A Few Useful Things to Know about Machine Learning

Pedro Domingos

Qingkai Kong (present)

Overview

- First published in 2012
- Summarize 12 key lessons relate to ML

1. LEARNING = REPRESENTATION + EVALUATION + OPTIMIZATION

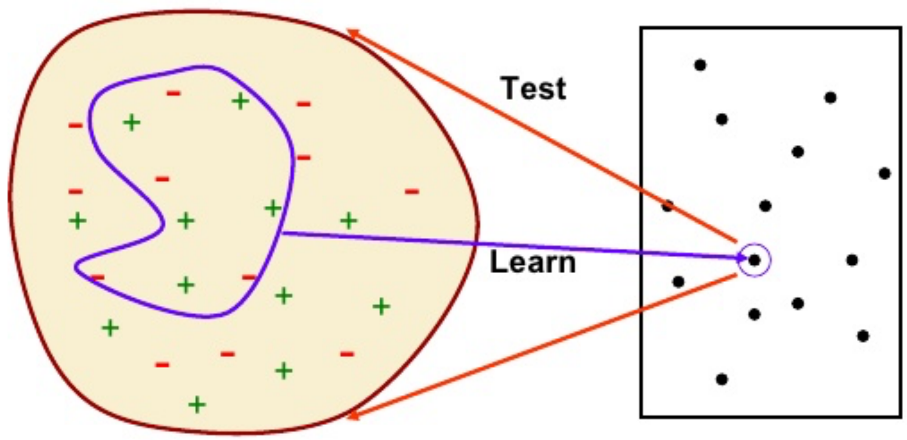
Table 1: The three components of learning algorithms.

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

2. GENERALIZATION THAT COUNTS

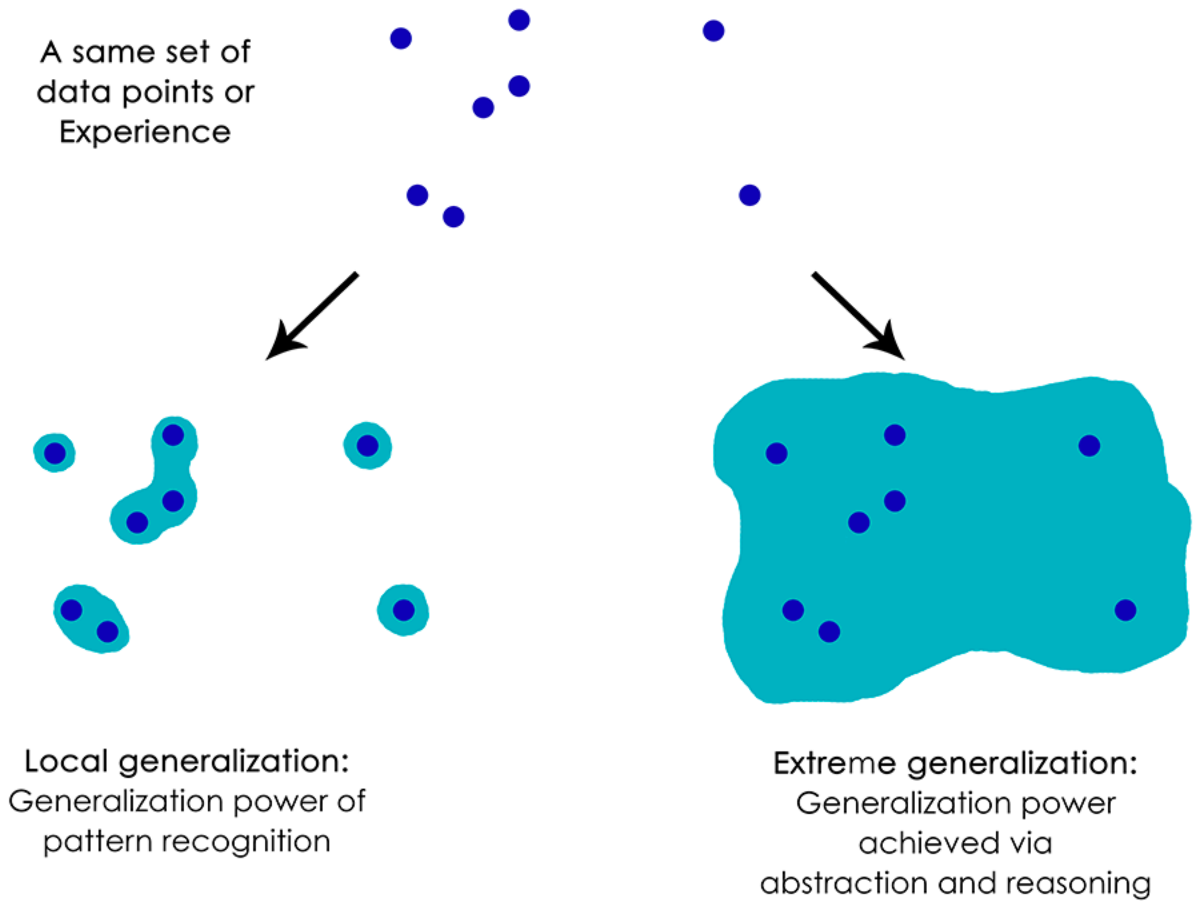
Generalization Error

- A hypothesis h is said to *generalize* well if it achieves low error on all examples in X



