Approximation and Genralizatoin error

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Generalization

- Some of the evaluation metrics we can use to measure the performance on the test set are the prediction accuracy and misclassification error in the context of classification models – we say that a good model has a "high generalization accuracy" or "low generalization error" (or, simply "good generalization performance").

Validation set

- In order to choose right model complexity and train the hyperparameters we cannot use:
  - training set: because we want to choose values that will generalize.
  - test set:, because that would be “cheating.” We’re only allowed to use the test set once, to report the final performance.
- A validation set, which is used to tune hyperparameters

Overfitting and Underfitting

- overfitting and underfitting are two terms that we can use to diagnose a machine learning model based on the training and test set performance.
- a model that suffers from underfitting does not perform well on the test and training set
- In contrast, a model that overfits can be usually recognized by a high training set accuracy, but low test set accuracy

Overfitting and Underfitting

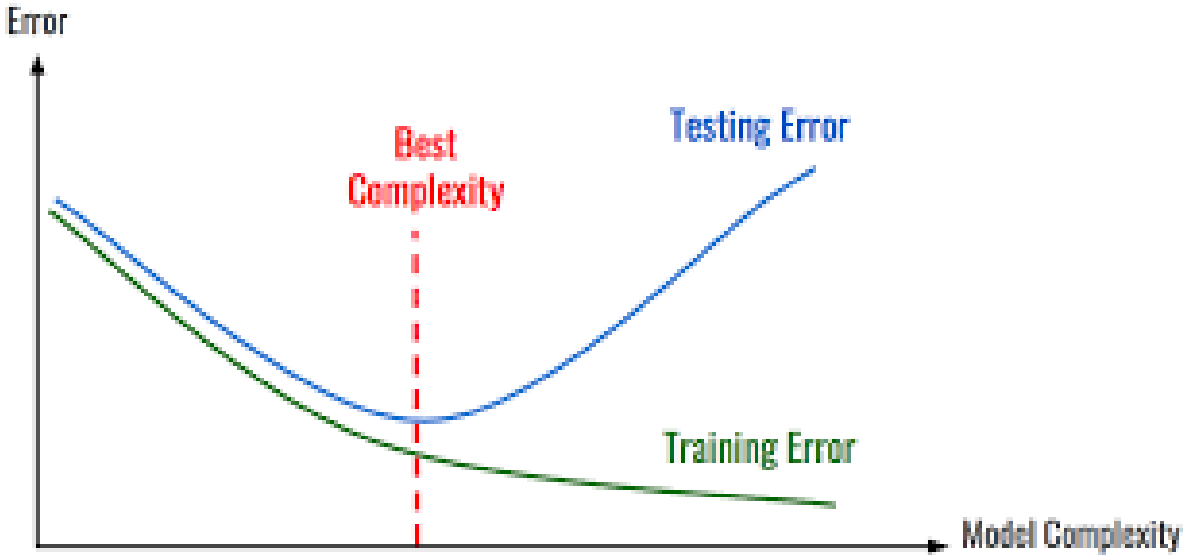

Figure: Overfitting and Underfitting in terms of train and test set error, Source

Bias-Variance Decomposition of the Squared Loss

- Let y be the true output value and $\hat{y}$ be the predicted output value that depends on the dataset generated by some unknown joint distribution and hence is random and we define the expected value w.r.t the data set as $\mathbb{E}[\hat{y}]$ the squared loss $(\hat{y} - y)^2$ can be decomposed as:
  (y -
  $\hat{y})^2 = ((y - \mathbb{E}[\hat{y}]) + (\mathbb{E}[\hat{y}] - \hat{y})^2 = (y - \mathbb{E}[\hat{y}])^2 + (\mathbb{E}[\hat{y}] - \hat{y})^2 + 2(y - \mathbb{E}[\hat{y}])(y - \mathbb{E}[\hat{y}])$
- Taking expected value w.r.t dataset we get:

$$\mathbb{E}[(y - \hat{y})^2] = (\underbrace{y - \mathbb{E}[\hat{y}]}_{\text{bias}})^2 + \underbrace{\mathbb{E}[(\mathbb{E}[\hat{y}] - \hat{y})^2]}_{\text{variance}}$$

Bias and Variance intuition

- The first term is the bias, which tells us how far off the model's average prediction is.
- The second term is the variance, which tells us about the variability in its predictions as a result of the choice of training set.

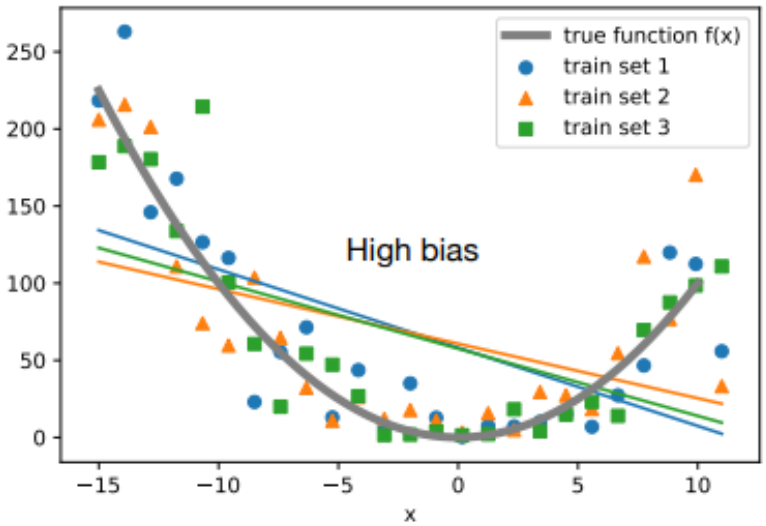

Figure: the bias is large because the difference between the true value and the predicted value, on average is large

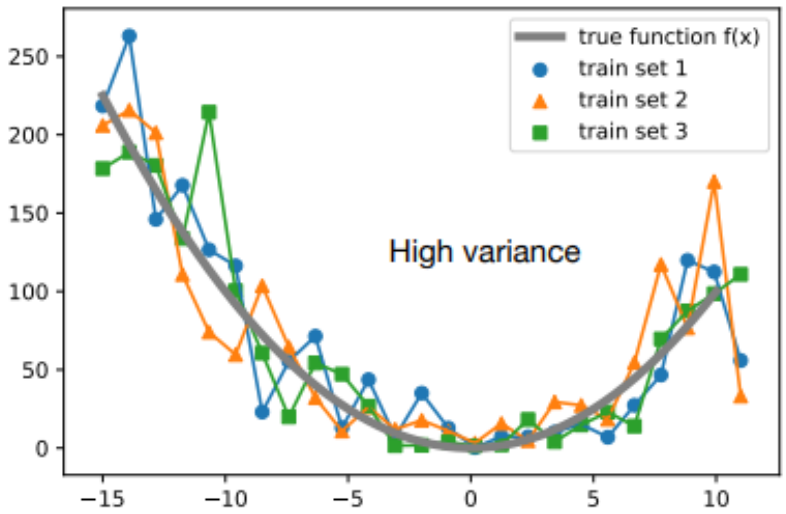

Figure: The variance is very high, since on average, a prediction differs a lot from the expectation value of the prediction

References

- https://www.cs.toronto.edu/ rgrosse/coursescsc321_2018/readings/L09%20Generalization.pdf
- https://sebastianraschka.com/pdf/lecture-notes/stat479fs19/08-model-eval-1-intro__notes.pdf

Thank You!

Any Question?