Regularization & Feature Selection

Data Science Immersive

FLATIRON SCHOOL

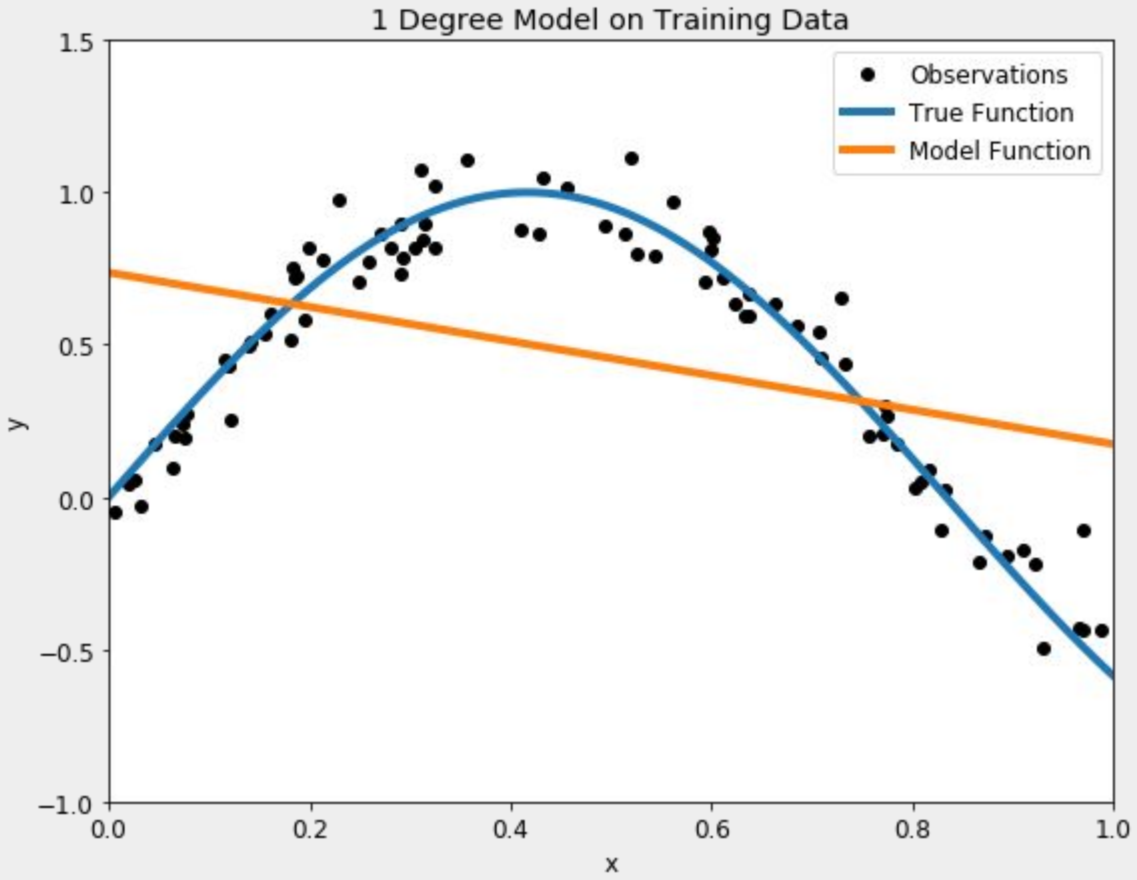
Learning Goals

After today, you'll be able to:

