Fundamentals of machine learning

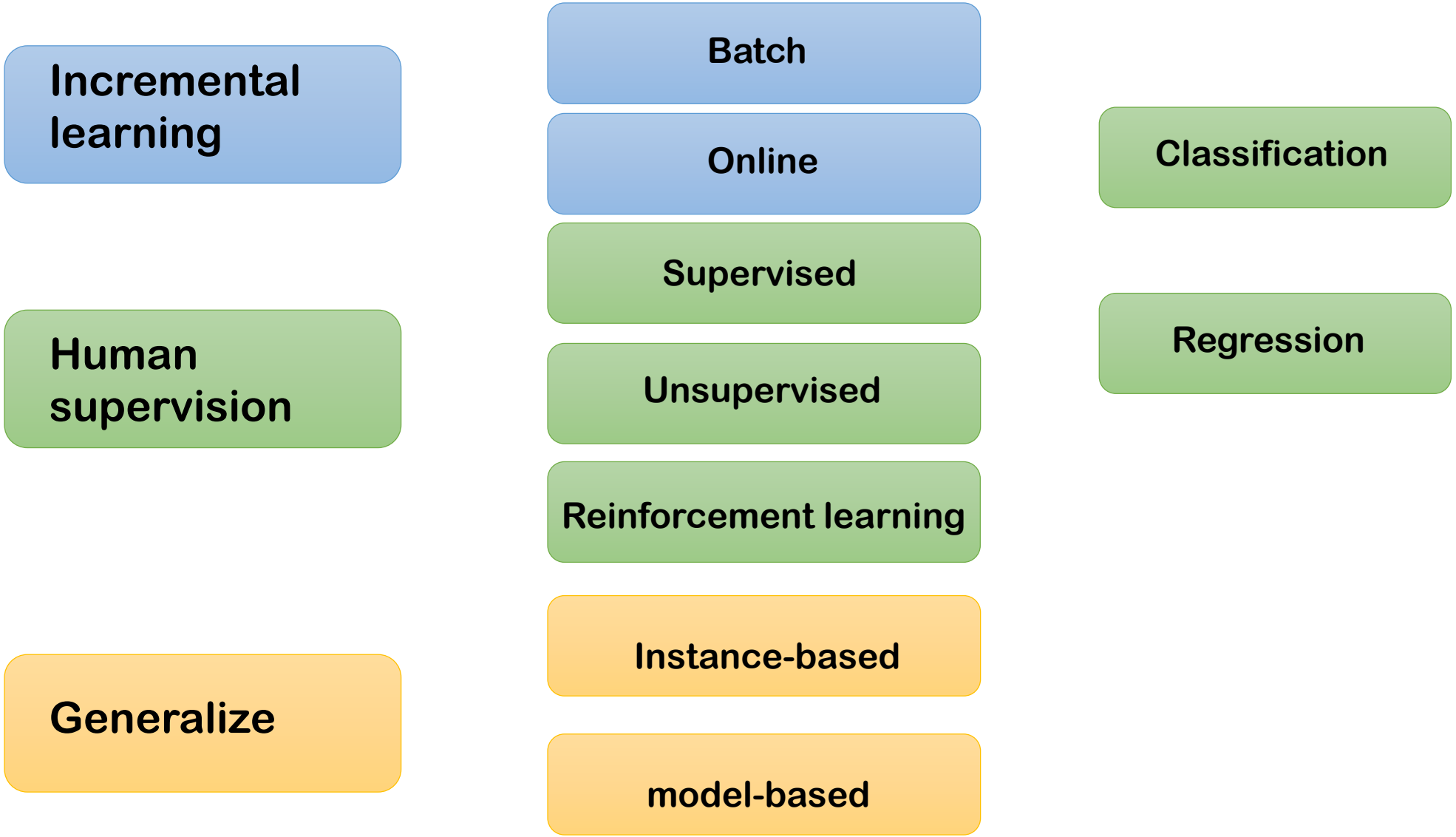
APSC 8280: Machine learning applied to plant science

