6.439 Problem Set 2

Xudong Sun

2.3

(a)

Compared to $\hat{\beta}$, the new $\hat{\beta}_{l1}$ is likely to contain 0s. As the L1 regularization has the property of producing many coefficients with zero values or very small values with few large coefficients, it can be used for feature selection.

(b)

The new $\hat{\beta}_{l2}$ does not contain 0s so no feature will be excluded. But L2 regularization can penalize the loss function. Therefore, the values of β are different.

(c)

The quadratic penalty term of the elastic net penalty makes the loss function strictly convex, and it therefore has a unique minimum. Therefore, the elastic net penalty helps overcome the limitation of L1 regularization that the solution is not unique. Also, we get a better outcome when doing regression on high-dimensional & few examples, or corelated data.

(d)

Likelihood function for observations under the logistic regression model:

$$L(\beta_0, \beta) = \prod_{i=1}^{n} P(Y = y_i | X = x_i)$$

$$= \prod_{i=\{1,\dots,n\}, y_i=1} P(Y = 1 | X = x_i) \cdot \prod_{i=\{1,\dots,n\}, y_i=0} P(Y = 0 | X = x_i)$$

$$= \prod_{i=\{1,\dots,n\}, y_i=1} P(Y = 1 | X = x_i) \cdot \prod_{i=\{1,\dots,n\}, y_i=0} (1 - P(Y = 1 | X = x_i))$$

$$= \prod_{i=1}^{n} P(Y = 1 | X = x_i)^{y_i} \big(1 - P(Y = 1 | X = x_i)\big)^{1-y_i}$$

Log-likelihood function:

$$\mathcal{L}(\beta_0, \beta) = \log\Big(\prod_{i=1}^{n} P(Y = 1 | X = x_i)^{y_i} \big(1 - P(Y = 1 | X = x_i)\big)^{1-y_i}\Big)$$

$$= \sum_{i=1}^{n} [y_i \log\big(P(Y = 1 | X = x_i)\big) + (1 - y_i) \log\big(1 - P(Y = 1 | X = x_i)\big)]$$

$$P(Y = 1 | X = x_i) = \frac{1}{1 + \exp(-\beta_0 - \beta^T x_i)}$$

$$\mathcal{L}(\beta_0, \beta) = \sum_{i=1}^{n} \log(1 + \exp(\beta_0 + \beta^T x_i)) + \sum_{i=1}^{n} [y_i(\beta_0 + \beta^T x_i)]$$

(e)

Log-likelihood function for observations:

$$\mathcal{L} = \log\left(\prod_{i=1}^{n} P(X = x_i, Y = y_i)\right)$$

$$\mathcal{L} = \log\left(\prod_{i=\{1,\dots,n\},y_i=1} P(X = x_i|Y = 1)\, P(Y = 1) \cdot \prod_{i=\{1,\dots,n\},y_i=0} P(X = x_i|Y = 0)P(Y = 0)\right)$$

$$= \log\left(\prod_{i=\{1,\dots,n\},y_i=1} P(X = x_i|Y = 1)\, \eta \cdot \prod_{i=\{1,\dots,n\},y_i=0} P(X = x_i|Y = 0)(1-\eta)\right)$$

$$= \log\left(\prod_{i=\{1,\dots,n\}} [P(X = x_i|Y = 1)\eta]^{y_i}\, [P(X = x_i|Y = 0)(1-\eta)]^{1-y_i}\right)$$

$$= \sum_{i=1}^{N}[y_i \log(\eta) + y_i \log[P(X = x_i|Y = 1)] + (1-y_i)\log(1-\eta) + (1-y_i)\log[P(X = x_i|Y = 0)]]$$

$$P(X = x_i|Y = 1) = Norm.PDF(\mu_1, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}}\exp(-\frac{(x_i-\mu_1)^2}{2\Sigma})$$

$$P(X = x_i|Y = 0) = Norm.PDF(\mu_0, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}}\exp(-\frac{(x_i-\mu_0)^2}{2\Sigma})$$