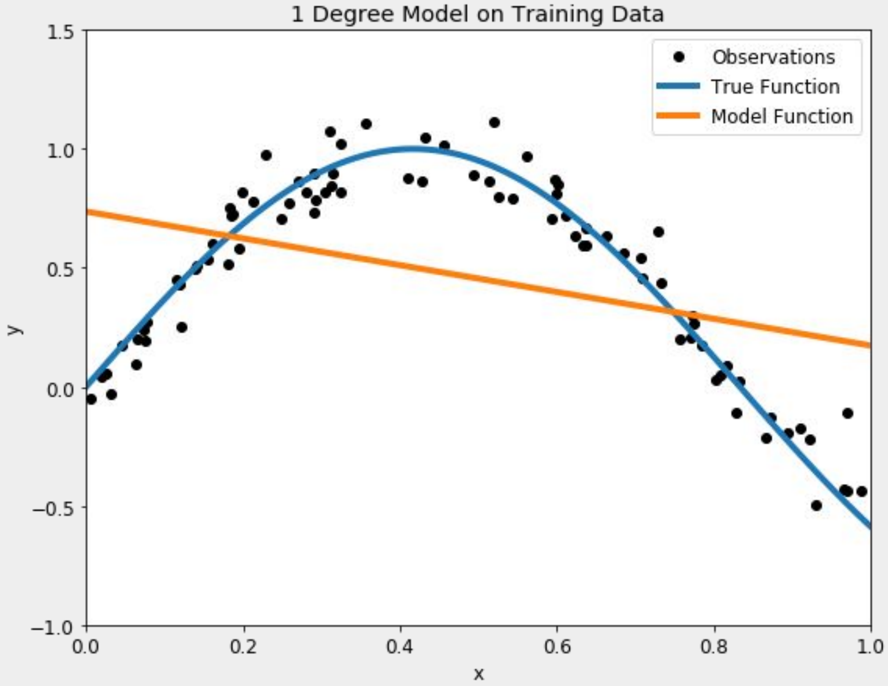
- Explain what the problems of over/underfitting are and how they relate to the bias/variance trade-off
- Explain how regularization addresses the over/underfitting problem
- Explain the differences between regularization methods and when it is appropriate to use them
- Explain the factors to consider when deciding which features to include
- Apply feature selection and regularization to prevent overfitting in a model

