

Overfitting versus Underfitting

Ms. Sandhya Harikumar,
Department of Computer Science & Engineering,
Amrita Vishwa Vidyapeetham, Amritapuri



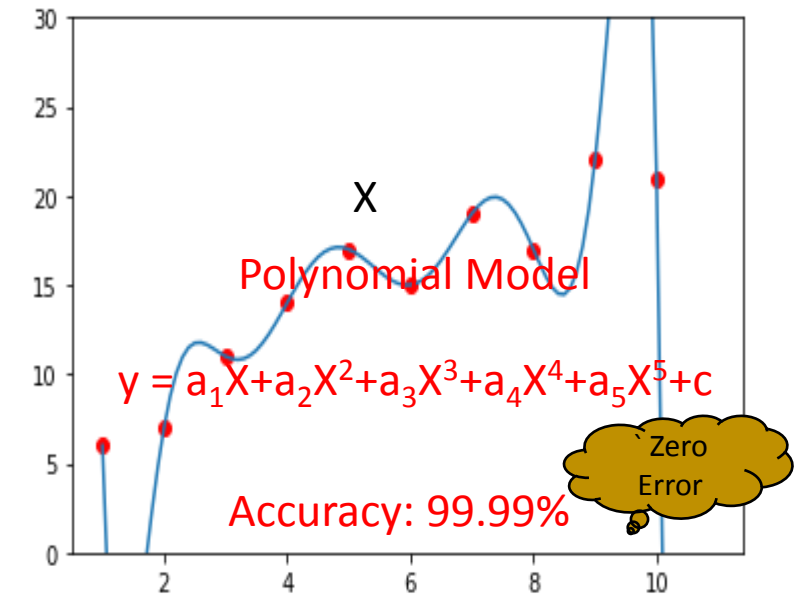
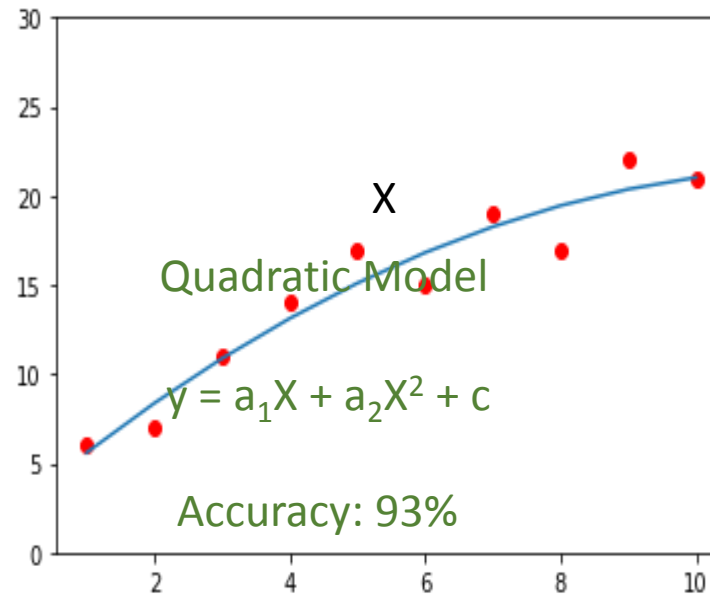
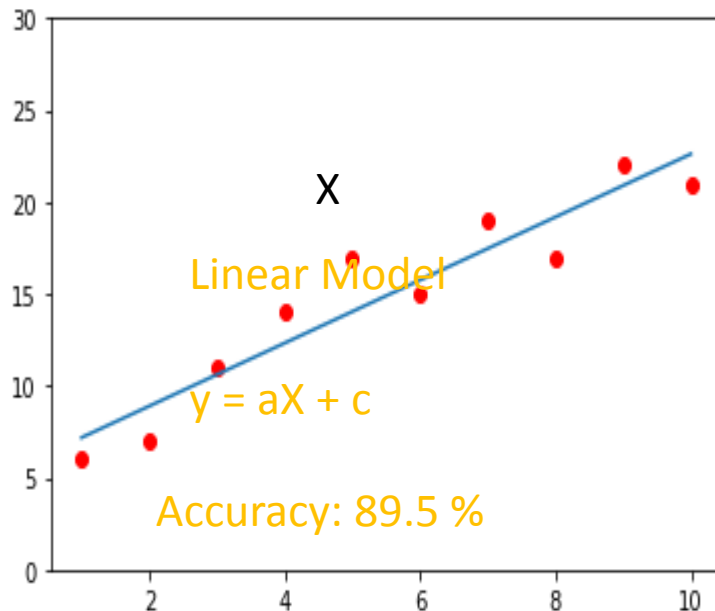
Supervised Machine Learning

- Algorithm learns a model from training data
- Goal is to best estimate the mapping function (f)
$$f(X) \rightarrow Y$$
- Inductive Learning : Learning general concepts from specific examples



Underfitting and Overfitting

- How well a machine learning model learns and generalizes to new data
- Fit : Refers to how well you approximate a target function.



