Regularization and trade-off between bias and variance

Farnoosh Khodakarami

This material is prepared with
Farnoosh khodakarami
and Ali Madani

Overfitting

Overfitting: Good performance on the training data, poor generalization to other data.

Underfitting: Poor performance on the training data and poor generalization to other data

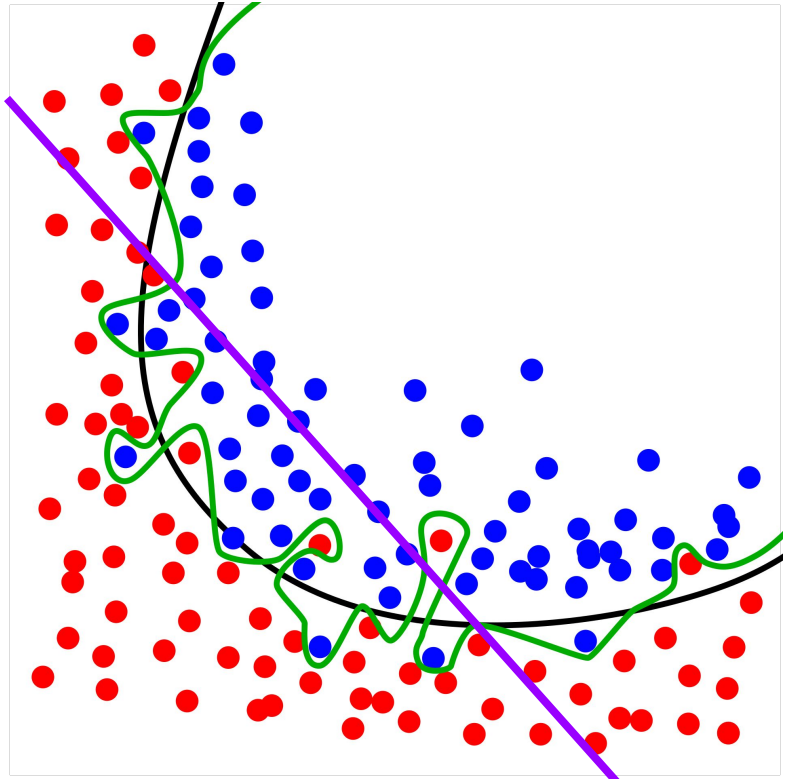
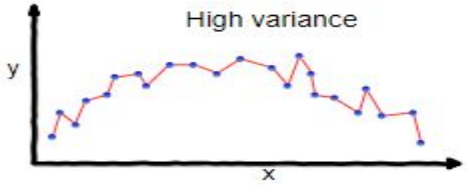
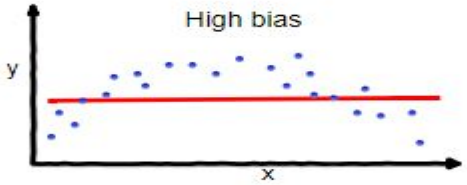