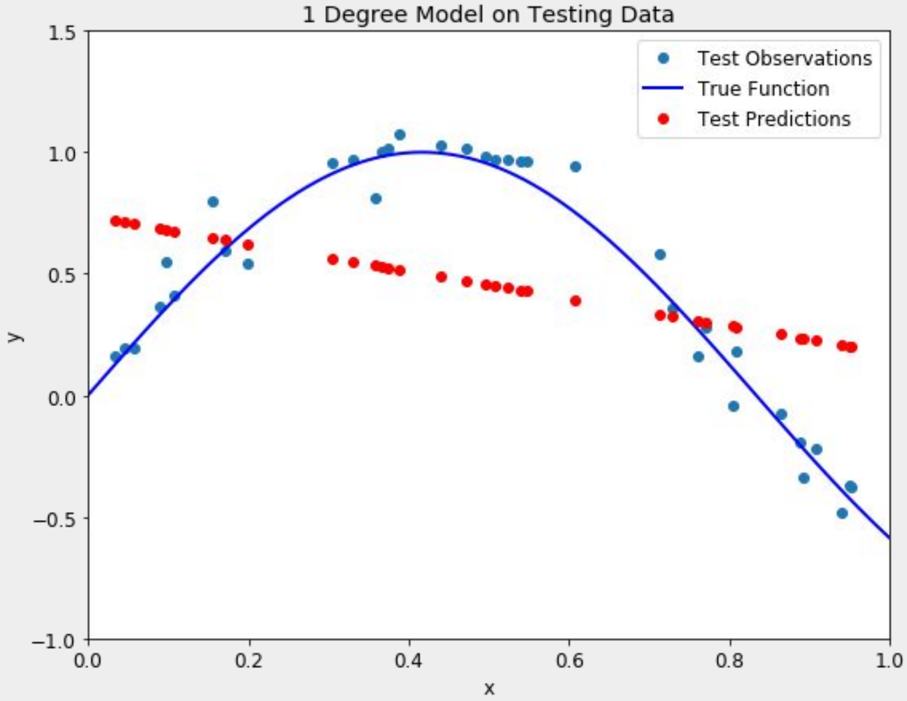
Question #1

- *What would the problem be if our model looked like this?*



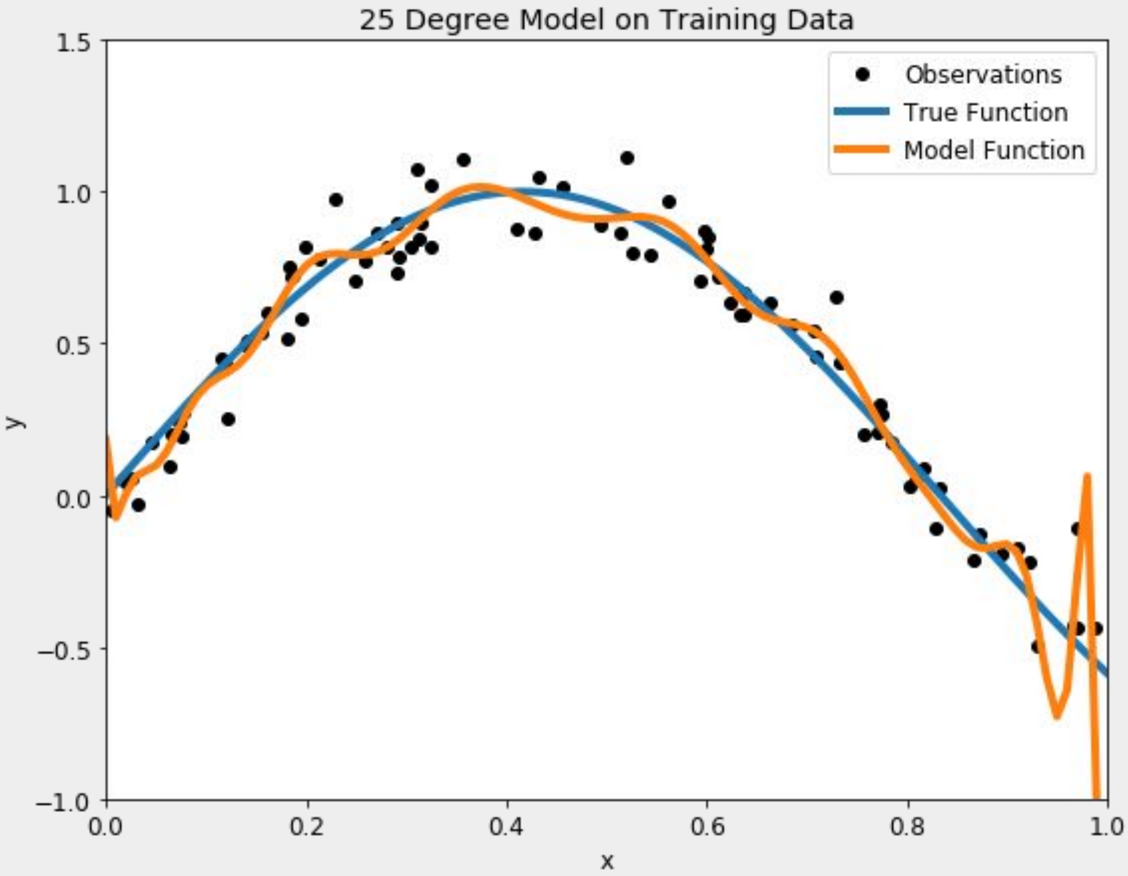
Question #1 - Review



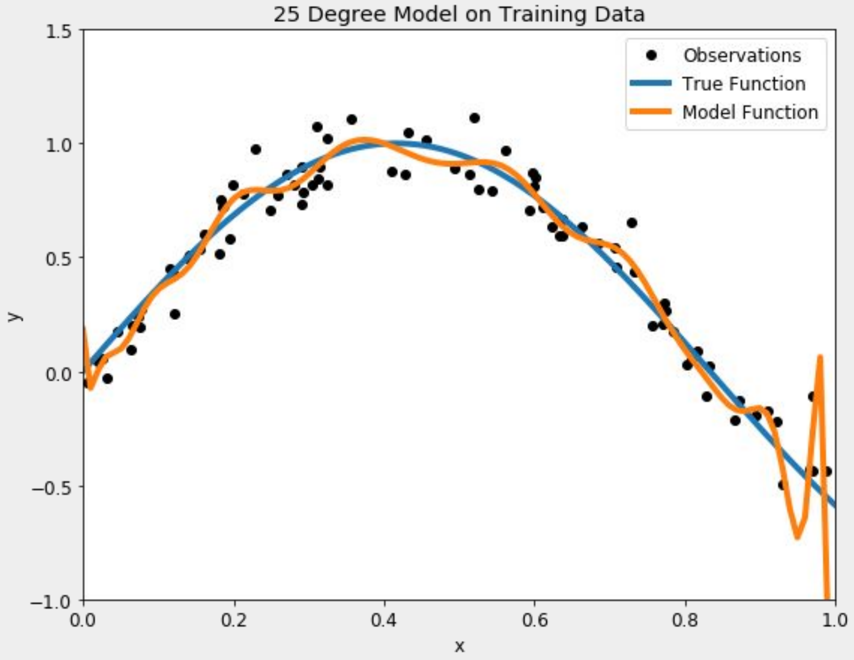


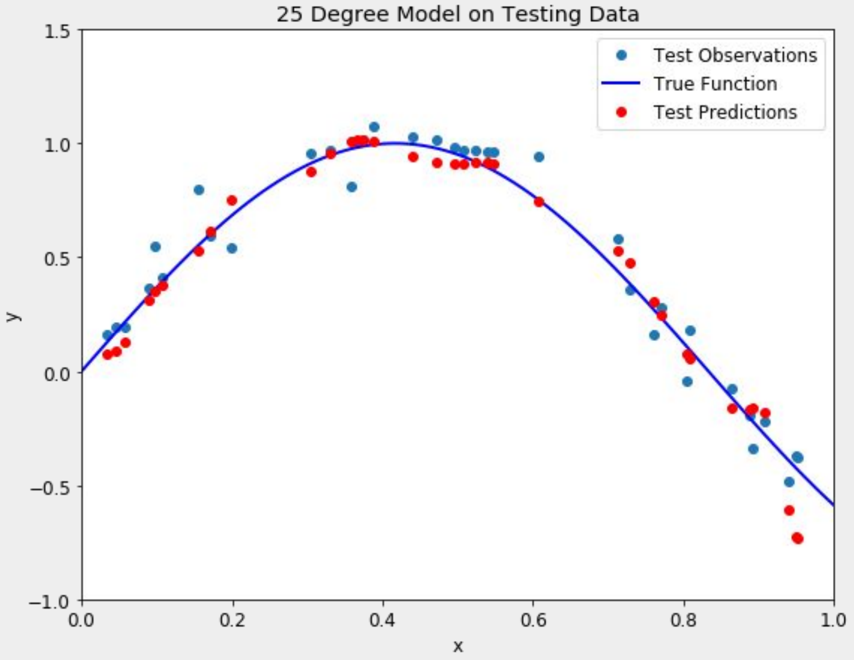
Question #2

- *What would the problem be if our model looked like this?*

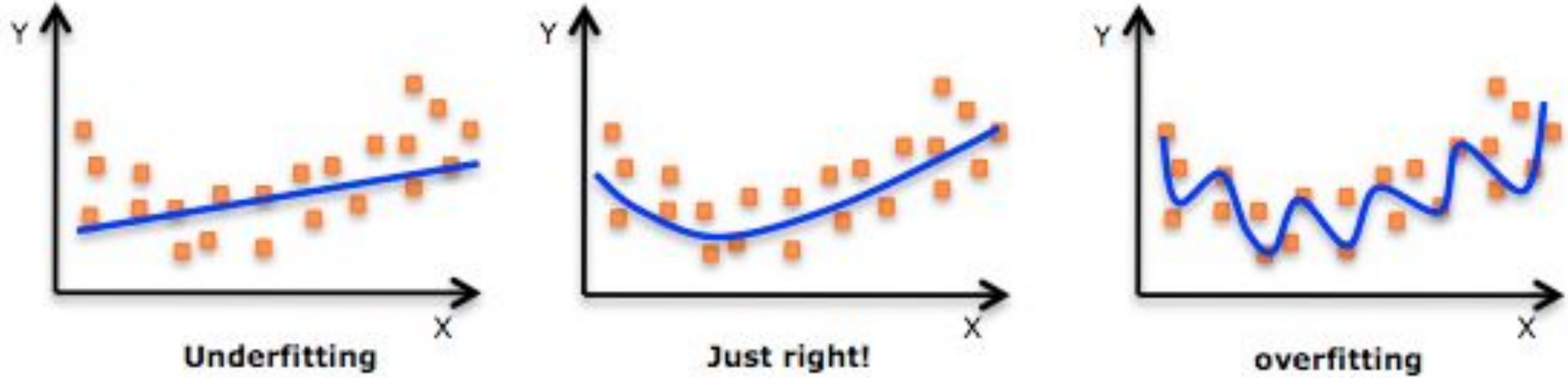


Question #2 - Review