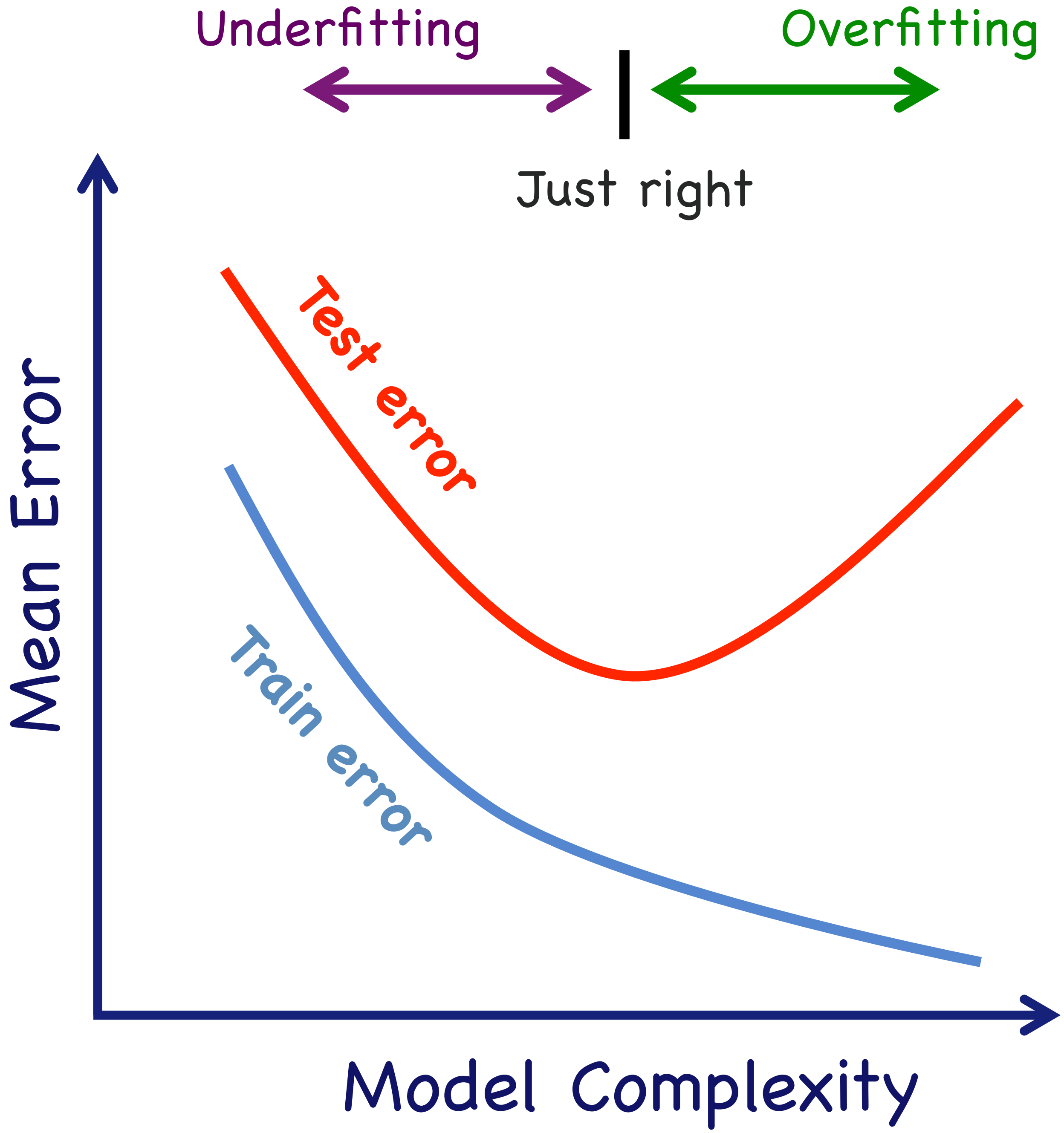
Example set X Hypothesis space H

2. GENERALIZATION THAT COUNTS

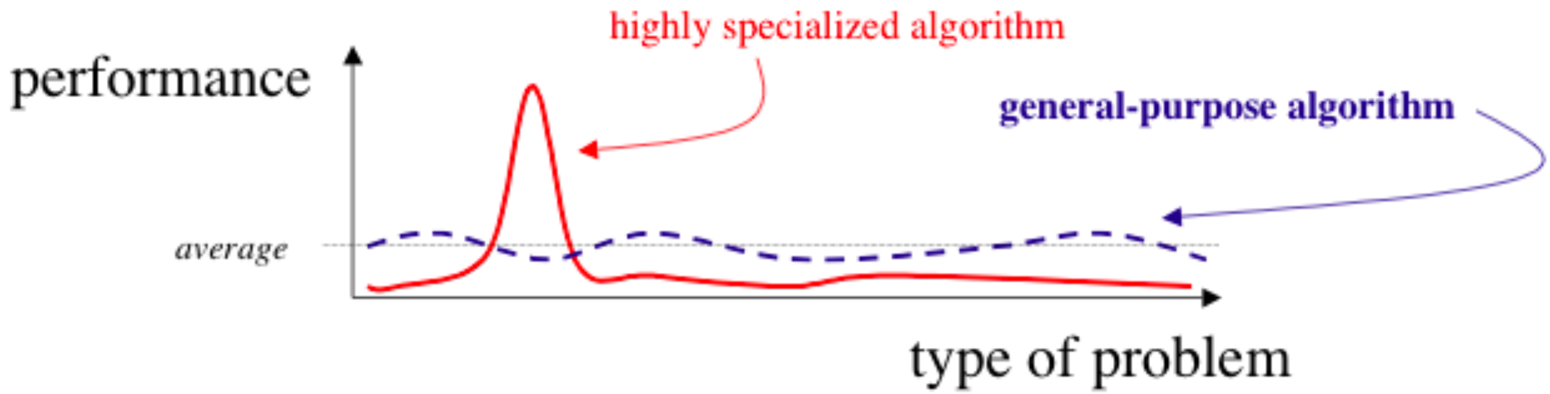


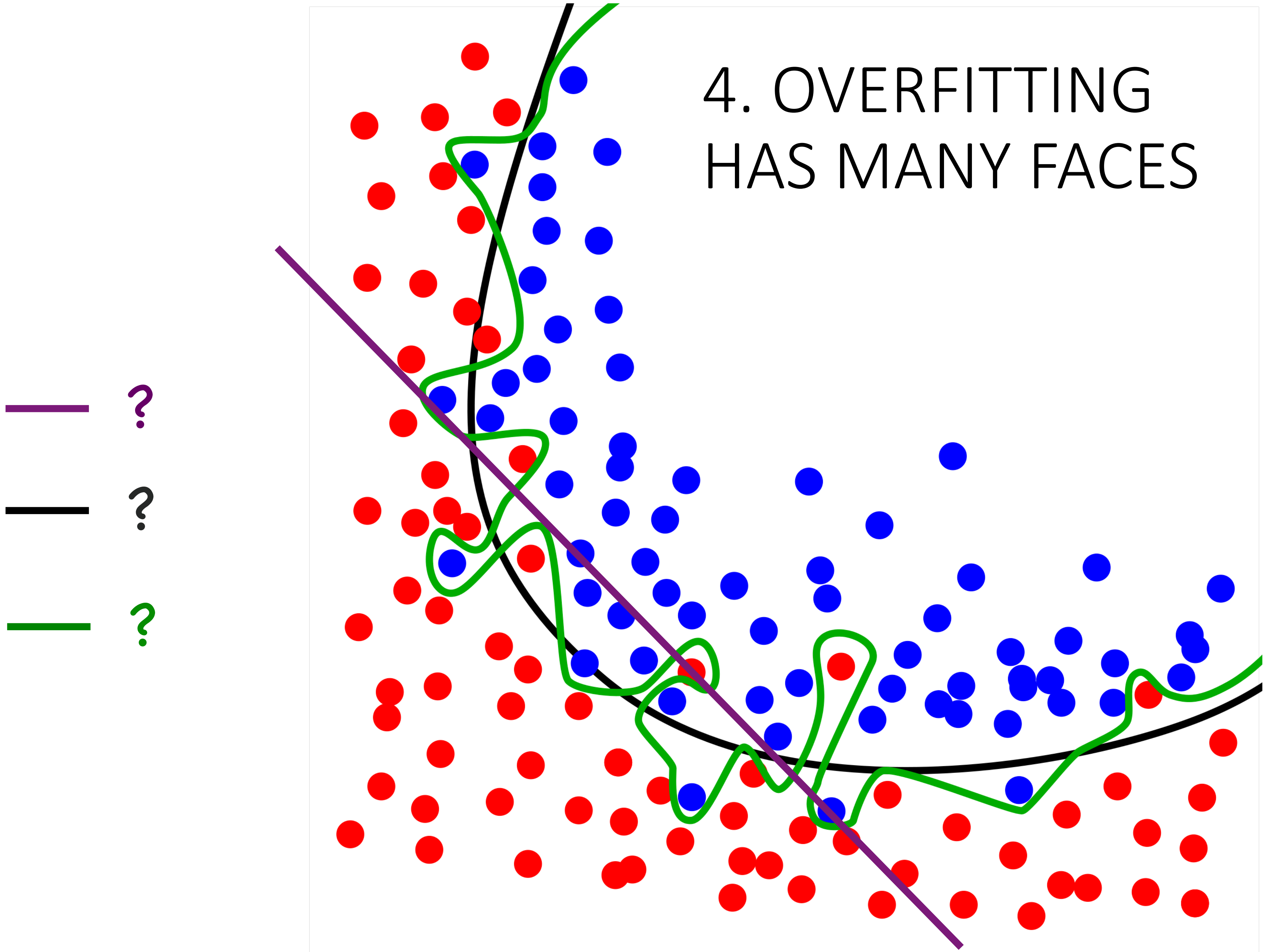