Image source wikipedia

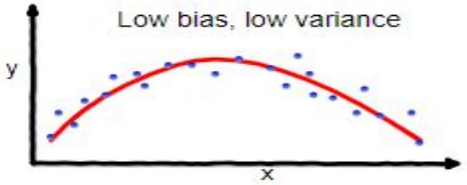
Bias–Variance Tradeoff



overfitting

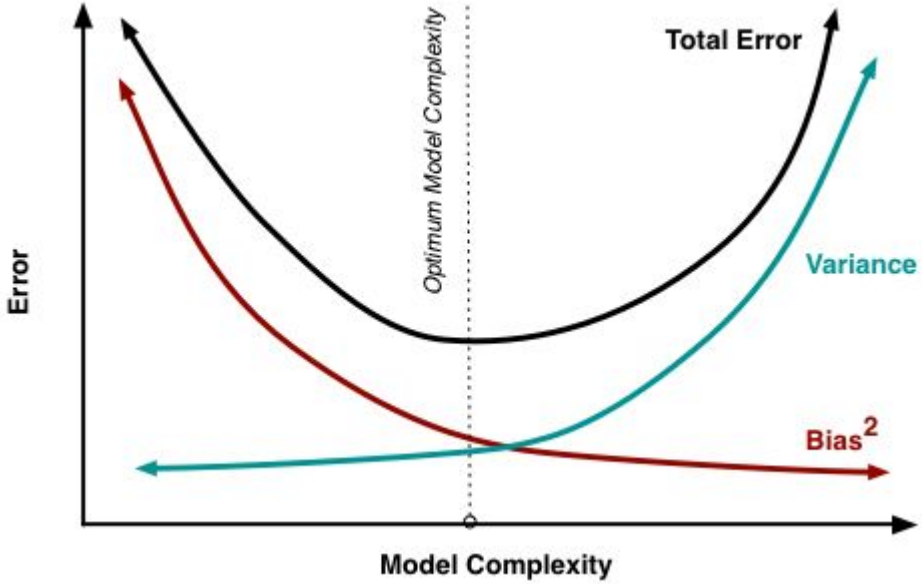


underfitting



Good balance

Bias–Variance Tradeoff



Training, validation and test set

Data	Use	Size
Training	Model training (parameter optimization)	Big
Validation	Assessing variance and hyperparameter optimization	Big or small Smaller than training
Test	Assessing variance	Smaller than validation

Training, validation and test set

Data	Use	Size
Training	Model training (parameter optimization)	Big
Validation	Assessing variance and hyperparameter optimization	Big or small Smaller than training
Test	Assessing variance	Smaller than validation
New data	I am the goal 😎	Very small

K-Folds Cross Validation to assess variance

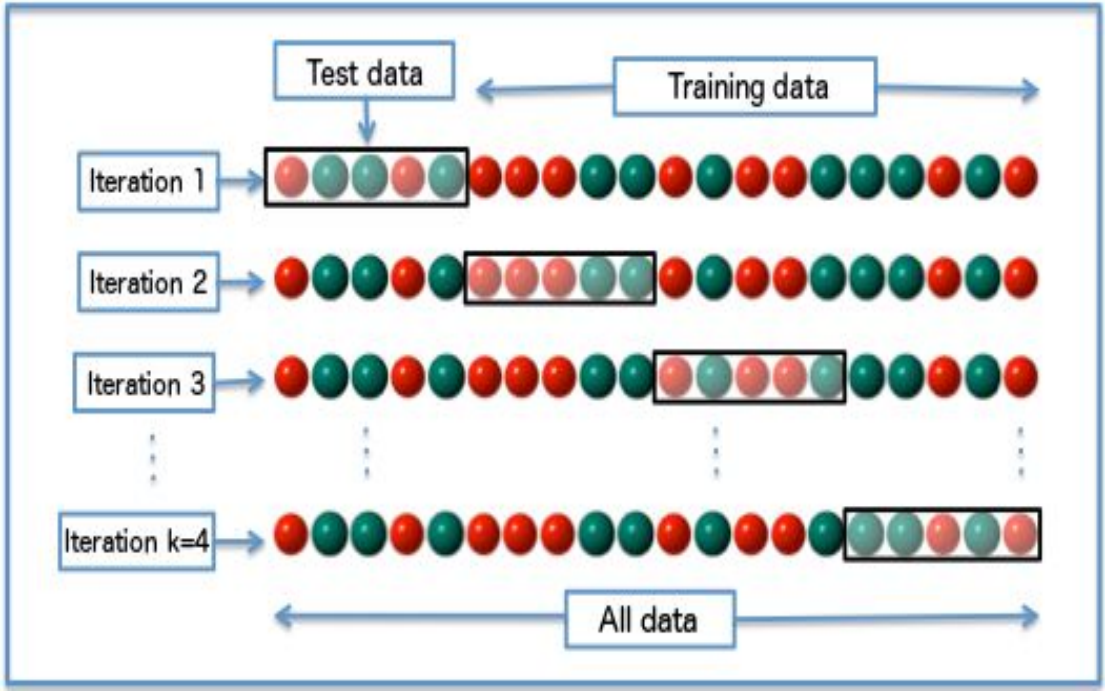
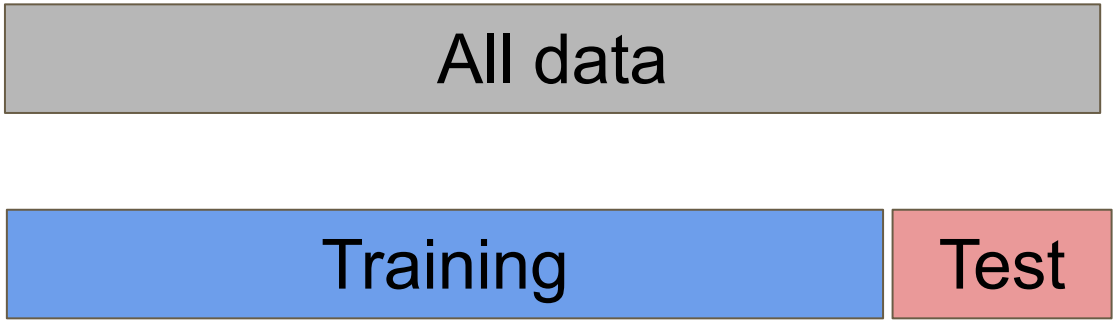


