11 | Regularization

Ivan Corneillet

Data Scientist

GENERAL ASSEMBLY

Learning Objectives

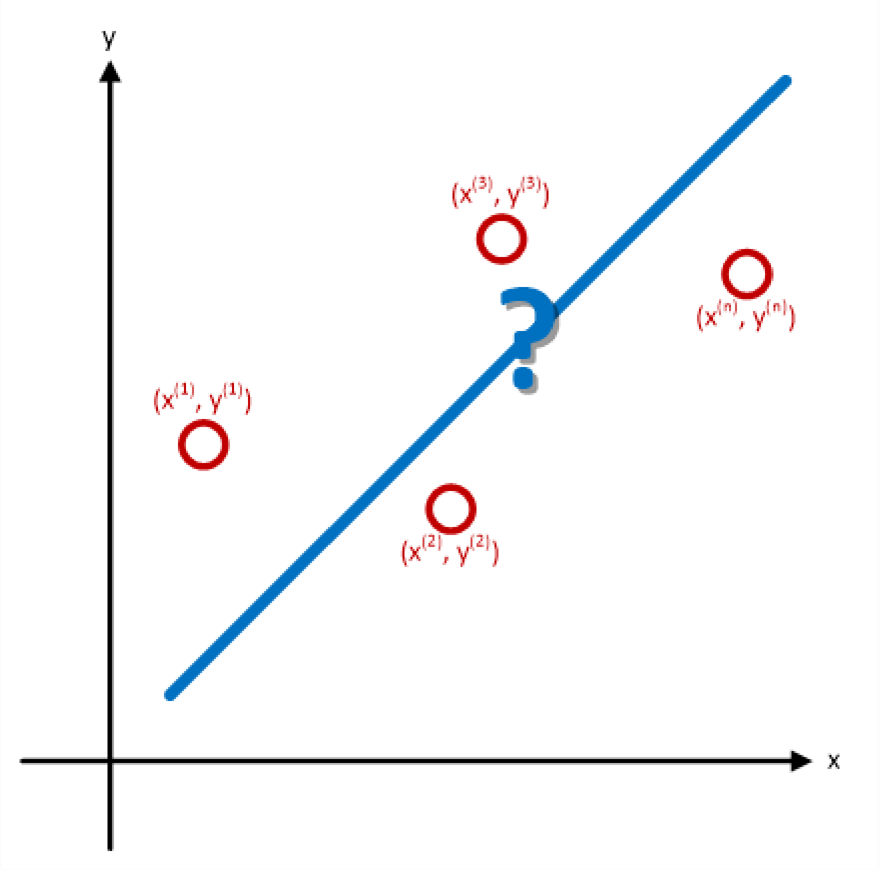
After this lesson, you should be able to:

- Use Ordinary Least Squares (OLS) and Loss Functions to also derive estimates for the coefficients
- Understand the Regularization Bias-Variance Trade-Off
- Understand LASSO and Ridge as a way to Regularize Linear Regression
- Use LASSO as a mean to select features

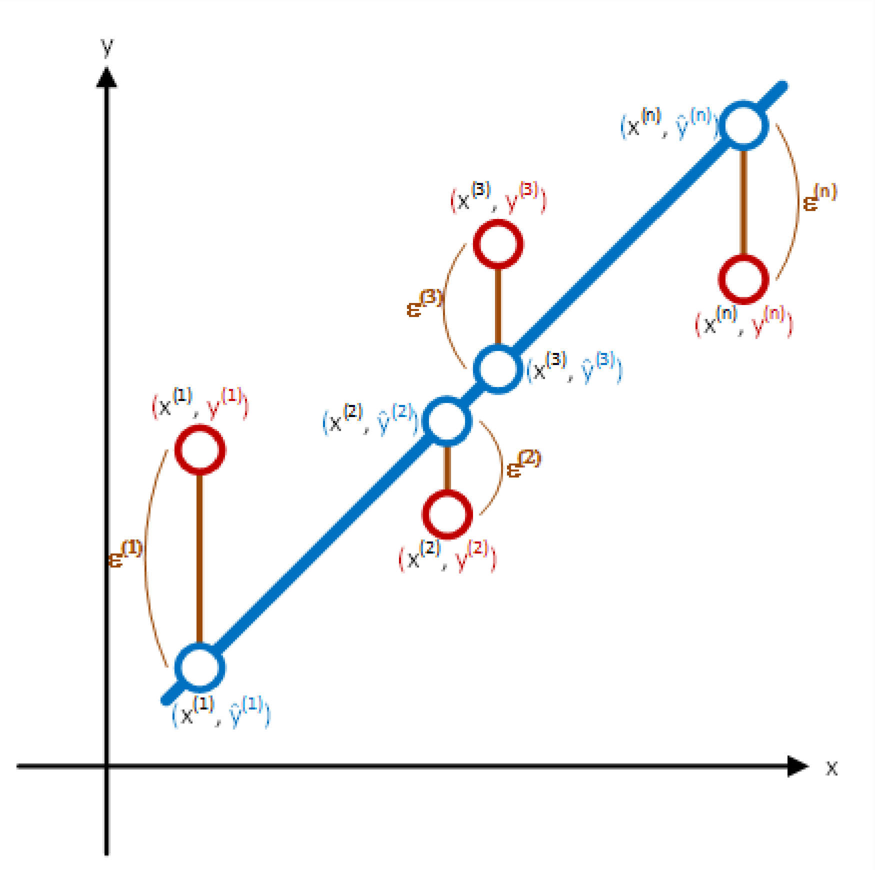
How to fit a linear regression model on a dataset?