Train/test dataset split

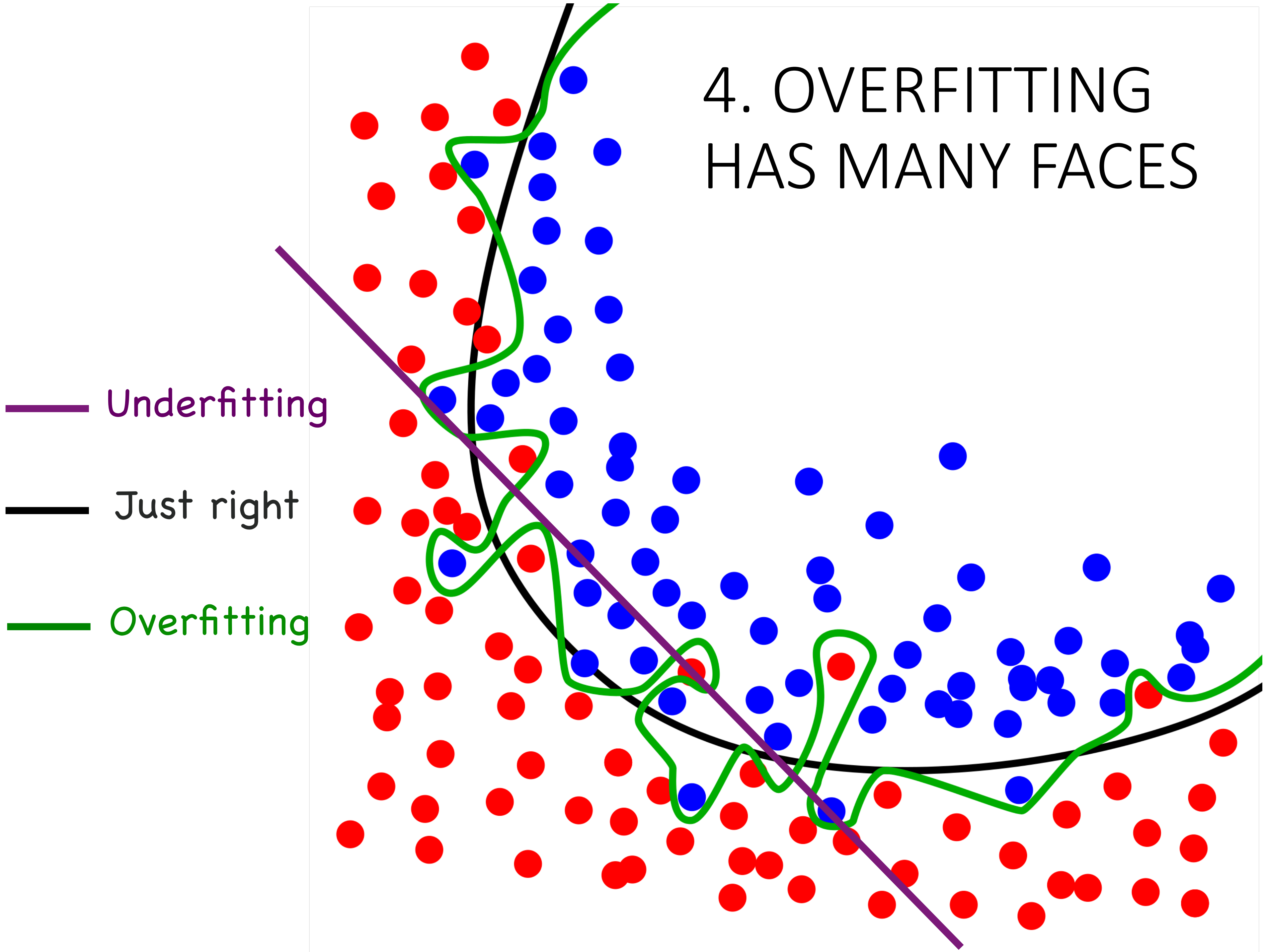
Train/test dataset split


Underfitting
Overfitting
Just right
Mean Error
Test error
Train error
Model Complexity