03/24/2020

Outline

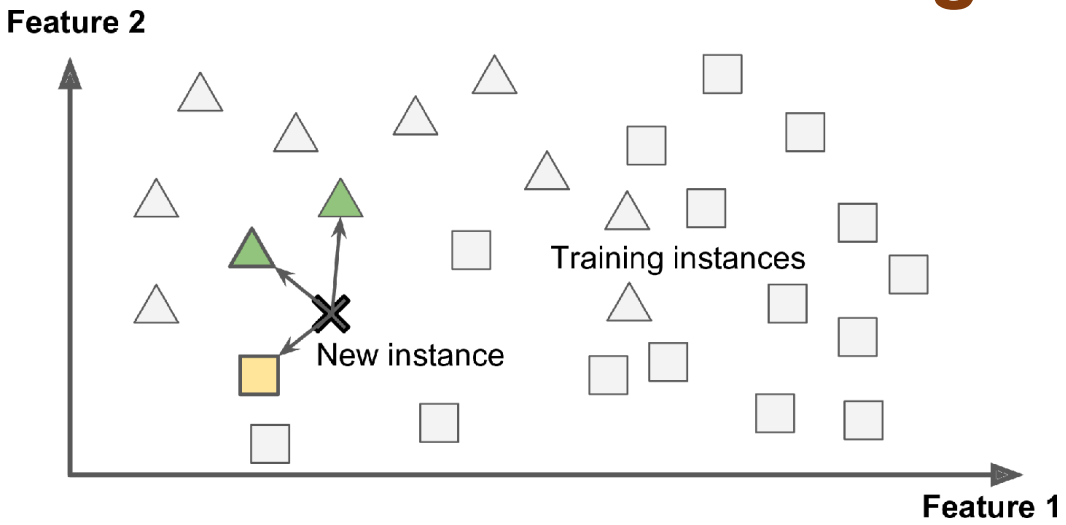
- Review expectations
- Laptop setup and basic R programming
- Types of machine learning systems
- Supervised machine learning
- Main challenges of machine learning
- Evaluation of machine learning systems

