Generalization Error

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Measuring Generalization

- Training Set: A set of training examples the model is trained on.
- Validation Set: This dataset is used to tune hyperparameters such as the number of hidden units, or the learning rate.
- Test Set: This dataset is used to measure the generalization performance.
- The losses on these subsets are called training , validation , and test loss, respectively.
- Cost Function: We can defined a cost function, the average loss over the training set as follows:

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(y_i, \hat{y}_i)$$

- Generalization Error: In addition to predicting the training examples correctly, the model should also be capable of generalizing to new cases.
- What is the purpose of the datasets above?

Bias + Variance

- Define Bias and Variance:
  - Bias: is commonly defined as the difference between the expected value of the estimator and the parameter that we want to estimate.
  - Variance: is defined as the difference between the expected value of the squared estimator minus the squared expectation of the estimator.

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta, \quad \text{Var}(\hat{\theta}) = \mathbb{E}\left[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2\right].$$

- Start with the squared error loss decomposition into bias and variance.

$$\begin{aligned} S &= (y - \hat{y})^2 \\ (y - \hat{y})^2 &= (y - \mathbb{E}[\hat{y}] + \mathbb{E}[\hat{y}] - \hat{y})^2 \\ &= (y - \mathbb{E}[\hat{y}])^2 + (\mathbb{E}[\hat{y}] - y)^2 + 2(y - \mathbb{E}[\hat{y}])(\mathbb{E}[\hat{y}] - \hat{y}). \end{aligned}$$

- Employing the expectation on both sides:

$$\mathbb{E}[S] = \mathbb{E}[(y - \hat{y})^2] = (y - \mathbb{E}[\hat{y}])^2 + \mathbb{E}[(\mathbb{E}[\hat{y}] - \hat{y})^2] = [\text{Bias}]^2 + \text{Variance}$$

Underfitting

- Underfitting is the opposite of overfitting: it occurs when your model is too simple to learn the underlying structure of the data.

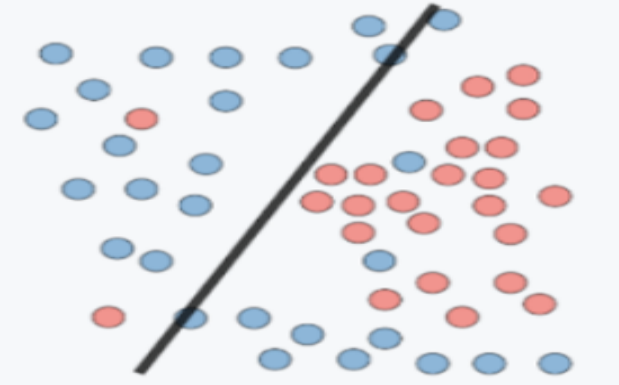

Figure: High Variance, Source

- The main options to fix this problem:
  - Selecting a more powerful model, with more parameters
  - Feeding better features to the learning algorithm (feature engineering)
  - Reducing the constraints on the model (e.g., reducing the regularization hyperparameter)

Overfitting

- Overfitting means the model works well on training data, but it doesn't generalize well.
- Overfitting occurs when there is too much complexity in the model in comparison to the amount and noise in the training data.

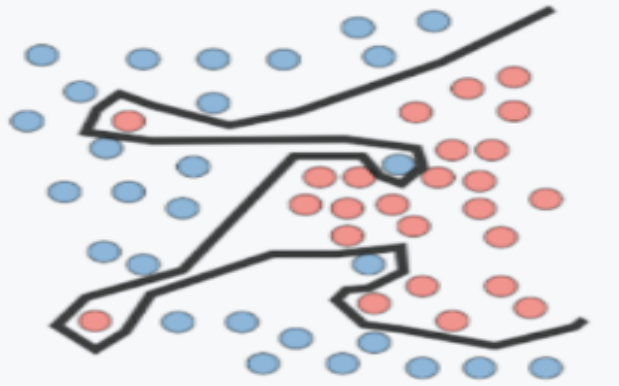

Figure: High Variance, Source

- Three practical solutions:
  - Simplify the model by selecting one with fewer parameters, reducing the number of attributes in the training data, or constraining the model.
  - Gather more training data.
  - Reduce the noise in the training data.