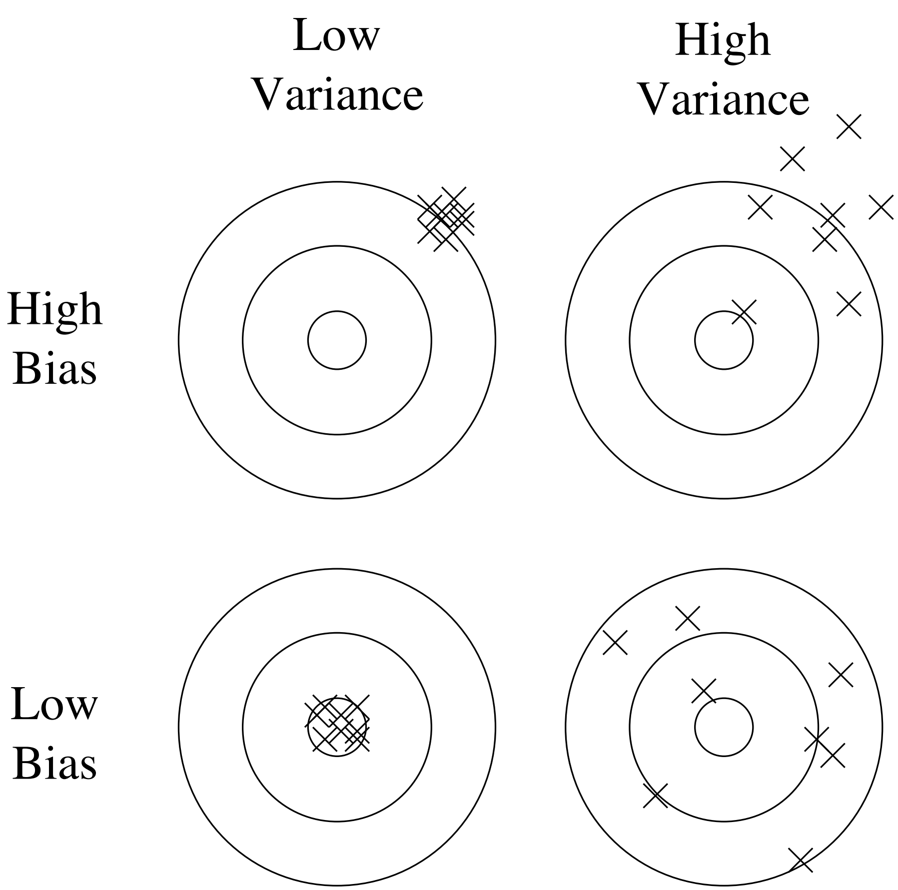
3. DATA ALONE IS NOT ENOUGH

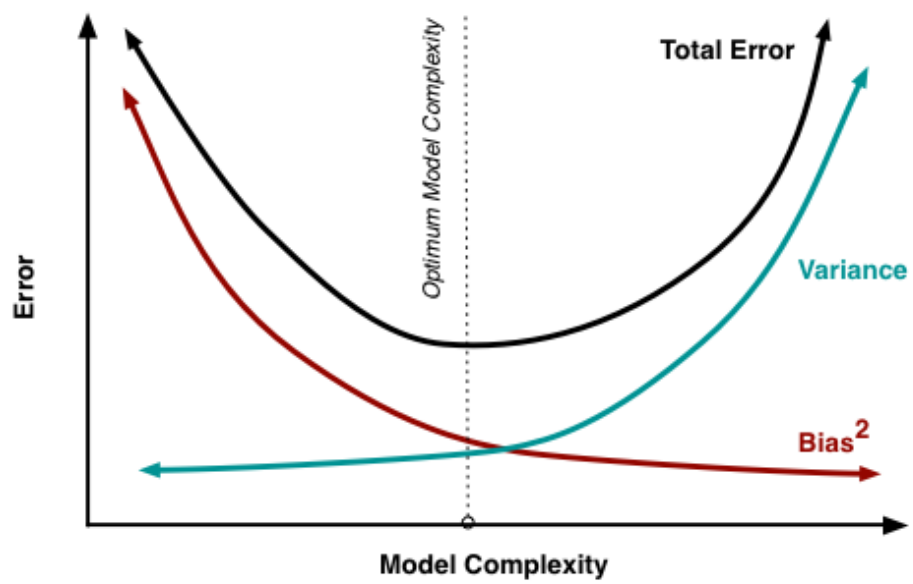



4. OVERFITTING
HAS MANY FACES
?
?
?

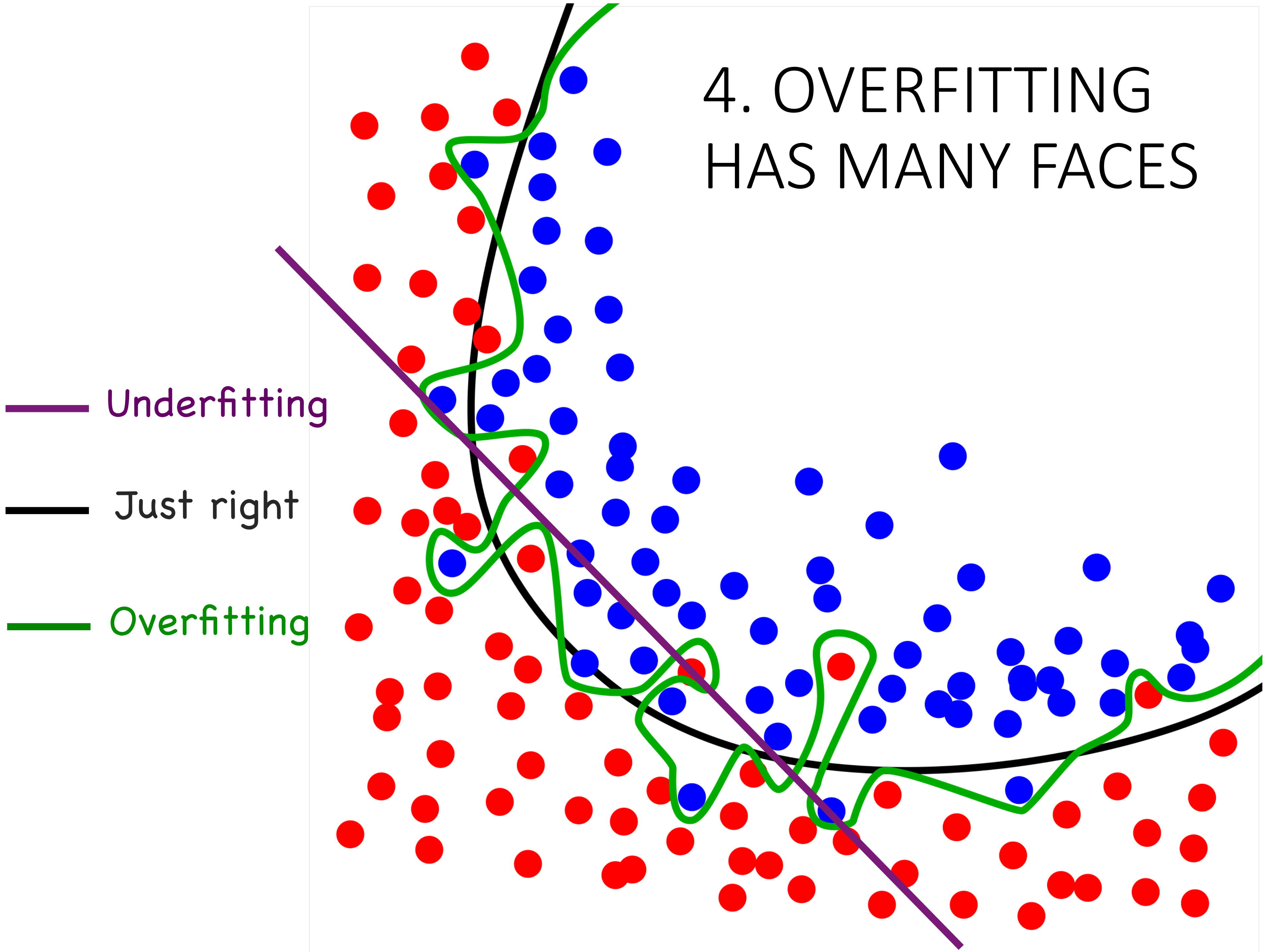

4. OVERFITTING
HAS MANY FACES
Underfitting
Just right
Overfitting

4. OVERFITTING HAS MANY FACES