Types of machine learning systems

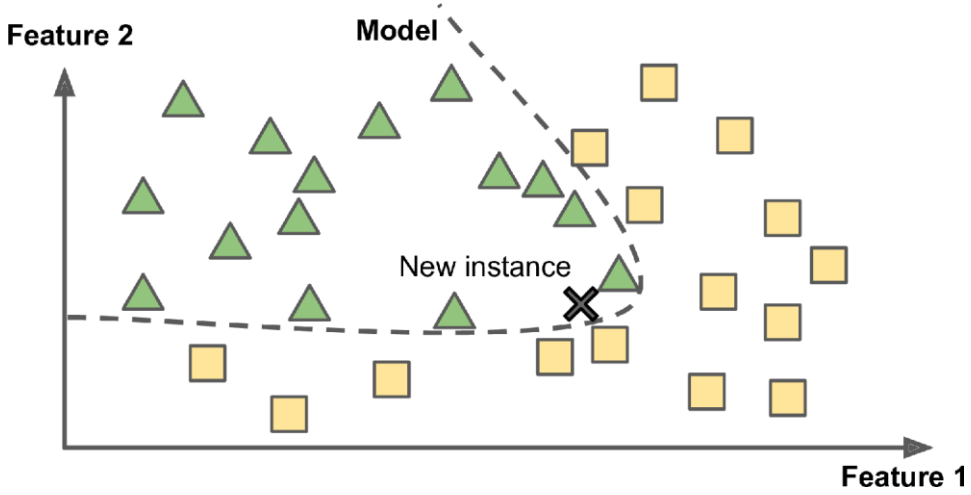


Types of machine learning systems

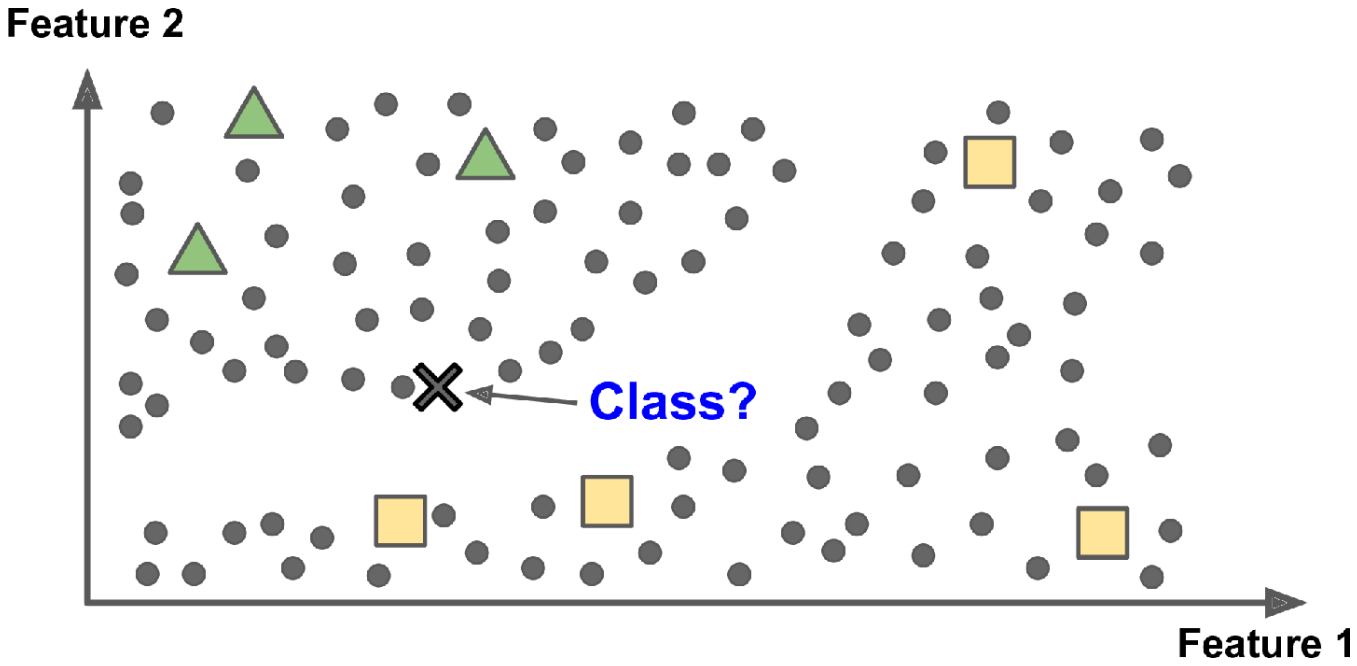
Instance-based learning



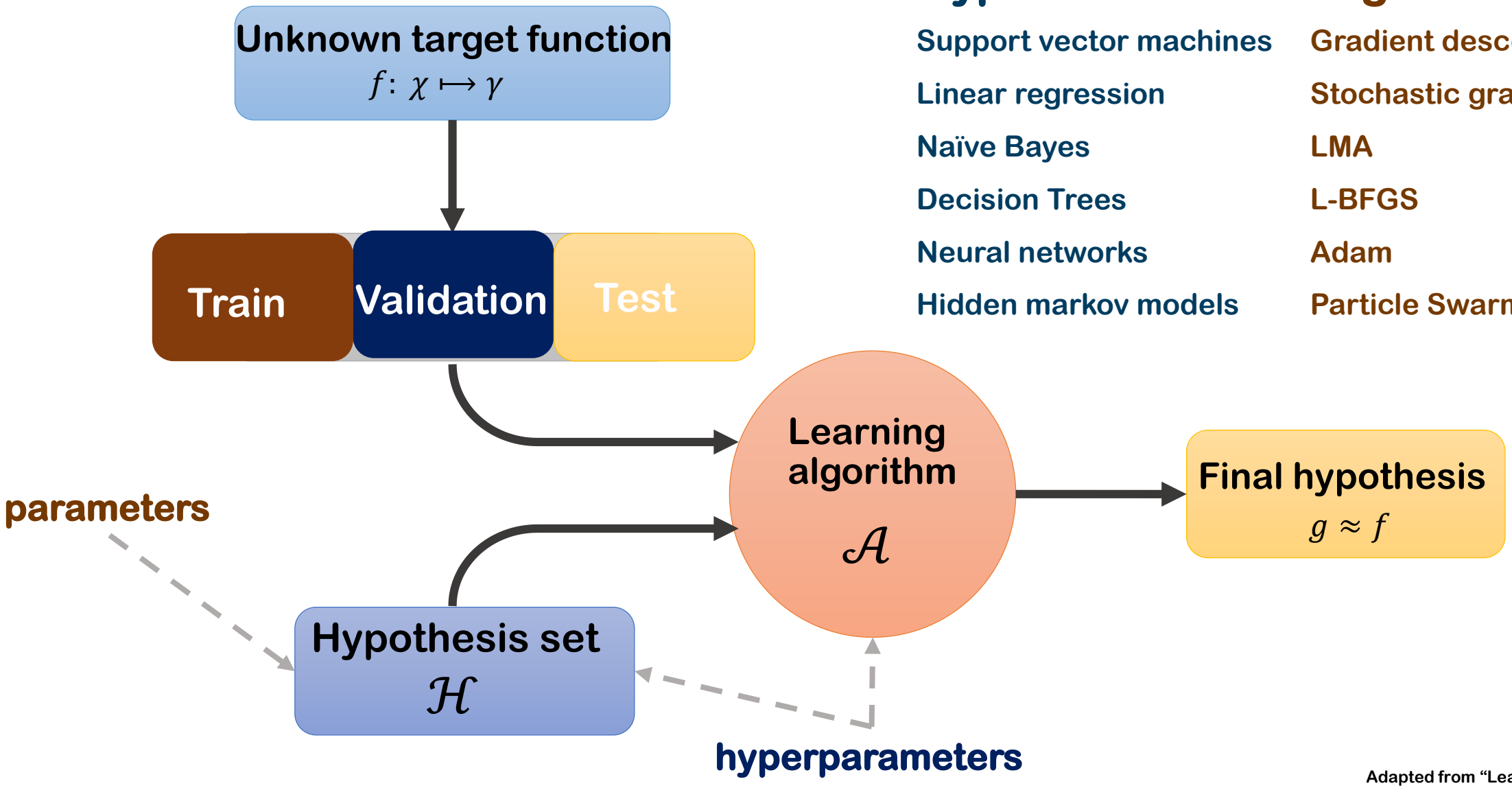
Model-based learning



Semi supervised learning



Supervised learning



Hypotheses

- Support vector machines
- Linear regression
- Naïve Bayes
- Decision Trees
- Neural networks
- Hidden markov models

Algorithms

- Gradient descent
- Stochastic gradient descent
