

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.

1.1.1 ALO for Logistic Regression with Lasso Penalty with intercept included

If we include intercept for the model but it will not be penalized, and rewrite $\tilde{x}_i = [1; x_i]$, $\tilde{X} = [\mathbf{1}, X]$ and $\tilde{\beta} = [\beta_0, \beta]$, the loss function stays in the same form as the case with no intercept, that is $l(\tilde{x}_i^\tau \tilde{\beta}; y_i) = -(y_i \tilde{x}_i^\tau \tilde{\beta} + \log(1 + \exp(\tilde{x}_i^\tau \tilde{\beta})))$, and the but the regularizer will be different, it becomes $R(\tilde{\beta}) = \lambda_1 \|D\tilde{\beta}\|_1$, where $D = [\mathbf{0}, I]$. Therefore, the corresponding dual optimal becomes $\hat{\theta} = y - \frac{e^{\tilde{X}\tilde{\beta}}}{1 + e^{\tilde{X}\tilde{\beta}}}$, the conjugate of the loss function will stay the same, but the conjugate for $R(\tilde{\beta})$ shall be changed.

$$R^*(\tilde{\beta}) = \begin{cases} 0 & \text{if } \beta_0 = 0 \text{ and } \beta_i \leq \lambda_1 \\ +\infty & \text{o.w.} \end{cases}$$

Given the new regularizer and loss function, we could derive the new $\tilde{X}_u = K^{-1}\tilde{X}$, and the corresponding Jacobian is therefore $J = I - \tilde{X}_{u,\tilde{E}}(\tilde{X}_{u,\tilde{E}}^\tau \tilde{X}_{u,\tilde{E}})^{-1} \tilde{X}_{u,\tilde{E}}^\tau$, where \tilde{E} denotes the columns of \tilde{X}_u such that $\tilde{E} = \{j : \tilde{X}_j^\tau \theta \leq \lambda_1\}$.

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 $X_u = K^{-1}X$, 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.

1.2.1 ALO for Logistic Regression with Elastic Net Penalty with intercept

The loss function for logistic regression with elastic net penalty with intercept is the same as the case without an intercept, however, the regularizer function is different. The dual optimal will also be $\hat{\theta} = y - \frac{e^{X\beta}}{1+e^{X\beta}}$. Using the same notations D and $\tilde{\beta}$ in the logistic regression with lasso penalty with intercept part, the regularizer now becomes $\lambda_1 \|D\beta\|_1 + \lambda_2 \|D\beta\|_2^2$, the conjugate of which is therefore:

$$R^*(\beta) = \begin{cases} \sum_{i:|\beta_i|>\lambda_1} \frac{(\lambda_1 - |\beta_i|)^2}{4\lambda_2} & \text{if } \beta_0 = 0 \\ +\infty & \text{o.w.} \end{cases}$$

First denote the set $\{E : |X_j^\tau \theta| > \lambda_1\}$ and $A = \frac{1}{2\lambda_2} \tilde{X}_{u,E} \tilde{X}_{u,E}^\tau$, the new Jacobian becomes:

$$J = (I + A)^{-1} - \frac{(I+A)^{-1} \mathbf{1} \mathbf{1}^\tau (I+A)^{-1}}{\mathbf{1}^\tau (I+A)^{-1} \mathbf{1}}$$

By plugging everything into (17), we could obtain the alo.

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.

2.1.1 ALO for Poisson Regression with Lasso Penalty with Intercept

Using the same notations \tilde{X} and $\tilde{\beta}$, the loss function now becomes $l(x_i^\tau \tilde{\beta}; y_i) = -y_i \tilde{x}_i^\tau \tilde{\beta} + e^{\tilde{x}_i^\tau \tilde{\beta}} + \log(y_i!)$, and the dual optimal now becomes $\hat{\theta} = y - e^{\tilde{X}\tilde{\beta}}$. The regularizer is also the same as the logistic regression with lasso penalty with intercept case, therefore the Jacobian is of the same formula, thus by plugging everything into (17), we could obtain the alo.

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.

2.2.1 ALO for Poisson Regression with Elastic Net Penalty with Intercept

Here the loss function is the same as that of Poisson Regression with Lasso penalty with intercept case and the regularizer is the same as Logistic Regression with elastic net penalty with intercept, thus the formula for the derivatives

of loss functions, the dual optimal as well as the Jacobian could be directly obtained using the same formula as mentioned before.