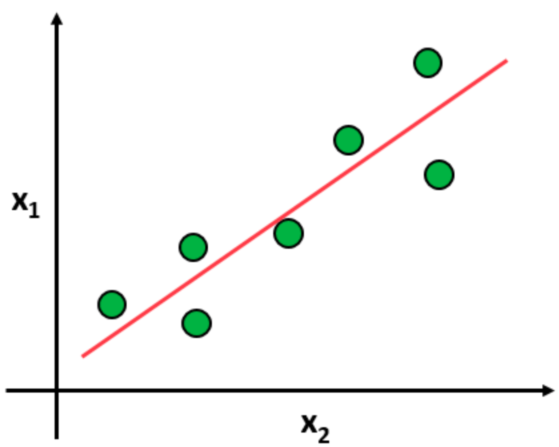


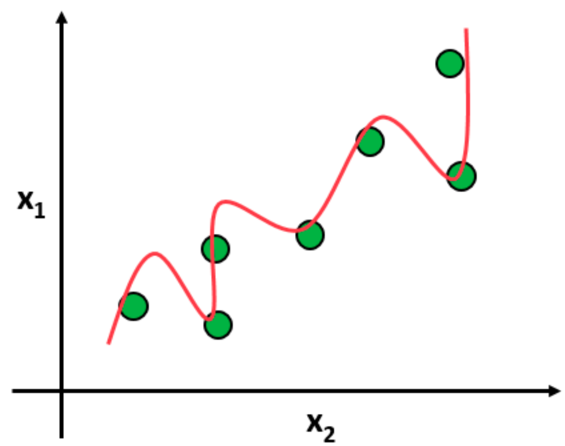
4. OVERFITTING HAS MANY FACES

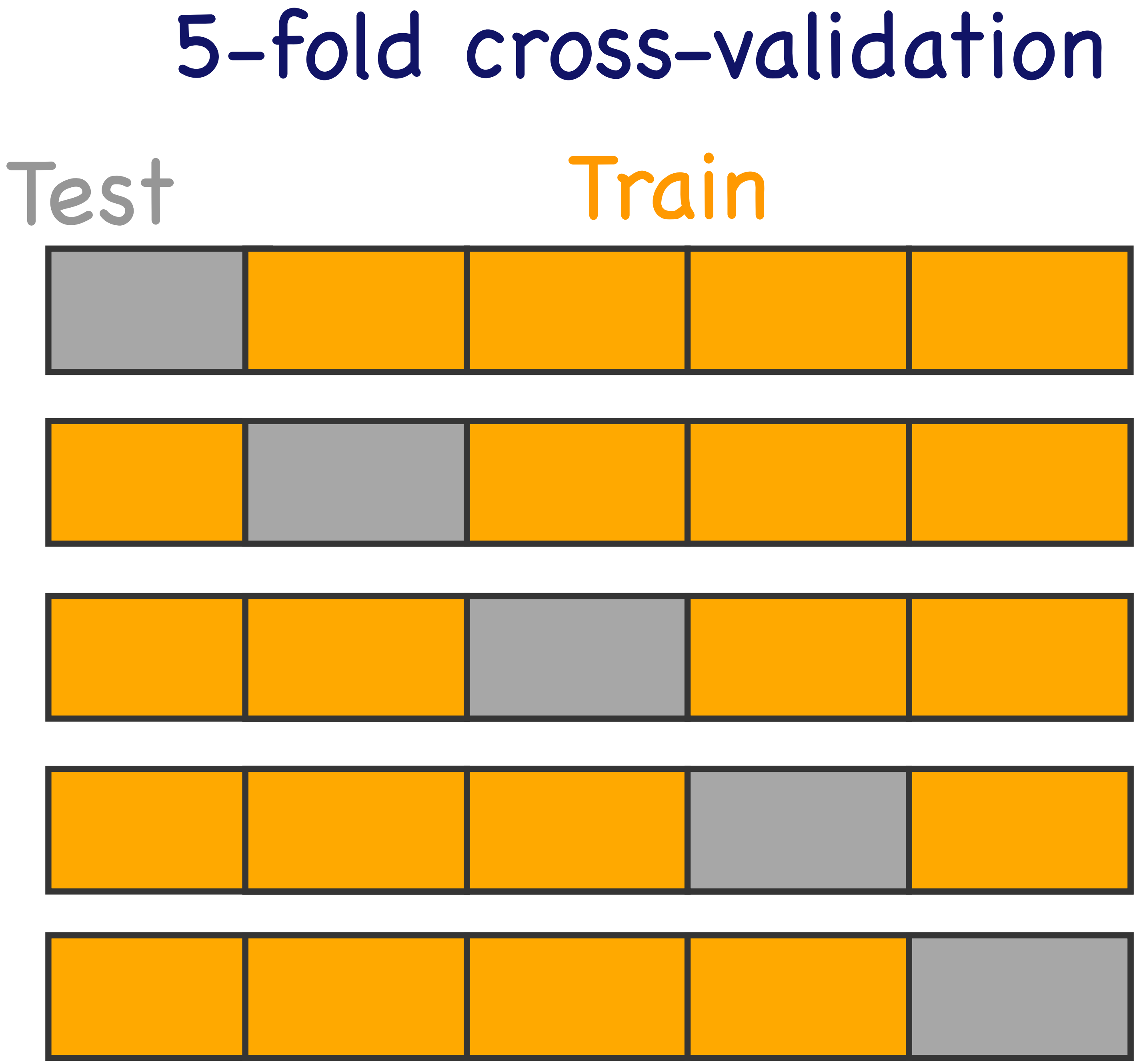


Ways to avoid overfitting

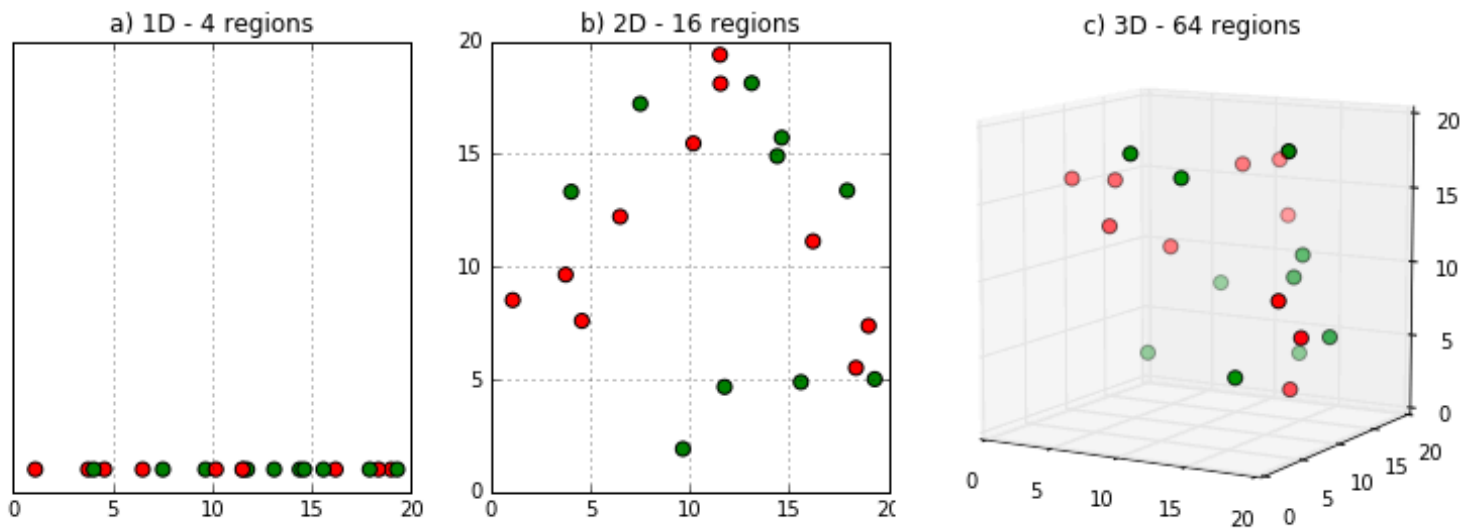
- More data
- Adding regularization term to the evaluation func.
- Cross-validation






