

1 ALO for Logistic Regression

1.1 ALO for Logistic Regression with Lasso penalty

First Let's rewrite the optimization problem with the loss functions separated for each observation, therefore the loss function goes:

$$-\sum_i (y_i x_i^\tau \beta + \log(1 + \exp(x_i^\tau \beta))) + \lambda_1 \|\beta\|_1$$

Where, separatively, the loss function is $l(x_i^\tau \beta; y_i) = y_i x_i^\tau \beta + \log(1 + \exp(x_i^\tau \beta))$ and the regularizer is $R(\beta) = \lambda_1 \|\beta\|_1$, from which we could derive the dual optimal $\hat{\theta} = y - \frac{e^{x\beta}}{1+e^{x\beta}}$, as well as the conjugate functions of them as $l^*(-\theta_i; y_i) = (y_i - \theta_i) \ln \frac{y_i - \theta_i}{1 - (y_i - \theta_i)} - \ln \frac{1}{1 - (y_i - \theta_i)}$,

$$R^*(\beta) = \begin{cases} 0 & \|\beta\|_\infty \leq \lambda_1 \\ \infty & o.w. \end{cases}$$

From the results of the conjugate functions above, we could also obtain the derivatives of the loss functions and the Jacobian of the regularizer:

$$\begin{aligned} \dot{l}^*(-\theta_i; y_i) &= \ln \frac{y_i - \theta_i}{1 - (y_i - \theta_i)} \\ \ddot{l}^*(-\theta_i; y_i) &= \frac{1}{(y_i - \theta_i)(1 - (y_i - \theta_i))} \end{aligned}$$

Recall (15) from the **General Smooth Loss**, the quadratic surrogate of the dual problem is $\min_u \frac{1}{2} \sum_i (u_i - \frac{\hat{\theta}_i \ddot{l}^*(-\hat{\theta}_i; y_i) + \hat{y}_i}{\sqrt{\ddot{l}^*(-\hat{\theta}_i; y_i)}})^2 + R^*(X^\tau K u)$, where $K = \text{diag}(\sqrt{\ddot{l}^*(-\hat{\theta}_i; y_i)})$ therefore the Jacobian at $y_u = \frac{\hat{\theta}_i \ddot{l}^*(-\hat{\theta}_i; y_i) + \hat{\theta}_i}{\sqrt{\ddot{l}^*(-\hat{\theta}_i; y_i)}}$ could locally be treated as the projection onto the orthogonal complement of the polyhedra $\{\|X^\tau K u\|_\infty \leq \lambda_1\}$, thus $J = I - X_{u,E} (X_{u,E}^\tau X_{u,E})^{-1} X_{u,E}$, where $X_{u,E}$ are the columns of $X_u = X^\tau K$, such that the columns in the set $E = \{|X_i^\tau \theta| = \lambda_1\}$ are selected. Take everything to (17), $y^i = K_{ii}(y_{u,i} - \frac{K_{ii} \hat{\theta}_i}{J_{ii}})$, we could obtain the alo for the i th observation.

1.2 ALO for Logistic Regression with Elastic Net Penalty

The optimization problem for logistic regression with elastic net penalty is:

$$-\sum_i (y_i x_i^\tau \beta + \log(1 + \exp(x_i^\tau \beta))) + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

The optimization problem is the same except the regularizer is changed, therefore the only thing different is the conjugate function of the regularizer, R^* and the corresponding Jacobian, here $R(\beta) = \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$:

$$R^*(\beta) = \sum_{|u_i| > \lambda_1} \frac{(\lambda_1 - |u_i|)^2}{4\lambda_2}$$

The corresponding Jacobian is $J = (I + \frac{1}{2\lambda_2} X_{u,E} X_{u,E}^\tau)$, where $X_{u,E}$ are the columns of $X_u = X^\tau K$, such that the columns in the set $E = \{|X_i^\tau \theta| = \lambda_1\}$ are selected. Take everything to (17), $y^{/i} = K_{ii}(y_{u,i} - \frac{K_{ii}\hat{\theta}_i}{J_{ii}})$, we could obtain the alo for the i th observation.

2 ALO for Poisson Regression

2.1 ALO for Poisson Regression with Lasso Penalty

The optimization function for Poisson regression with lasso penalty is:

$$\sum_i -y_i x_i^\tau \beta + e^{x_i^\tau \beta} + \log(y_i!) + \lambda_1 \|\beta\|_1$$

The regularizer is the same as the logistic regression with the lasso penalty case, thus the Jacobian will also be the same, therefore we only have to focus on the loss function $l(x_i^\tau \beta; y_i) = -y_i x_i^\tau \beta + e^{x_i^\tau \beta} + \log(y_i!)$. The optimal solution for the dual problem $\hat{\theta} = y - e^{X\beta}$ and the conjugate of the loss function is $l^*(-\theta_i; y_i) = (y_i - \theta_i) \ln(y_i - \theta_i) - (y_i - \theta_i)$, the corresponding derivatives are therefore:

$$\begin{aligned} \dot{l}^*(-\theta_i; y_i) &= \ln(y_i - \theta_i) \\ \ddot{l}^*(-\theta_i; y_i) &= \frac{1}{y_i - \theta_i} \end{aligned}$$

By plugging everything into (17), we obtain the alo for Poisson regression with the lasso penalty.

2.2 ALO for Poisson Regression with Elastic Net Penalty

The loss function for Poisson regression with elastic net penalty is the same as that of Poisson regression with the lasso penalty and the regularizer of it is the same as that of logistic regression with elastic net penalty, thus by plugging everything into (17), we could obtain the alo for Poisson regression with elastic net penalty.