Overview

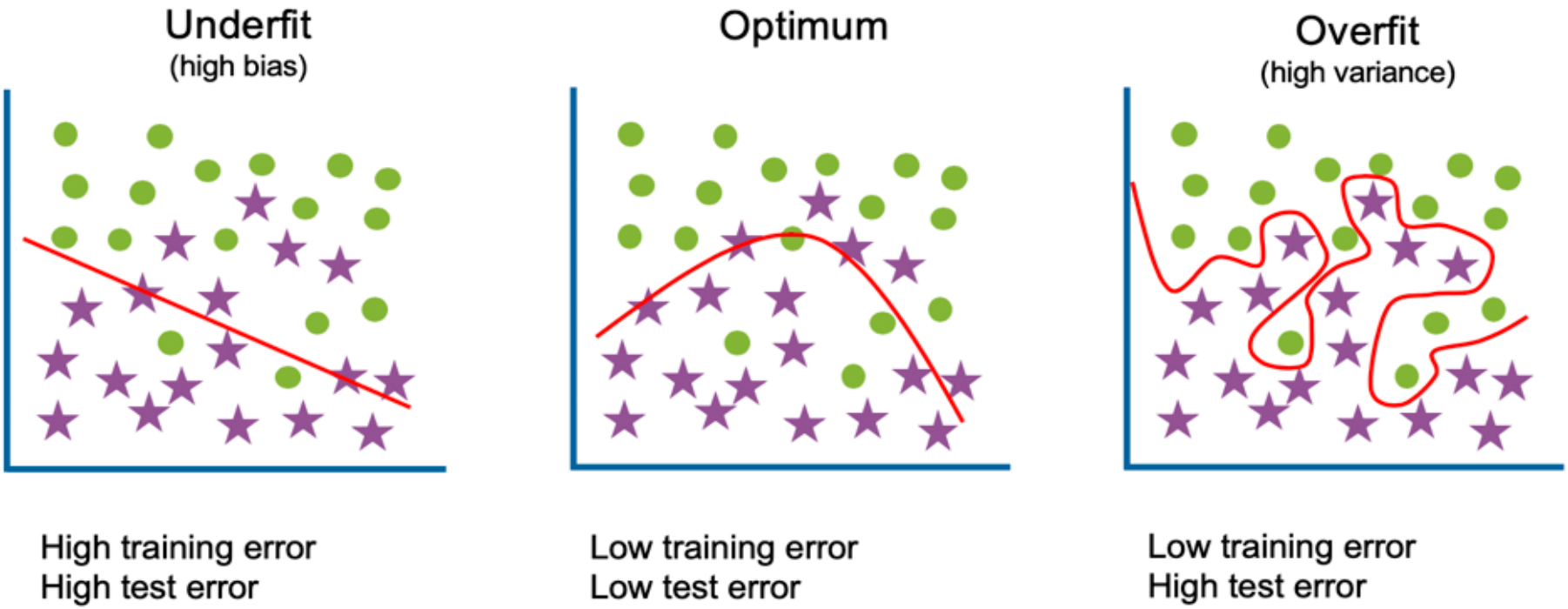

Figure: Overfitting vs Underfitting, Source

Overview

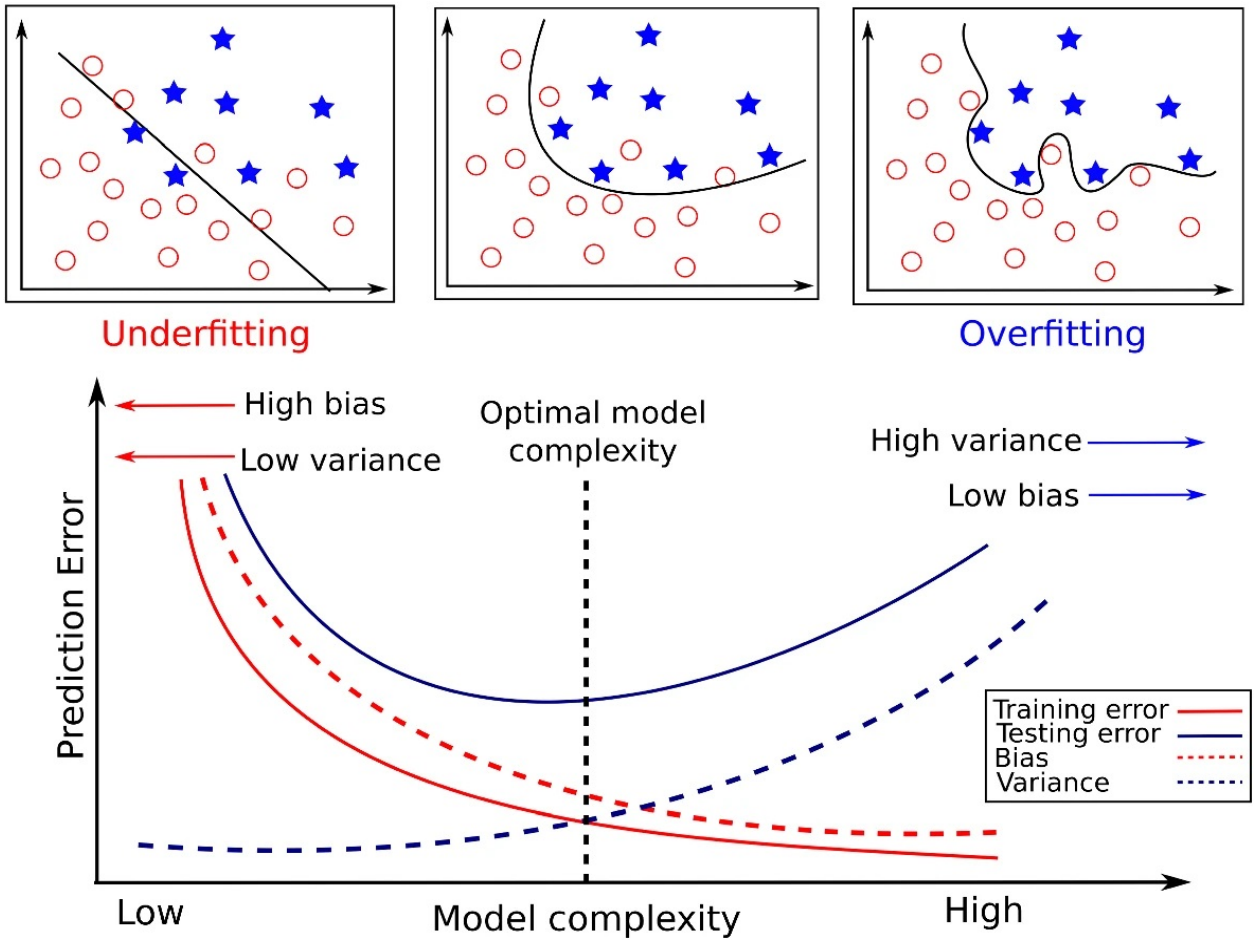

Figure: Overfitting vs Underfitting, Source

Thank You!

Any Question?