- LMA
- L-BFGS
- Adam
- Particle Swarm Optimization

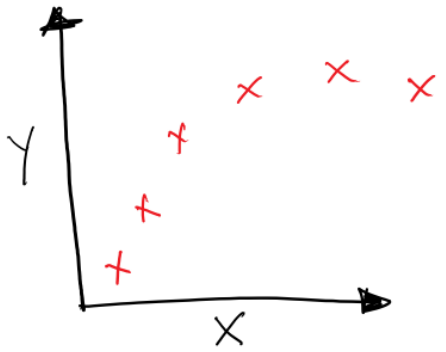
Adapted from "Learning from Data"

Main challenges of machine learning

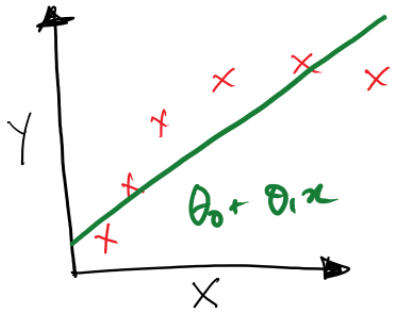
- Insufficient quantity of training data
- Nonrepresentative training data
- Poor quality data
- Irrelevant features
- Overfitting
- Underfitting

Bias variance tradeoff

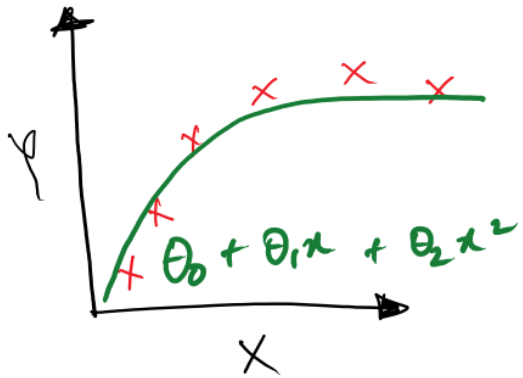
data



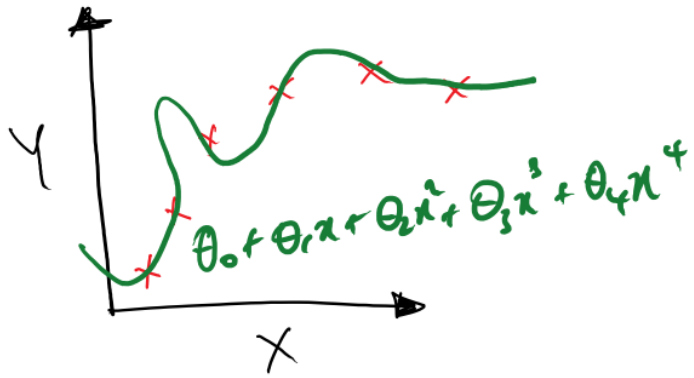
High bias (underfit)



