CSIS3764
DATA SCIENCE

Mr WSJ Marais

Underfitting vs Overfitting

Bias vs Variance

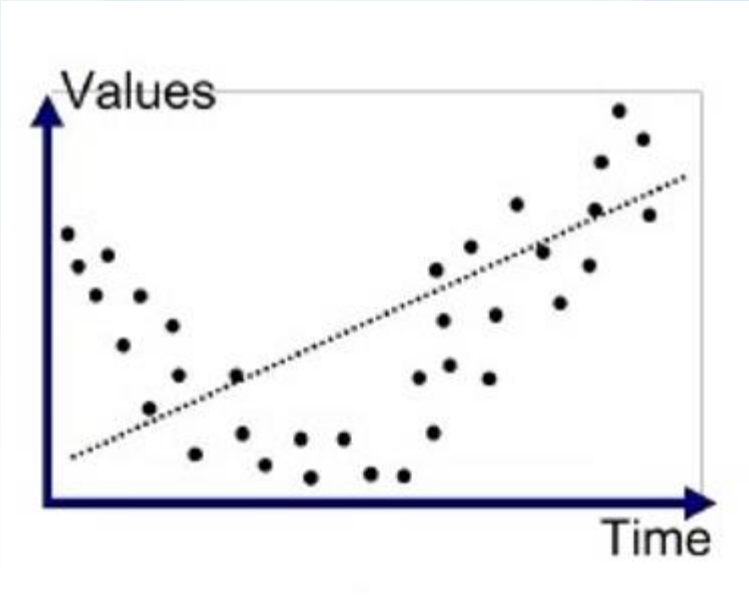
T: 051 401 2754 csi@ufs.ac.za **www.ufs.ac.za/csi**

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS·UV
NATURAL AND
AGRICULTURAL SCIENCES
NATUUR- EN
LANDBOUWETENSKAPPE

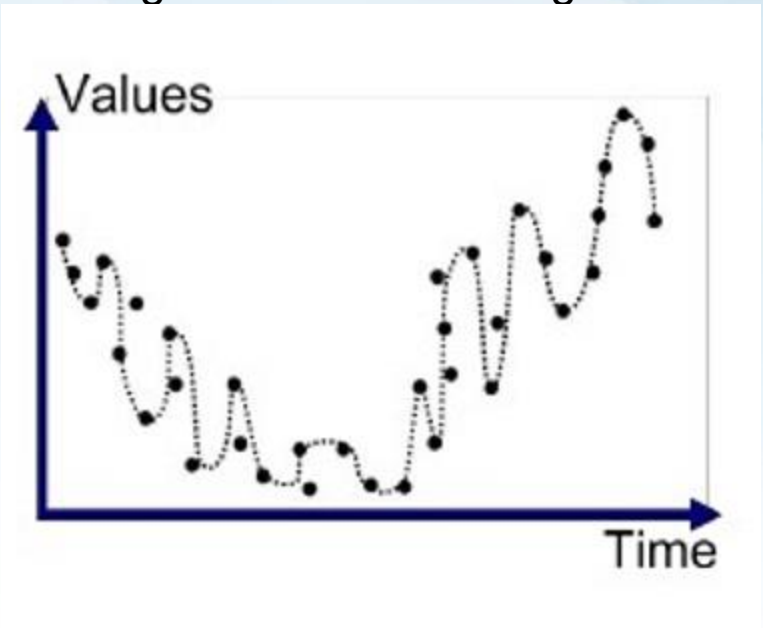
UNDERFITTING

- The model is simple
- The model poorly represents the relationship in the data
- Produces poor accuracy results with the training data
- Produces similar accuracy results with test/new data
- The sums of squares are substantial but similar for training and test data
- The model has a high bias
- The model has a low variance



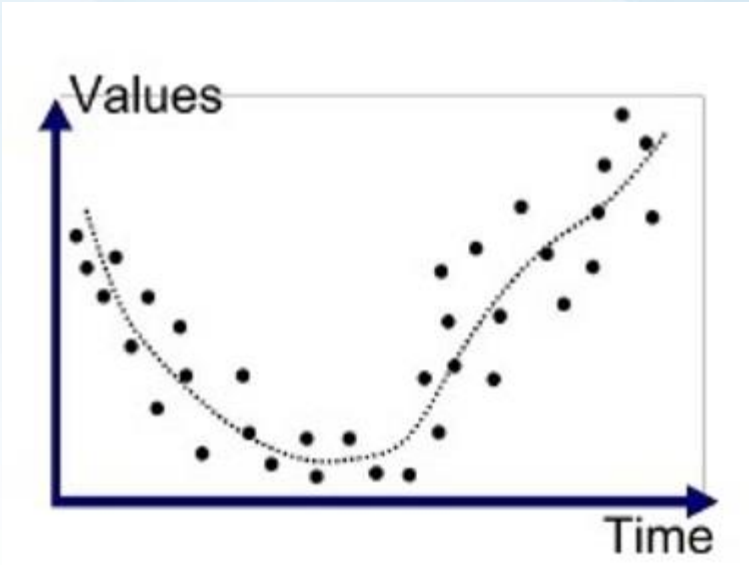
OVERFITTING

- The model is complex
- The model adapts itself too much to the training data
- Produces very good accuracy results with the training data
- Produces poor accuracy results with test/new data
- The sums of squares is basically zero for training data but much larger for the test/new data
- The model has a low bias
- The model has a high variance



Good Fit

- The model is balanced between complex and simple
- The model represents the relationship in the data quite good but not exactly
- Produces good accuracy results with the training data
- Produces good accuracy results with test/new data
- The model has a low bias
- The model has a low variance



UFS UV

VIDEOS

- Machine Learning Fundamentals: Bias and Variance
- Bias/Variance

PREVENTION

- If model has high bias
  - Use bigger neural network / more complex algorithms
  - Train longer
  - Try other algorithms
- If model has high variance
  - More data
  - Regularization
  - Try other algorithms

Trade Off

- Usually there is a trade off between bias and variance
- Modern machine/deep learning has less trade off between bias and variance

VIDEOS

- Basic Recipe for Machine Learning