

Linear regression - key messages

- The linear regression (LR) model is linear w.r.t. the parameters $\beta_0, \beta_1, \dots, \beta_n$ and β_0 is called the intercept/bias of the model.
- To make LR more robust to noise, Ridge regression appends the loss function by the L_2 loss of β (excluding the intercept) whose strength depends on the penalty term λ . The penalty term λ has to be estimated from the training data, e.g., by cross-validation.
- However, the L_2 loss is dominated by outliers and as an alternative LASSO regression uses the L_1 loss. As there is no closed form solution for LASSO, gradient-descent methods (using the smooth Huber function instead of L_1) are used in practice.