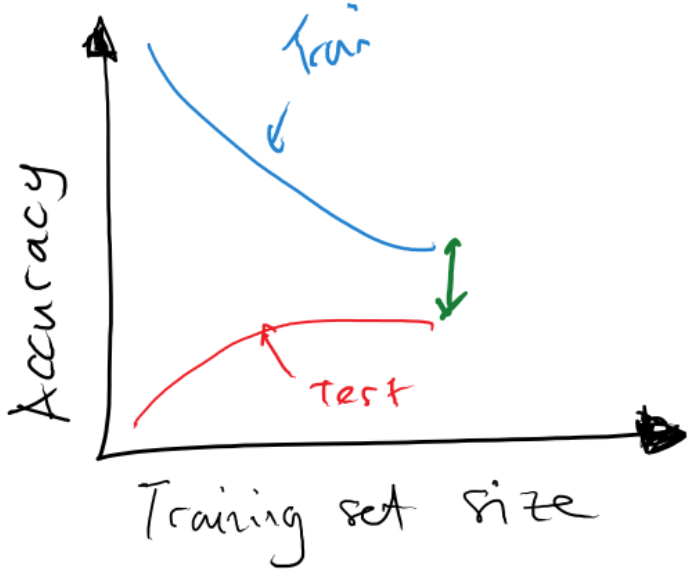
'just right'



High variance (overfit)



