} \cdot e^{-\frac{1}{2\sigma^2} \cdot |y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \cancel{+\epsilon_i})|^2 - \frac{|\beta|}{b}}$$

c). $\log(f(\beta|x,y)) = \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \frac{1}{2b} - \left(\frac{1}{2\sigma^2} \cdot |y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \cancel{+\epsilon_i})|^2 + \frac{|\beta|}{b}\right)$

To maximize posterior, s.t.

$\underset{\beta}{\operatorname{argmax}} (\log(f(\beta|x,y)))$

As the first part in $\log(f(\beta|x,y))$ does not contain any β,

it's equal to $\underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2\sigma^2} \cdot |y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij} \cancel{+\epsilon_i})|^2 + \frac{|\beta|}{b}\right)$

$$= \underset{\beta}{\operatorname{argmin}} \left(\underbrace{|y_i - (\beta_0 + \sum_{j=1}^{p} x_{ij})|^2}_{RSS} + \frac{2\sigma^2}{b}|\beta|\right).$$

so if let $\frac{2\sigma^2}{b} = \lambda$, it's equal to $\operatorname{argmin}(RSS + \lambda|\beta|)$.

which is Lasso Regression.

d) $f(\beta|X,y) \propto P(Y,X,\beta)\, P(\beta)$

$$= L(\theta|\beta)\, P(\beta)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \cdot |y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j|^2} \cdot \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi c}} e^{-\frac{\beta_i^2}{2c}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{n+p} \cdot \frac{1}{\sigma^n} \cdot \frac{1}{\sqrt{c^p}} \cdot e^{-\frac{1}{2\sigma^2}|y_i - \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j|^2 - \frac{1}{2c}\sum_{j=1}^{p} \beta_j^2}$$

e) Similar to C.

$\cancel{\text{to}}$ $\underset{\beta}{\mathrm{Argmax}}\,(f(\beta \mid X, y))$.

is $\underset{\beta}{\mathrm{Argmin}}\left(\frac{1}{2\sigma^2}\cdot\left|y_i-(\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j)\right|^2+\frac{1}{2C}\sum_{j=1}^{p}\beta_j^2\right)$

$= \mathrm{Argmin}\left(\underbrace{\left|y_i-(\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j)\right|^2}_{RSS.}+\frac{\cancel{2}\sigma^2}{C}\sum_{j=1}^{p}\beta_j^2\right)$

Let $\lambda = \frac{\sigma^2}{C}$, $= \mathrm{Argmin}\,(RSS + \lambda\sum_{j=1}^{p}\beta_j^2)$.

As $\beta: \beta_1 \cdots \beta_p$ are iid. and $\sim N(0, c)$, mode $= \frac{\sigma^2}{C}$, and it's the posterior mean.