Machine Learning Network-Constrained Regression of Epigenetic Data

Sivo Vladimirov Daskalov

Corpus Christi College

28 June 2017

Outline

Epigenetic background

Penalized regression methods

$$\text{Lasso } \lambda \sum_{i=1}^{p} |\beta_i|$$

$$\text{Elastic Net } \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sqrt{\sum_{i=1}^{p} \beta_i^2}$$

$$\text{Grace } \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{u \sim v} \left(\frac{\beta_u}{\sqrt{d_u}} - \frac{\beta_v}{\sqrt{d_v}} \right)^2 w(u,v)$$

$$\text{aGrace } \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{u \sim v} \left(\frac{sign(\tilde{\beta}_u)\beta_u}{\sqrt{d_u}} - \frac{sign(\tilde{\beta}_v)\beta_v}{\sqrt{d_v}} \right)^2 w(u,v)$$

$$\text{GBLasso } \lambda \sum_{u \sim v} \left[\left(\frac{|\beta_u|}{\sqrt{d_u}} \right)^\gamma + \left(\frac{|\beta_v|}{\sqrt{d_v}} \right)^\gamma \right]^{1/\gamma}$$

$$\text{Linf } \lambda \sum_{u \sim v} \max \left(\frac{|\beta_u|}{\sqrt{d_u}}, \frac{|\beta_v|}{\sqrt{d_v}} \right)$$

$$\text{aLinf } \lambda \sum_{u \sim v} \left| \frac{sign(\tilde{\beta}_u)\beta_u}{\sqrt{d_u}} - \frac{sign(\tilde{\beta}_v)\beta_v}{\sqrt{d_v}} \right|$$

$$\text{TTLP } \lambda_1 \sum_{i=1}^{p} J_\tau |\beta_i| + \lambda_2 \sum_{u \sim v} \left| J_\tau \left(\frac{|\beta_u|}{w_u} \right) - J_\tau \left(\frac{|\beta_v|}{w_v} \right) \right|$$

$$\text{LTLP } \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{u \sim v} \left| J_\tau \left(\frac{|\beta_u|}{w_u} \right) - J_\tau \left(\frac{|\beta_v|}{w_v} \right) \right|$$

Composite voting regression

Orchestrated hyperparameter tuning

Model evaluation

Regression method similarities

Breast cancer dataset