

Machine Learning Network-Constrained Regression of Epigenetic Data

Sivo Vladimirov Daskalov

Corpus Christi College

28 June 2017

Outline

Epigenetic background

Penalized regression methods

Composite voting regression

Orchestrated hyperparameter tuning

Model evaluation

Regression method similarities

Breast cancer dataset

Epigenetic background

Penalized regression methods

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

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

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

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

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

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

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

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

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

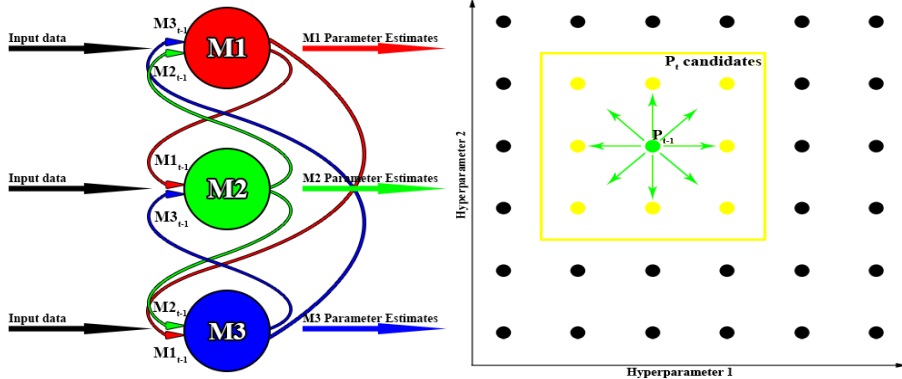
Composite voting regression

	X_1	X_2	...	X_p
<i>Method 1</i>	$M_1(\beta_1)$	$M_1(\beta_2)$...	$M_1(\beta_p)$
<i>Method 2</i>	$M_2(\beta_1)$	$M_2(\beta_2)$...	$M_2(\beta_p)$
...
<i>Method k</i>	$M_k(\beta_1)$	$M_k(\beta_2)$...	$M_k(\beta_p)$

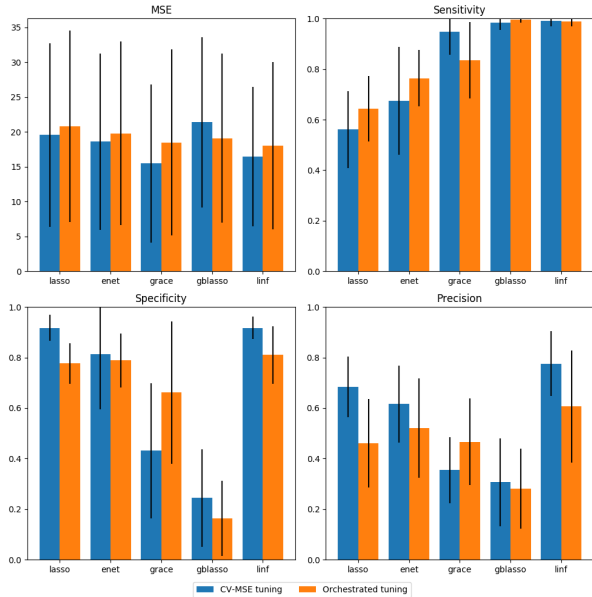
$$X_j = \begin{cases} \text{important,} & \text{if } \frac{\sum_{i=1}^k [M_i(\beta_j) \neq 0]}{k} \geq \text{fraction of votes threshold} \\ \text{unrelated,} & \text{otherwise} \end{cases}$$

Final model obtained from OLSE on the set of important predictors

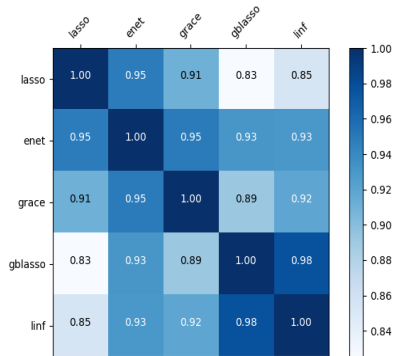
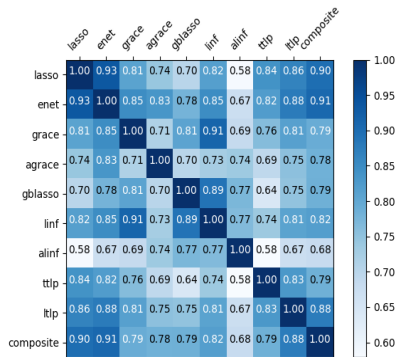
Orchestrated hyperparameter tuning



Model evaluation



Regression method similarities



Breast cancer dataset