Image source wikipedia

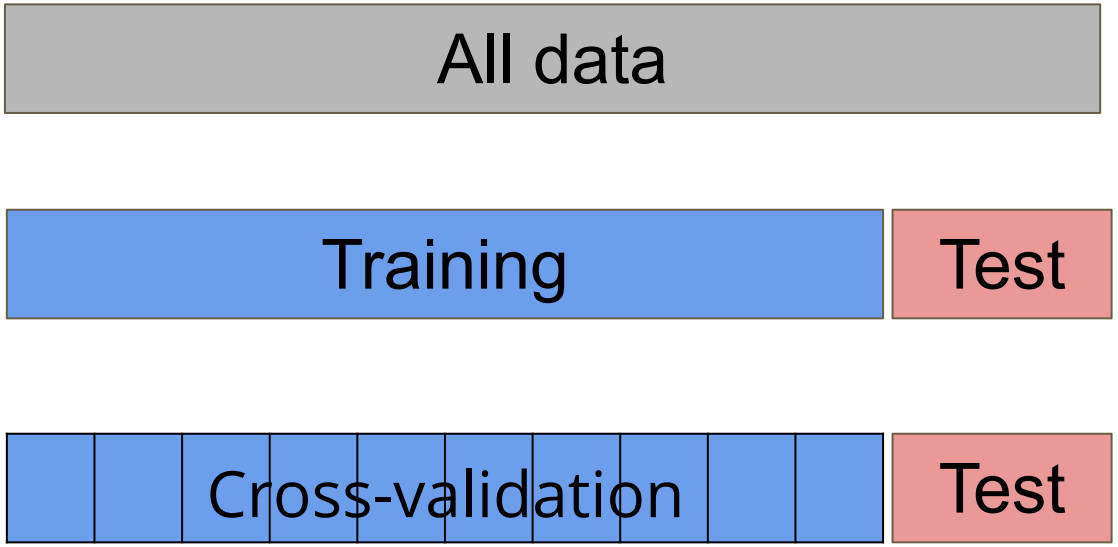
Training, validation and test splits (one dataset)

Training, validation and test splits (one dataset)

Holding out part of the data for assessing the model



Training, validation and test splits (one dataset)