Ordinary Least Squares (OLS) and Loss Functions

How do we estimate $\hat{\beta}$?



We can estimate $\hat{\beta}$ with Ordinary Least Squares (OLS)



- Hypothesis

$$\hat{y}(x) = x \cdot \hat{\beta}$$

- Parameters

$$\hat{\beta}$$

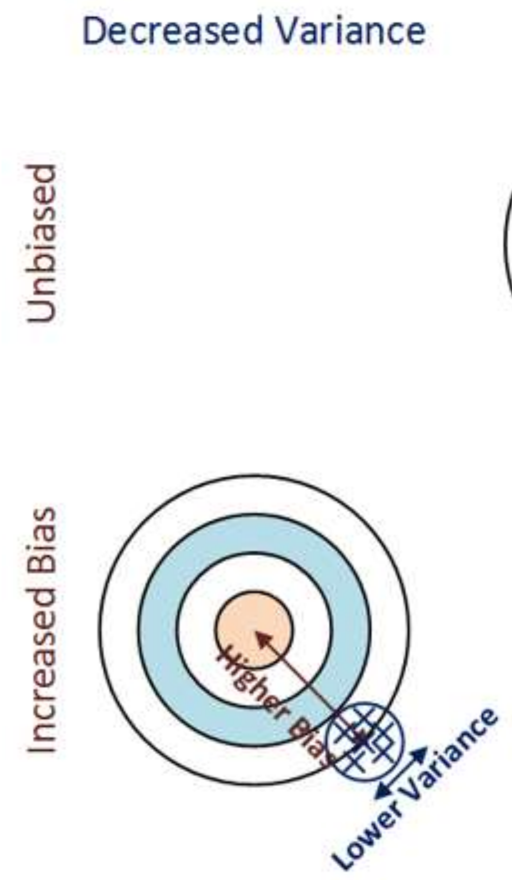
- Goal

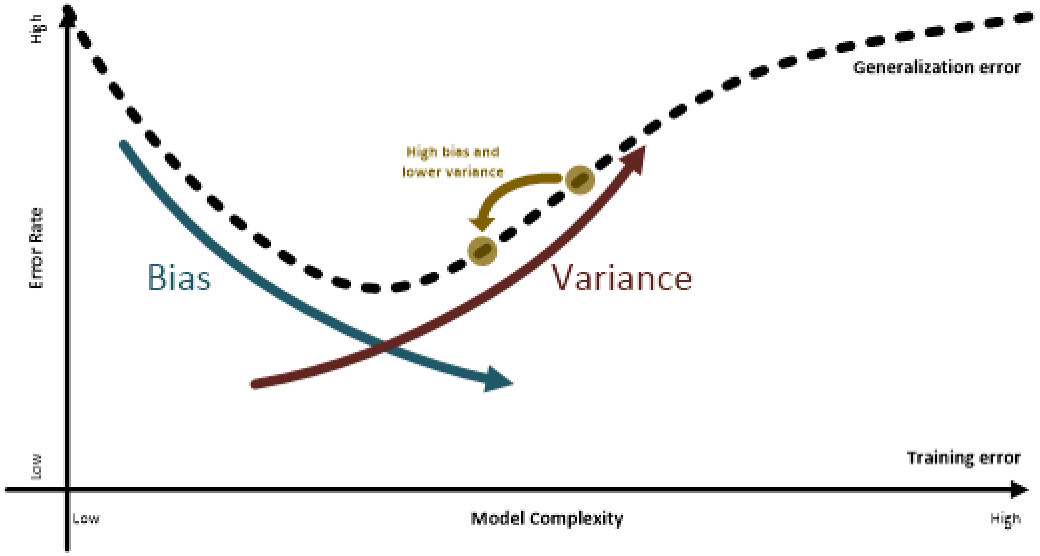
$$\underset{\hat{\beta}}{\operatorname{argmin}} \underbrace{\left\| y_{train} - \overbrace{X_{train} \cdot \hat{\beta}}^{\hat{y}_{train}} \right\|^2}_{L(\hat{\beta})}$$

(i.e., minimizing the least square errors)

Regularization

OLS yields unbiased estimators at the cost of high variance. Can we trade some (higher) bias for lower variance and get ahead on the bias-variance trade-off?





About Loss Functions

- Loss functions are a powerful tool to optimize the fit of machine learning algorithms
- Loss functions are not limited to linear regression- and regularization-based models
  - E.g., training a logistic regression algorithm (while also leveraging linear regression) is also modeled and fitted with loss functions