5-fold cross-validation
Test
Train

5. INTUITION FAILS IN HIGH DIMENSIONS

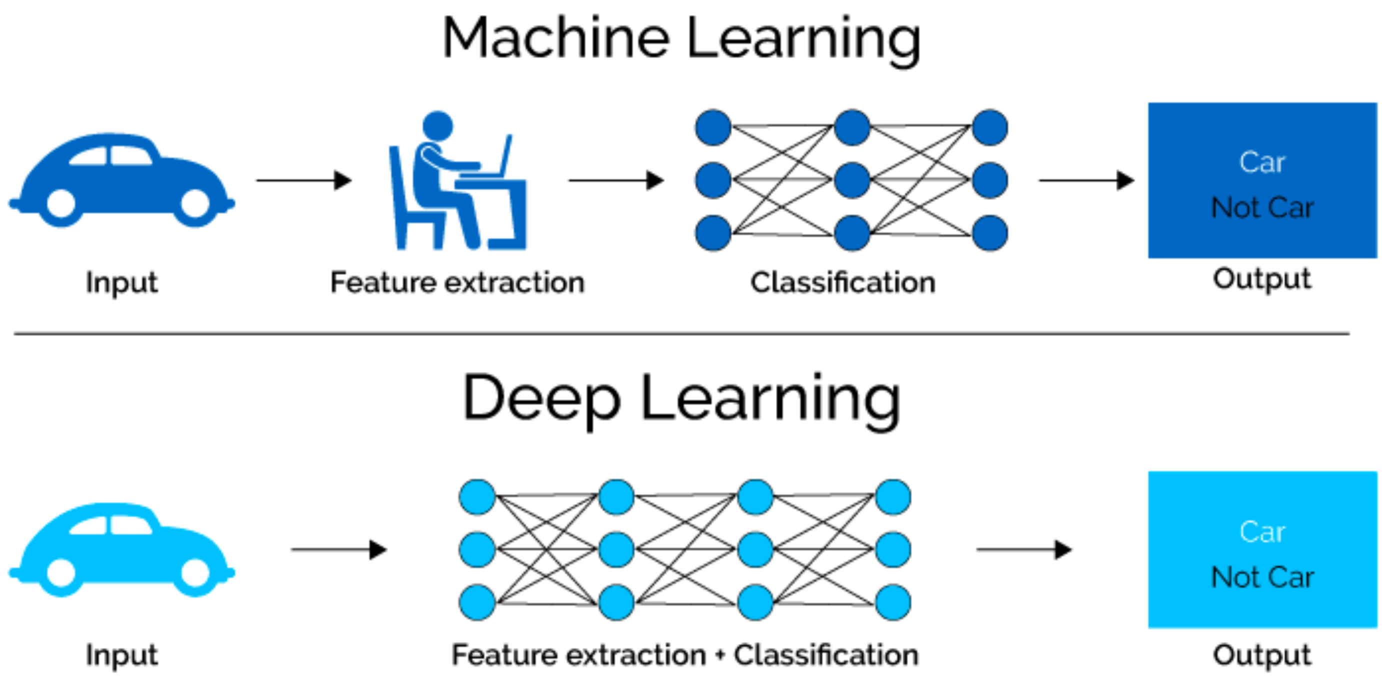


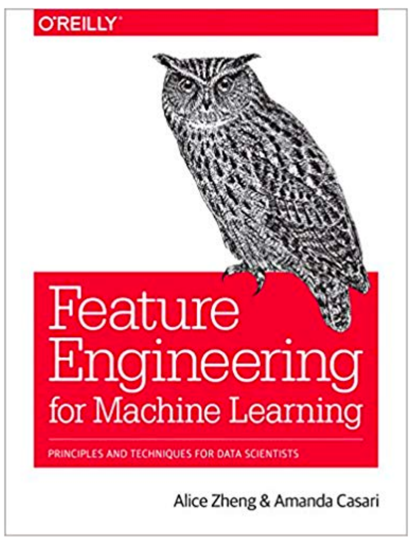
http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/

6. THEORETICAL GUARANTEES ARE NOT WHAT THEY SEEM

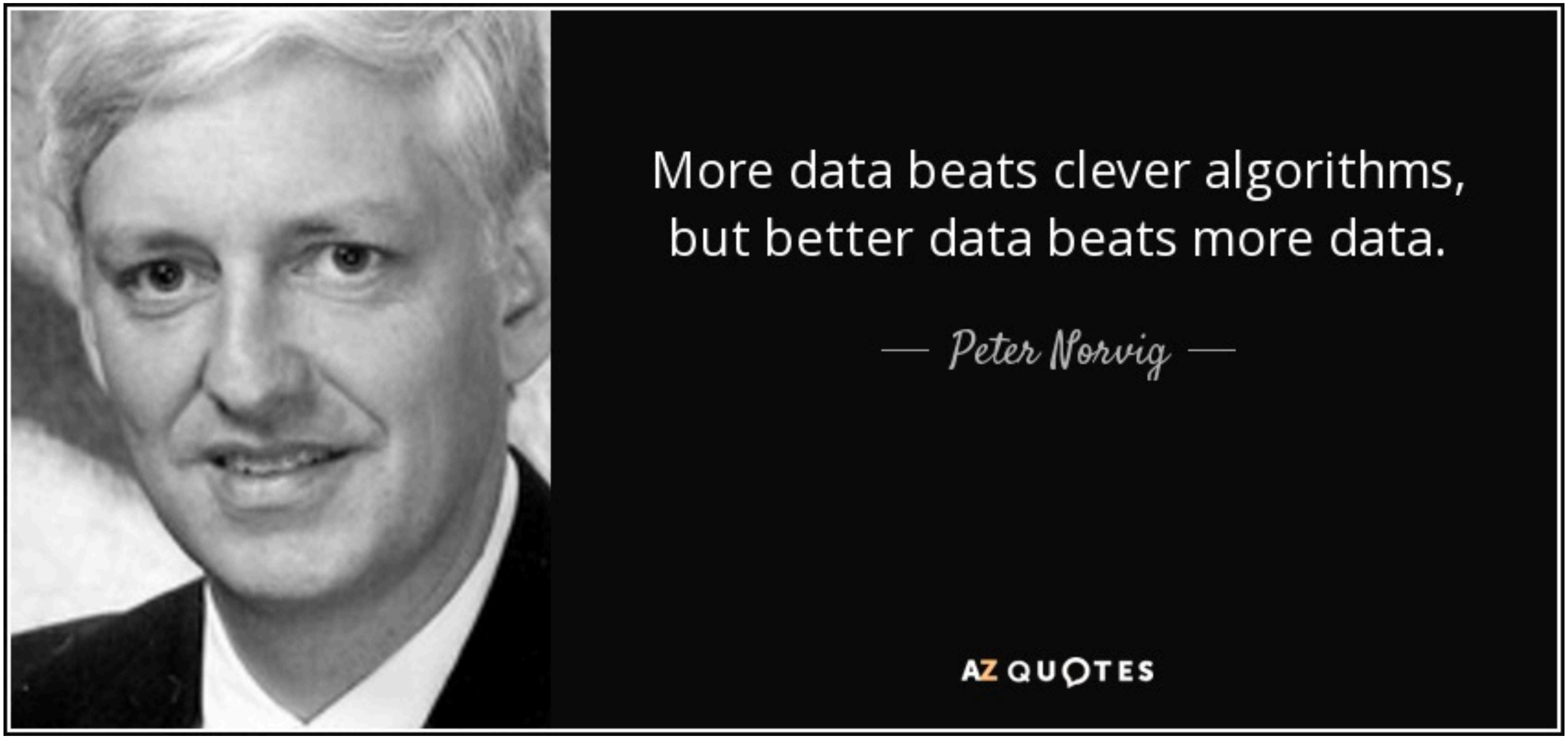
The main role of theoretical guarantees in machine learning is not as a criterion for practical decisions, but as a source of understanding and driving force for algorithm design.