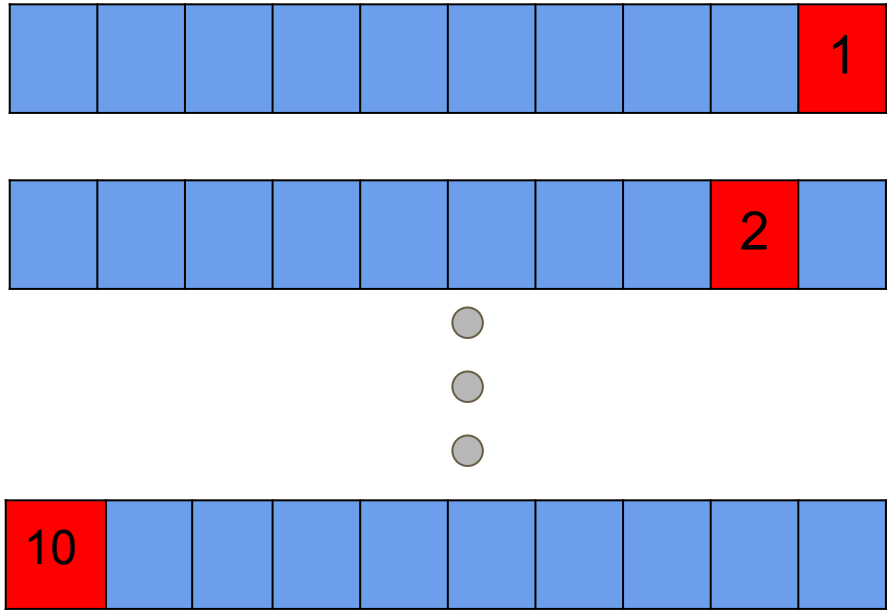
Holding out part of the data for assessing the model

Full:

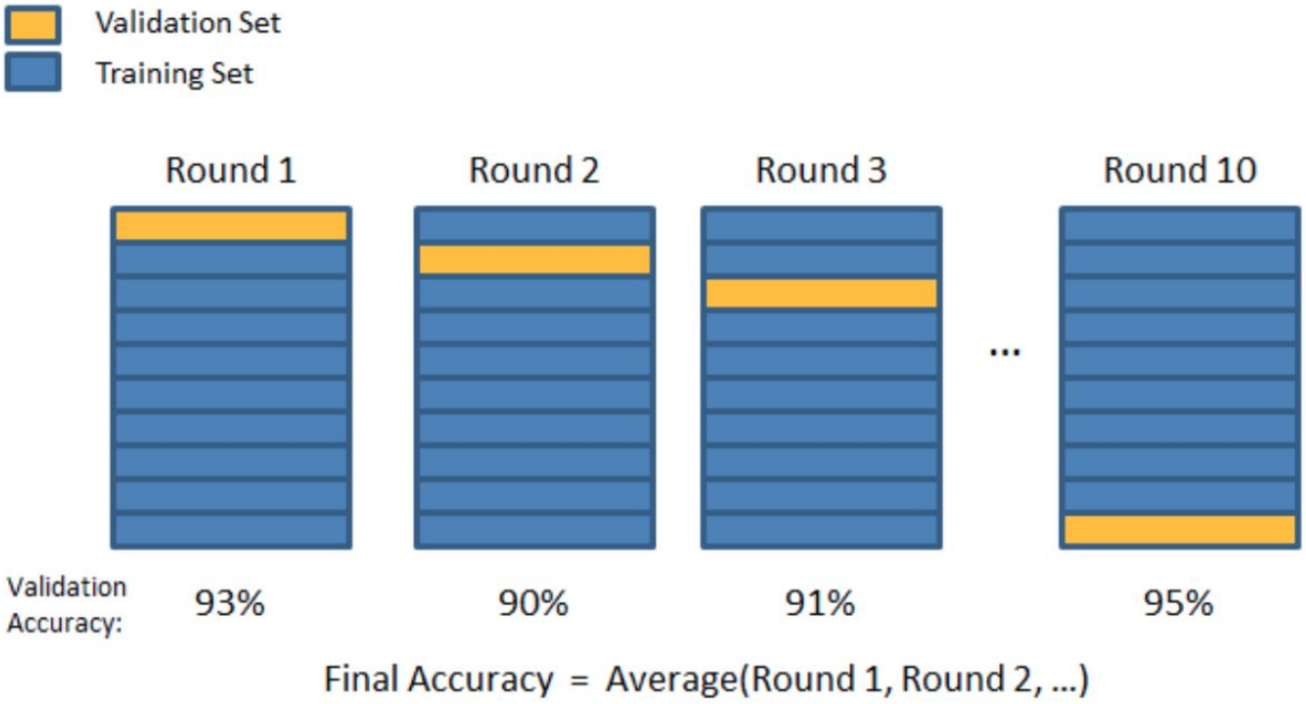
- Model training
- Variance assessment and hyperparameter optimization
- Testing the model