Model Evaluation

- To better understand machine learning algorithms
- To get better performance on your data.
- Overfitting or underfitting the data - Cause of poor performance in machine learning
- Overfitting – Less Generalization
- Underfitting – More assumptions



Training Error

Guessing: ~50%

Underfitting

Overfitting

Mr. know it all
~98%



Quiz based on
class work

Good fit

Problem solving approach:
~92%

Professor



A



B



C

Testing Error



Guessing: ~47%

A



Mr. know it all
~69%

B



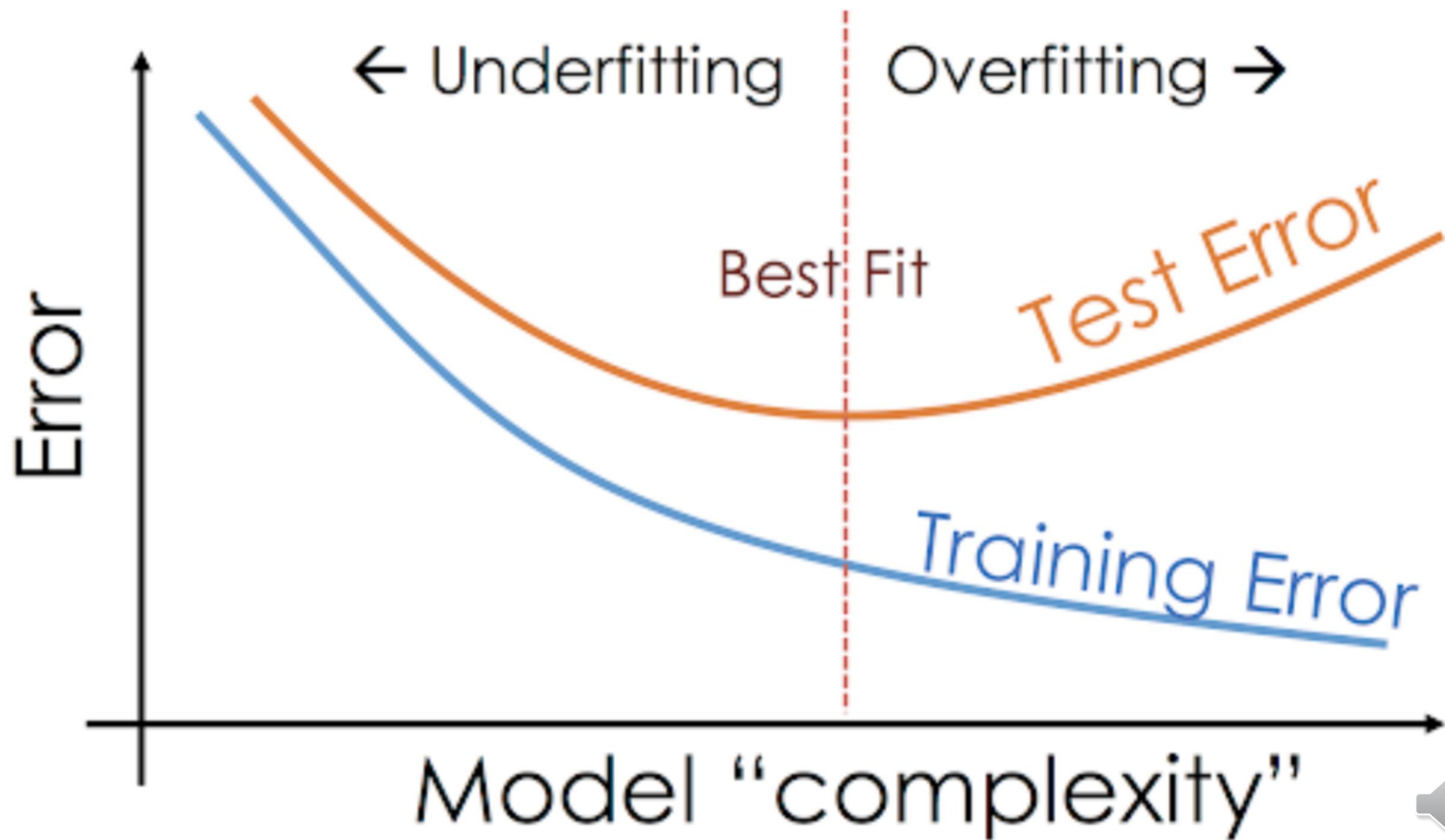
Problem solving approach:
~89%



Semester Exam

Professor





Takeaways

- Overfitting – Less Generalization and complex model
- Underfitting – Less Generalization and simple model



References

- <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>
- <http://scott.fortmann-roe.com/docs/BiasVariance.html>
- <https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/#:~:text=Bias%20is%20the%20simplifying%20assumptions,the%20bias%20and%20the%20variance.>
- <https://www.analyticsvidhya.com/blog/2020/02/underfitting-overfitting-best-fitting-machine-learning/>