

Epilogue

Gianluca Campanella

Contents

Use the right tools for the job

Don't fool yourself

Don't be evil

Use the right tools for the job

What is Data Science?

A **problem-solving approach**
based on the scientific method

Use the right tools for the job

Remember that the focus is on **adding value**

- Find and use the right tools for the job
 - Even if it happens to be Excel!

Don't try to run before you can walk

THE DATA SCIENCE **HIERARCHY OF NEEDS**

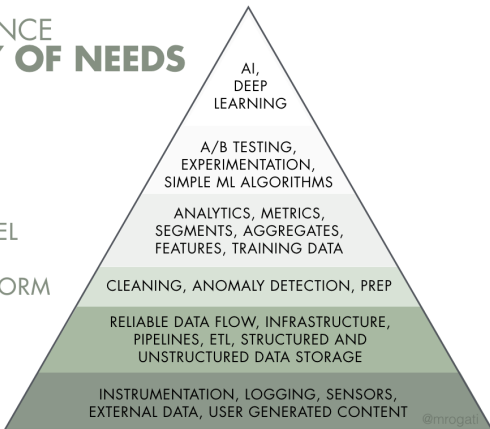
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



From M. Rogati

Machine Learning models

	Categorical	Numerical
Supervised	Classification	Regression
Unsupervised	Clustering	Dimensionality reduction

Don't fool yourself

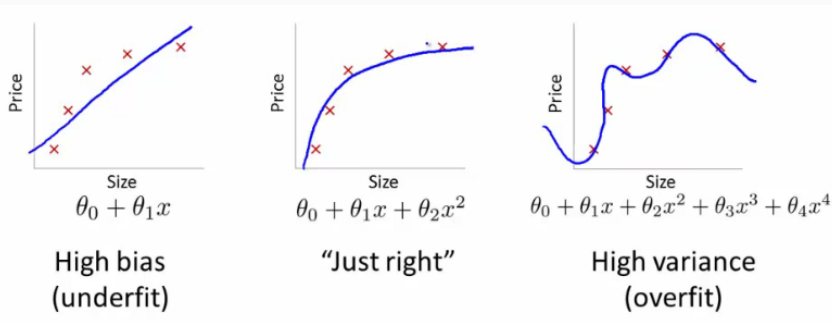
Don't fool yourself

It's only numbers after all!

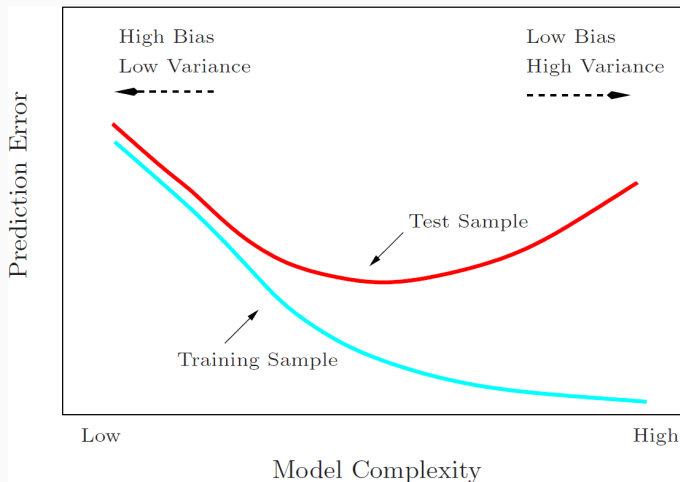
- There is no causation without manipulation
- If it sounds too good to be true...

Bias-variance trade-off

- Irreducible variance due to the stochastic process
- Bias in approximating f^* using f
- Estimation variance of \hat{f}



Bias-variance trade-off and generalisability



From *The Elements of Statistical Learning*

Don't be evil

The biggest threat is not from AI...

The biggest threat is not from AI...

...but from the humans who control it