Accuracy in validation

10-fold cross-validation



Accuracy in validation



Regularization to overcome overfitting

Ridge regression

Objective function:

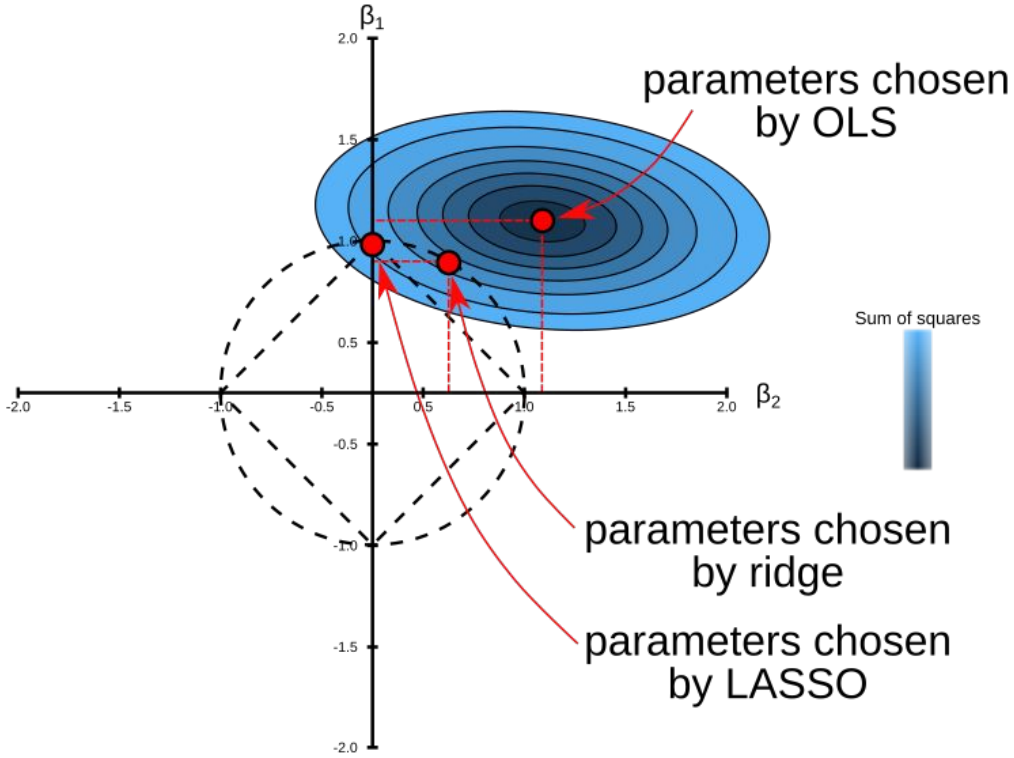
$$\underset{\beta_0,\beta}{\text{minimize}} \left\{ \frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\}$$

Constraint for regularization:

