Question 1.

# HDDA - Spring 2020

$$\underset{z}{\text{minimize}} \; \|z\|_1 \quad \text{subject to} \quad Ax - z = b.$$

a Decomposing (1)

$$\begin{cases} f = 0 \\ g = \|.\|_1 \end{cases} \leftrightarrow \begin{cases} f(x) = 0 \\ g(z) = \|z\|_1 \end{cases}$$

Augmented Lagrian:

$$\mathcal{L}(x, z, u, \rho) = f(x) + g(z) + u^T(Ax - z - b) + \frac{1}{2}\|Ax - z - b\|_2^2$$

b) ADMM update:

$$x^{k+1} := (A^TA)^{-1} A^T(b + z^k - u^k)$$

$$z^{k+1} := S_{1/\rho}(Ax^{k+1} - b + u^k)$$

$$u^{k+1} := u^k + Ax^{k+1} - z^{k+1} - b$$

Least Absolute error for Regression problem:

$y = \mathbb{R}^n \leftarrow$ Response

$X = \mathbb{R}^{n \times p} \leftarrow$ features

$z =$ error, meaured by $\beta X - y = z$

objective cost function

$$\underset{z}{\text{minimize}} \; \|z\|_1 \quad \text{subject to} \quad \beta X - y - z = 0$$

problem 2:

$$L(\beta) = \prod_{i=1}^{n} (1 - \pi_i)^{1-y_i} (\pi_i)^{y_i}$$

$$\log L(\beta) = \log\left( \prod_{i=1}^{n} (1 - \pi_i)^{1-y_i} (\pi_i)^{y_i} \right)$$

$$= \sum_{i=1}^{n} \log\left[ (1 - \pi_i)^{1-y_i} (\pi_i)^{y_i} \right]$$

$$= \sum_{i=1}^{n} \left[ (1-y_i)(1-\pi_i) + y_i(\pi_i) \right]$$

$$= \sum_{i=1}^{n} \left[ (1-y_i)\left(1 - \frac{\cancel{\exp}e^{(\beta_0 + \beta^T x)}}{1+\exp(\beta_0+\beta^T x)}\right) + y_i \frac{\exp(\beta}{1+\exp(\beta} \right.$$

$$\log L(\beta) = \sum_{i=1}^{n} \left[ (1-y_i) \frac{1}{1+\exp(\beta_0+\beta^T x_i)} + y_i \frac{\exp(\beta_0+\beta^T x}{1+\exp(\beta_0+\beta} \right.$$

$$\log L(\beta) = \sum_{i=1}^{n} \left( \frac{1-y_i + y_i(\exp(\beta_0+\beta^T x_i))}{1+\exp(\beta_0+\beta^T x_i)} \right)$$

objective function of Ridge LR:

$$\min_{\beta_0, \beta_1} + \log \mathcal{L}(\beta) + \lambda(\|\beta_0\|_1 + \|\beta_1\|)$$

$$= \min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( \frac{1 - y_i + y_i(\exp(\beta_0 + \beta_1^T x_i))}{1 + \exp(\beta_0 + \beta_1^T x_i)} \right) + \lambda(\|\beta_0\|_1 + \|\beta_1\|_1)$$

d) Lasso:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( \frac{1 - y_i + y_i(\exp(\beta_0 + \beta_1^T x_i)}{1 + \exp(\beta_0 + \beta_1^T x_i)} \right) + \lambda(\|\beta_0\|_2^2 + \|\beta_1\|_2^2)$$

e. Adaptive Lasso

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( \frac{1 - y_i + y_i(\exp(\beta_0 + \beta_1^T x_i))}{1 + \exp(\beta_0 + \beta_1^T x_i)} \right) + \lambda_n \sum_{j=1}^{n} \hat{w}_j |\beta_j|$$

Where $\hat{w}_j = \frac{1}{|\hat{\beta}_j^{ridge}|}$ etc.

f) I will choose adaptive Lasso based on confusion matrix and accuracy. $(0 \to 277 / 1 \to 23)$

① We have severely unbalanced data set, thus, Ridge Regression tends to shrink every coefficients to 0 to predict all 0 labels, while Adaptive Laso can retain non-0 coefficients problem 4:

$$E(U,V) = \sum_{(u,i) \in M} \left( M_{u,i} - U_u^T V_i \right)^2$$

$$E(U,V) = \sum_{(u,i) \in M} \left( M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right)^2$$

a) updating for $U_{v,k}$

$$U_{v,k} \leftarrow U_{v,k} - \mu \frac{\partial E(U,V)}{\partial U_{v,k}}$$

$$\Leftrightarrow U_{v,k} \leftarrow U_{v,k} + 2\mu \left( M_{u,i} - \sum_{j=1}^{r} U_{u,j} V_{i,j} \right) \left( V_{i,k} \right)$$

b. Updating for $V_{i,k}$

$$V_{i,k} \leftarrow V_{i,k} - \mu \frac{\partial E(U,V)}{\partial V_{i,k}}$$

$$\Leftrightarrow V_{i,k} \leftarrow V_{i,k} + 2\mu \underset{(u,i) \in M}{\left( M_{u,i} - \sum_{j=1}^{r} U_{u,j} V_{i,j} \right)} \left( U_{u,k} \right)$$

The generalized [illegible]

$$E(u,v) = \sum_{(u,i)\in M} \left( M_{u,i} - \sum_{k=1}^{r} u_{u,k} V_{i,k} \right)^2 + \lambda \sum_{u,k} u_{u,k}^2 + \lambda \sum_{i,k}$$

a) updating $u_{u,k}$ :

$$u_{u,k} = u_{u,k} - \mu \frac{E(u,v)}{\partial u_{u,k}}$$

$$u_{u,k} = u_{u,k} - \mu \Big[ 2 \Big( M_{u,i} - \sum_{j=1}^{r} u_{u,j} V_{i,j} \Big) (- V_{i,k}) + (2(\lambda)(u_{u,k}) \Big]$$

b. updating $V_{u,k}$

$$V_{i,k} = V_{i,k} - \mu \Big[ 2 \Big( M_{u,i} - \sum_{j=1}^{r} u_{u,j} V_{i,j} \Big) (- u_{u,k}) + (2(\lambda)(V_{i,k})) \Big]$$