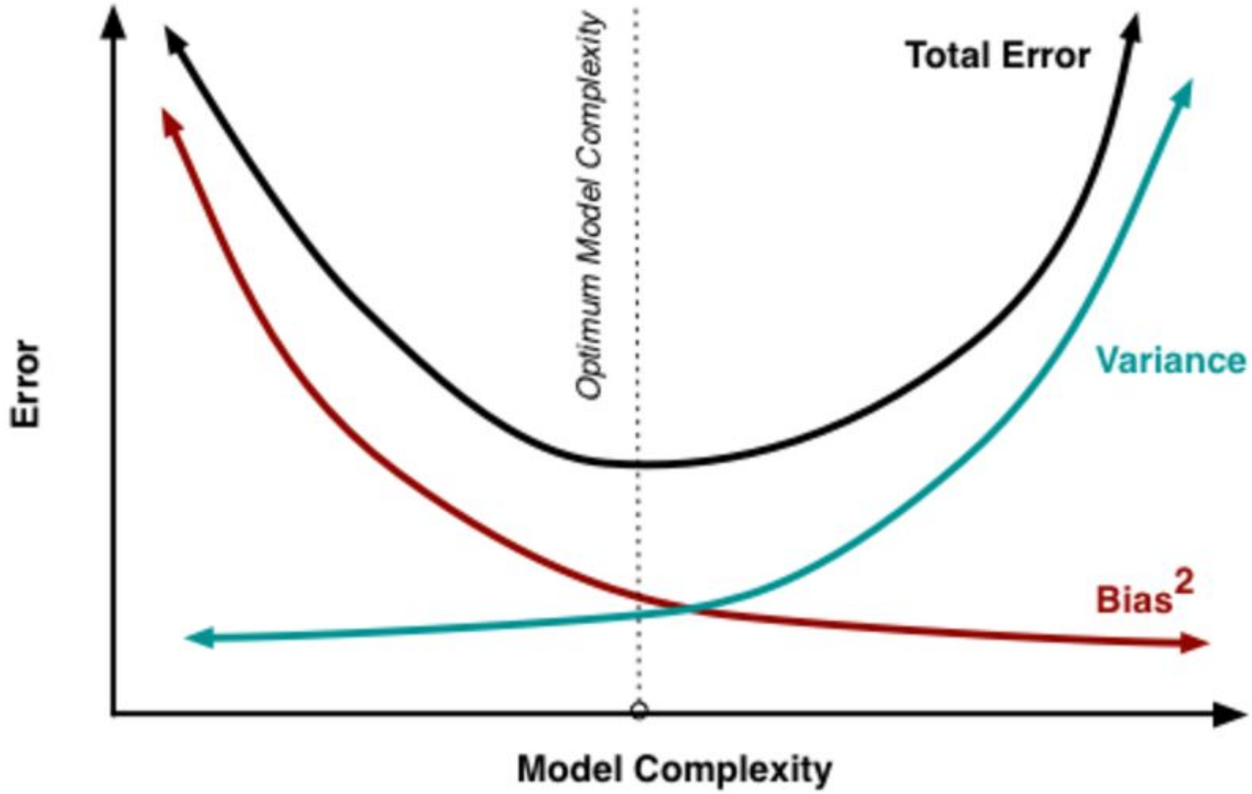




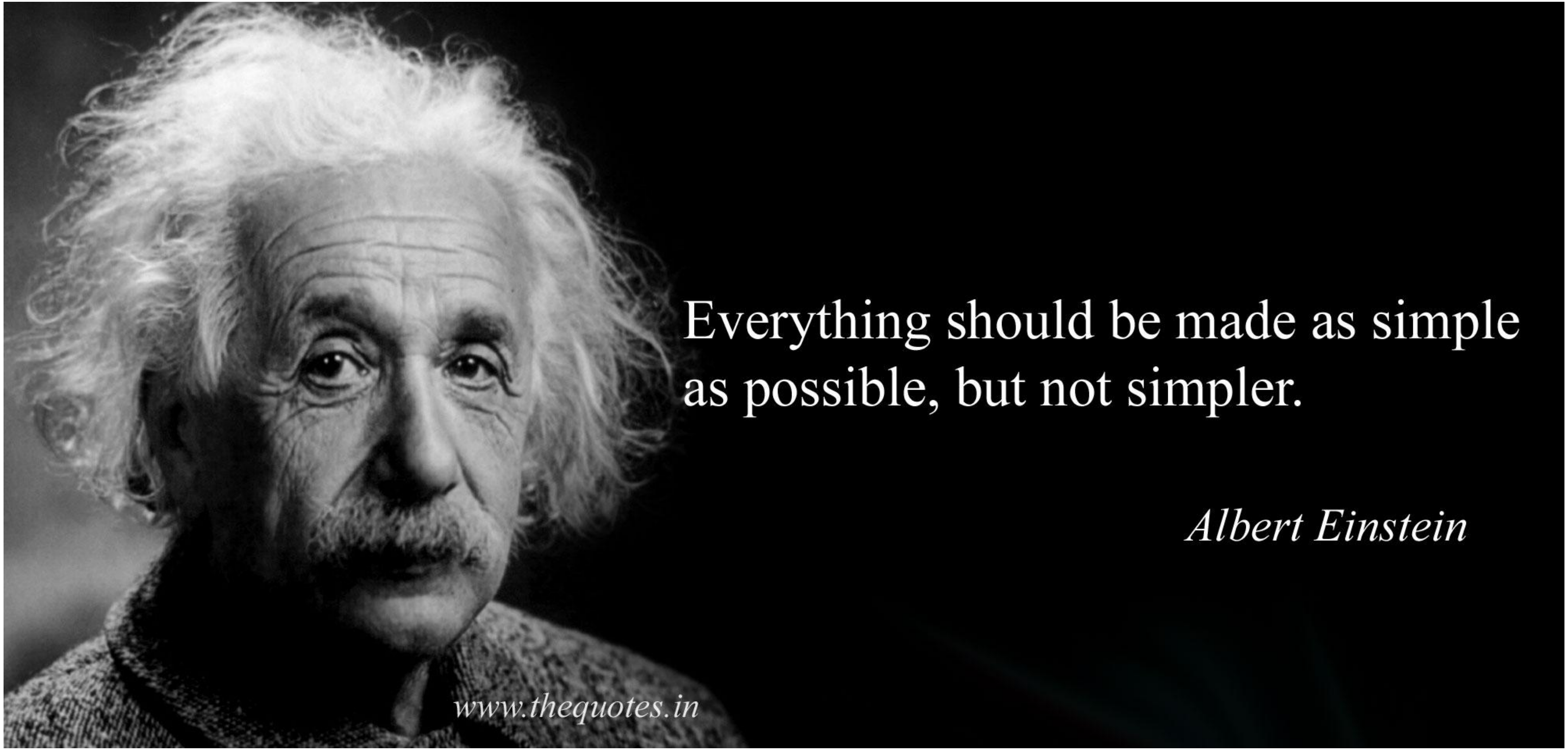
Underfitting vs. Overfitting



Bias Variance Trade-off



Regularization



Regularization

For simple linear regression, the cost function is represented as:

$$\text{cost\_function} = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}(y_i - bx + b_0)^2$$

For linear regression with multiple predictors, the cost function is expressed as so:

$$\text{cost\_function} = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{k}(m_j x_{ij}) + b)^2$$

Ridge

Ridge regression applies a penalizing parameter

$$\text{cost\_function\_ridge} = \sum_{i=1}^{n}(y_i - \hat{y})^2 + \lambda \sum_{j=1}^{p} m_j^2$$

$$= \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k}(m_j x_{ij} + b)\right)^2 + \lambda \sum_{j=1}^{p} m_j^2$$

Lasso

Lasso regression applies a penalizing parameter

$$\text{cost\_function\_lasso} = \sum_{i=1}^{n}(y_i - \hat{y})^2 + \lambda \sum_{j=1}^{p} | m_j|$$

$$= \sum_{i=1}^{n}(y_i - \sum_{j=1}^{k}(m_j x_{ij} + b))^2 + \lambda \sum_{j=1}^{p} | m_j|$$

Questions to ponder?

- When should you use a regularized model instead of a normal model?

- How do regularized models differ from normal linear regression?

- How does a Ridge model differ from a Lasso Model?