7. FEATURE ENGINEERING IS THE KEY

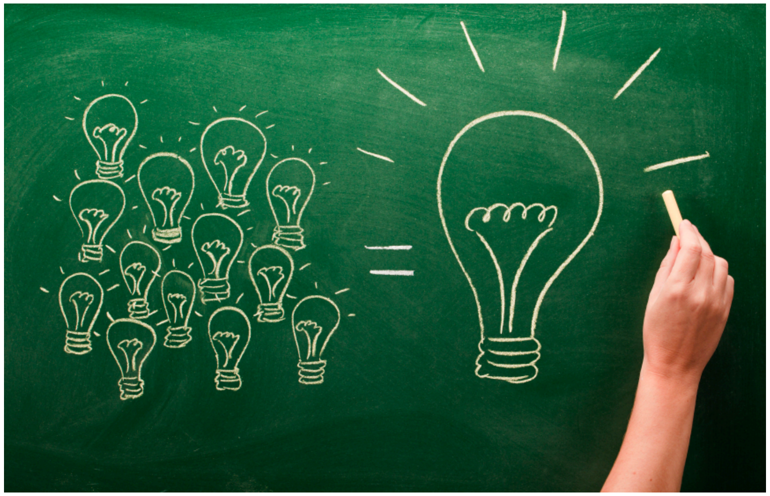


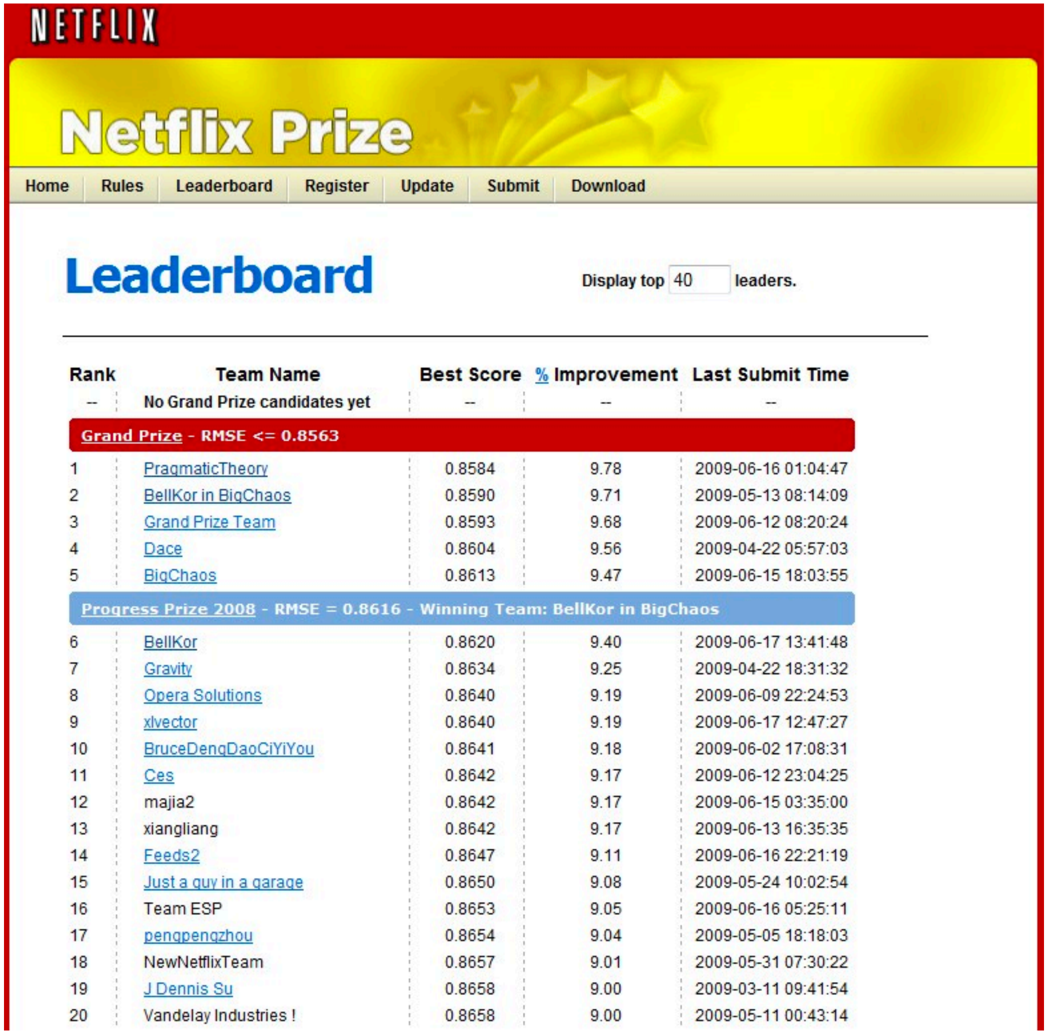


8. MORE DATA BEATS A CLEVERER ALGORITHM



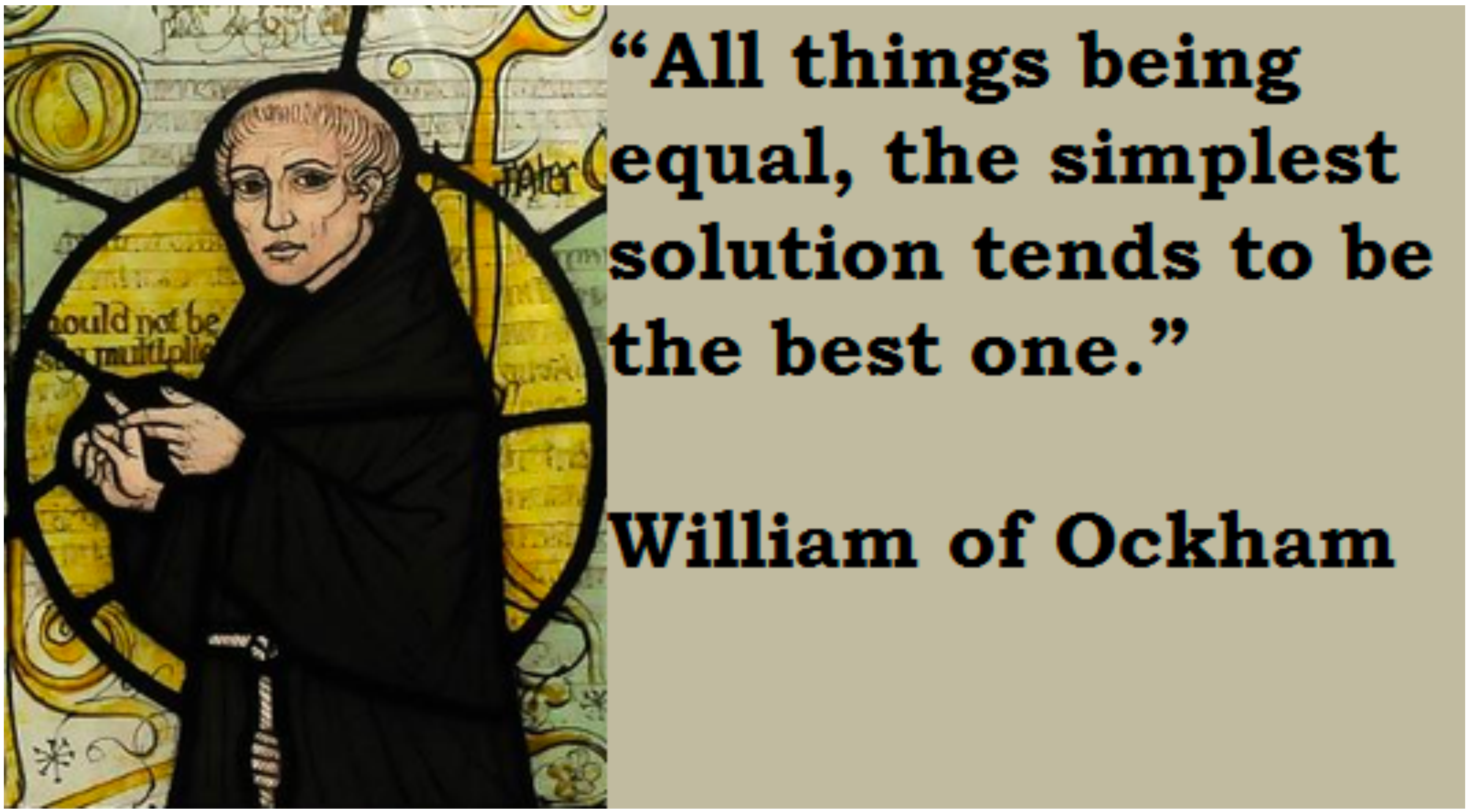
9. LEARN MANY MODELS, NOT JUST ONE





Rank	Team Name	Best Score	% Improvement	Last Submit Time
--	No Grand Prize candidates yet	--	--	--
Grand Prize - RMSE <= 0.8563				
1	PragmaticTheory	0.8584	9.78	2009-06-16 01:04:47
2	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
3	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
4	Dace	0.8604	9.56	2009-04-22 05:57:03
5	BigChaos	0.8613	9.47	2009-06-15 18:03:55
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
6	BellKor	0.8620	9.40	2009-06-17 13:41:48
7	Gravity	0.8634	9.25	2009-04-22 18:31:32
8	Opera Solutions	0.8640	9.19	2009-06-09 22:24:53
9	xlvector	0.8640	9.19	2009-06-17 12:47:27
10	BruceDengDaoCiYiYou	0.8641	9.18	2009-06-02 17:08:31
11	Ces	0.8642	9.17	2009-06-12 23:04:25
12	majia2	0.8642	9.17	2009-06-15 03:35:00
13	xiangliang	0.8642	9.17	2009-06-13 16:35:35
14	Feeds2	0.8647	9.11	2009-06-16 22:21:19