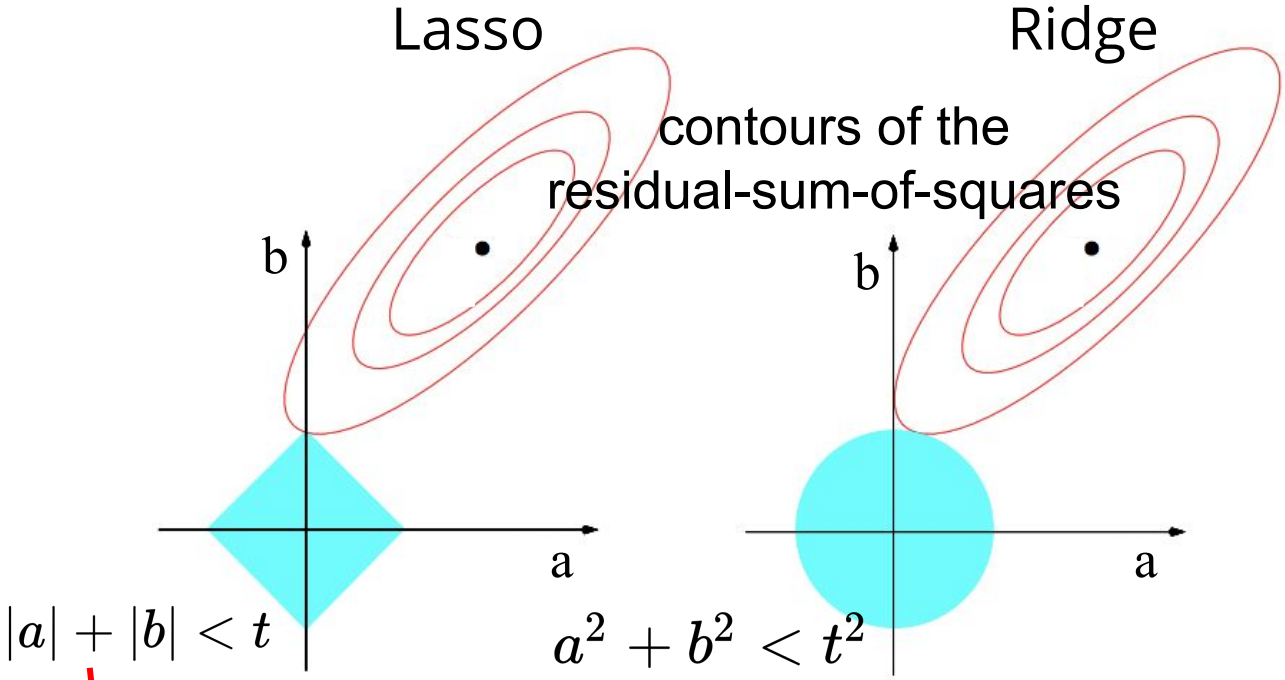
$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 \leq t^2.$$

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC

Optimization space in regularization

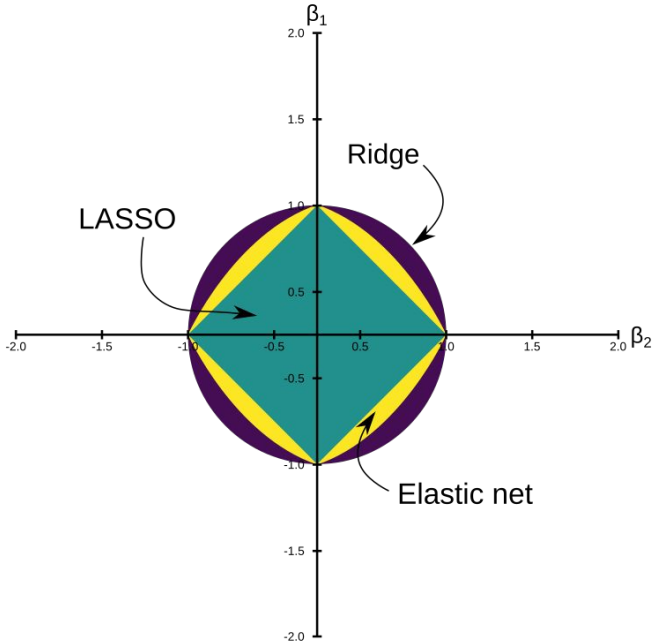


Optimization space in regularization



Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC

Linear Regression/Ridge Regression/ Lasso/Elastic Net



An image visualising how ordinary regression compares to the Lasso, the Ridge and the Elastic Net Regressors. Image Citation: Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net.