

Модель LASSO

В методе LASSO коэффициенты линейной модели

$$y = X\beta + \varepsilon$$

находятся из решения задачи минимизации среднеквадратичной ошибки. При этом в функцию потерь добавляется штраф на величину коэффициентов:

$$loss(\beta, \lambda) = \|y - X\beta\|_2^2 + \lambda \sum_{i=1}^k |\beta_i|,$$

где λ — параметр регуляризации, который задаёт баланс между качеством подгонки модели и её сложностью. Для того, чтобы получить аналитическое решение этой задачи, необходимо для каждого коэффициента минимизировать функцию потерь для двух случаев раскрытия модуля.

Особенностью регрессии LASSO является её склонность занулять большие коэффициенты, из-за чего она часто используется для отбора переменных. Эта склонность определяется значением параметра регуляризации λ , который подбирается с помощью кроссвалидации.