15	Just a guy in a garage	0.8650	9.08	2009-05-24 10:02:54
16	Team ESP	0.8653	9.05	2009-06-16 05:25:11
17	pengpengzhou	0.8654	9.04	2009-05-05 18:18:03
18	NewNetflixTeam	0.8657	9.01	2009-05-31 07:30:22
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	Vandelay Industries !	0.8658	9.00	2009-05-11 00:43:14

10. SIMPLICITY DOES NOT IMPLY ACCURACY

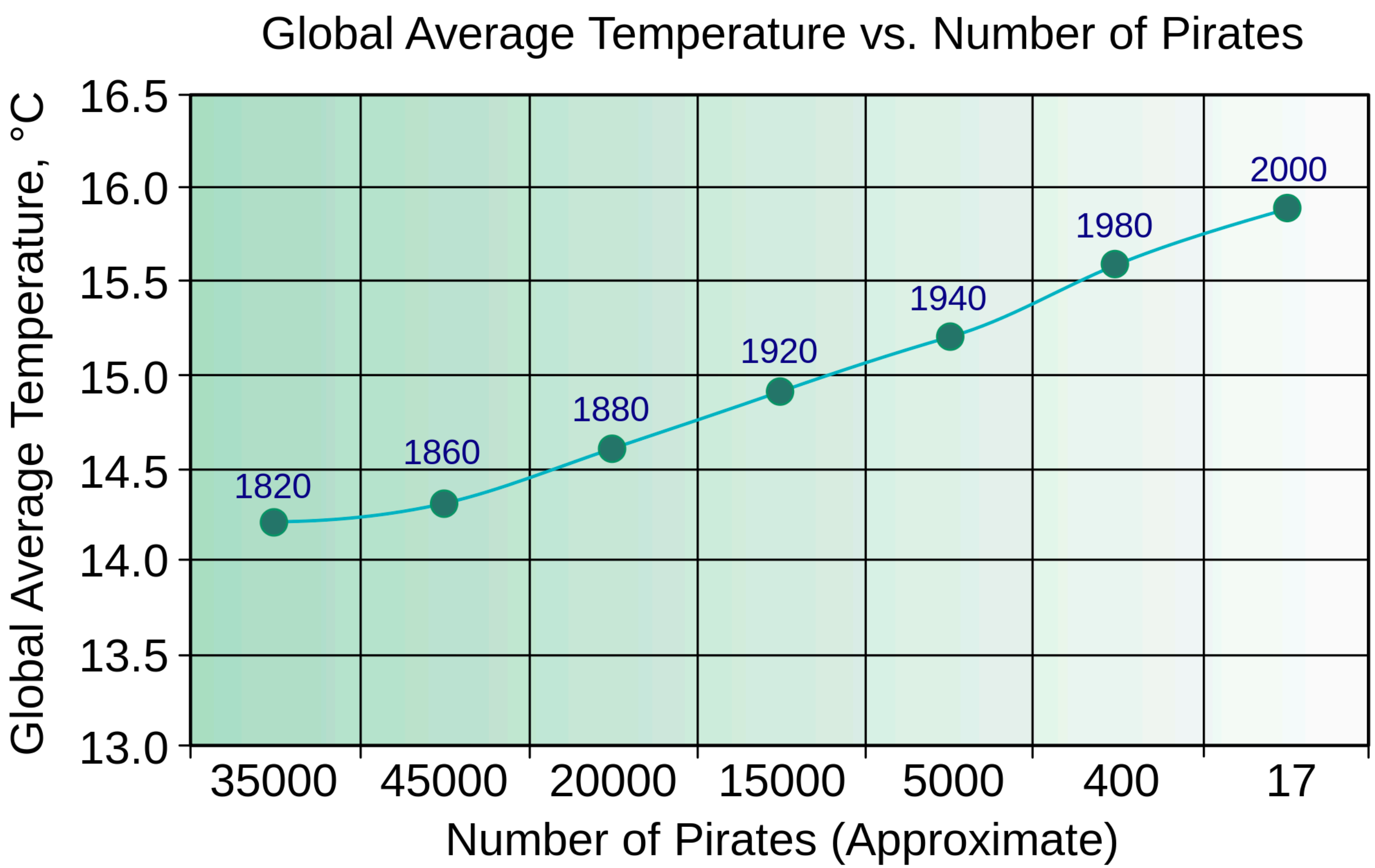


11. REPRESENTABLE DOES NOT IMPLY LEARNABLE

- "... just because a function can be represented does not mean it can be learned …"
  - Lack of data
  - Representation is too large
  - Model availability given other requirements
  - Continuous vs. discrete functions

- Therefore, the key question is not "Can it be represented?", to which the answer is often trivial, but 'Can it be learned'

12. CORRELATION DOES NOT IMPLY CAUSATION



Occam's Razor

The cyclic multiverse has multiple branes - each a universe - that collided, causing Big Bangs. The universes bounce back and pass through time, until they are pulled back together and again collide, destroying the old contents and creating them anew.

God did it.