Learning curves

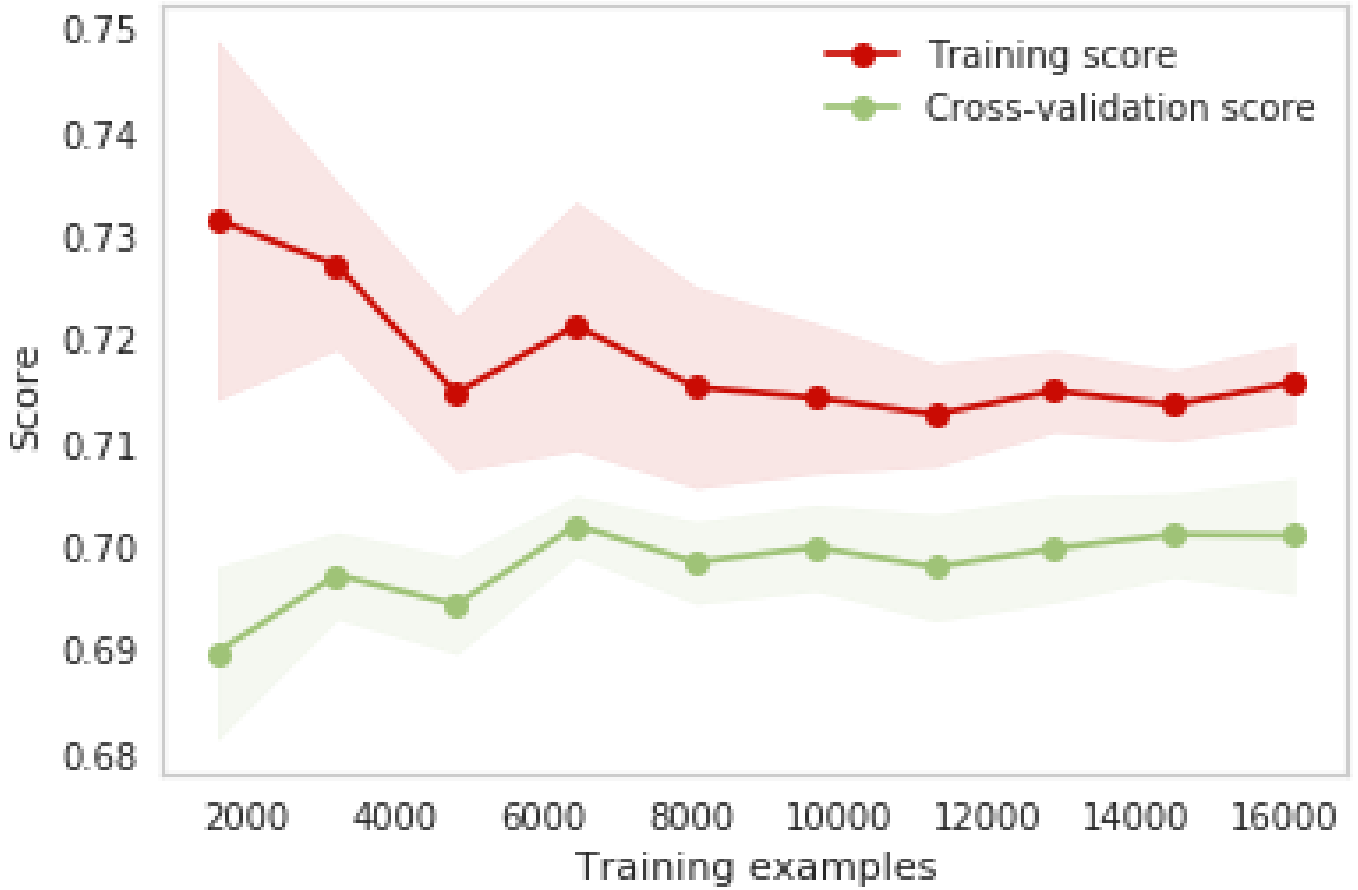


Bias variance tradeoff

High bias (underfit)

High variance (overfit)

'real life' example

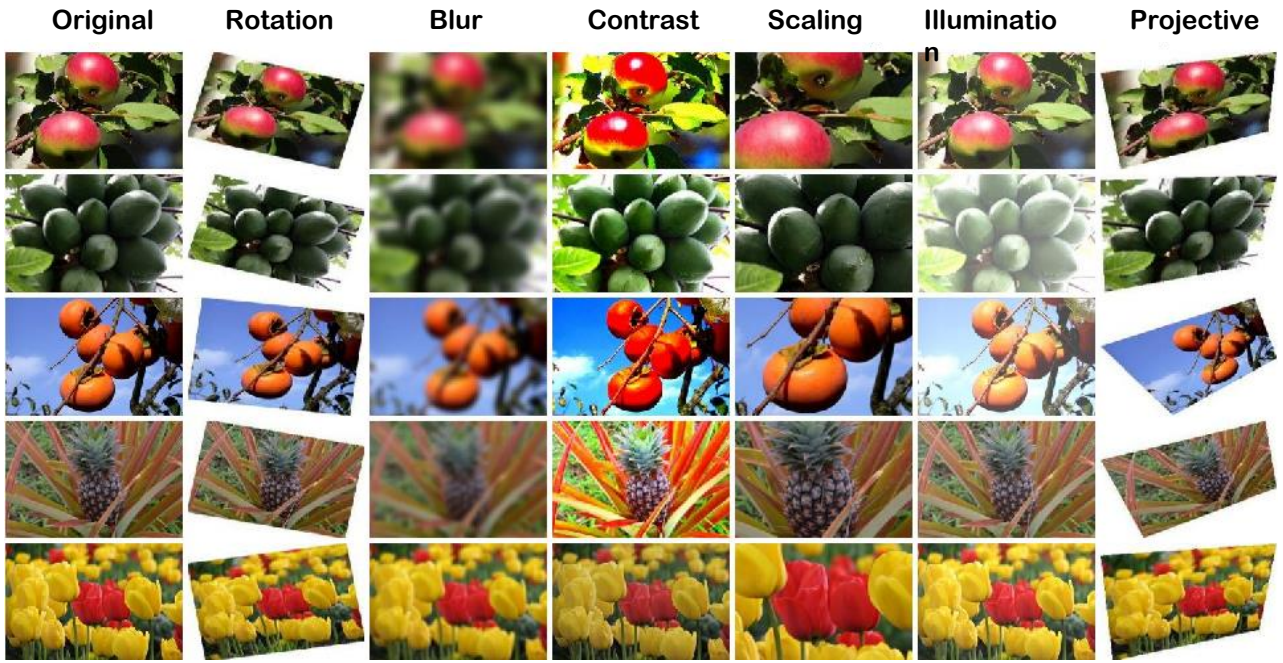


When machine learning fails…

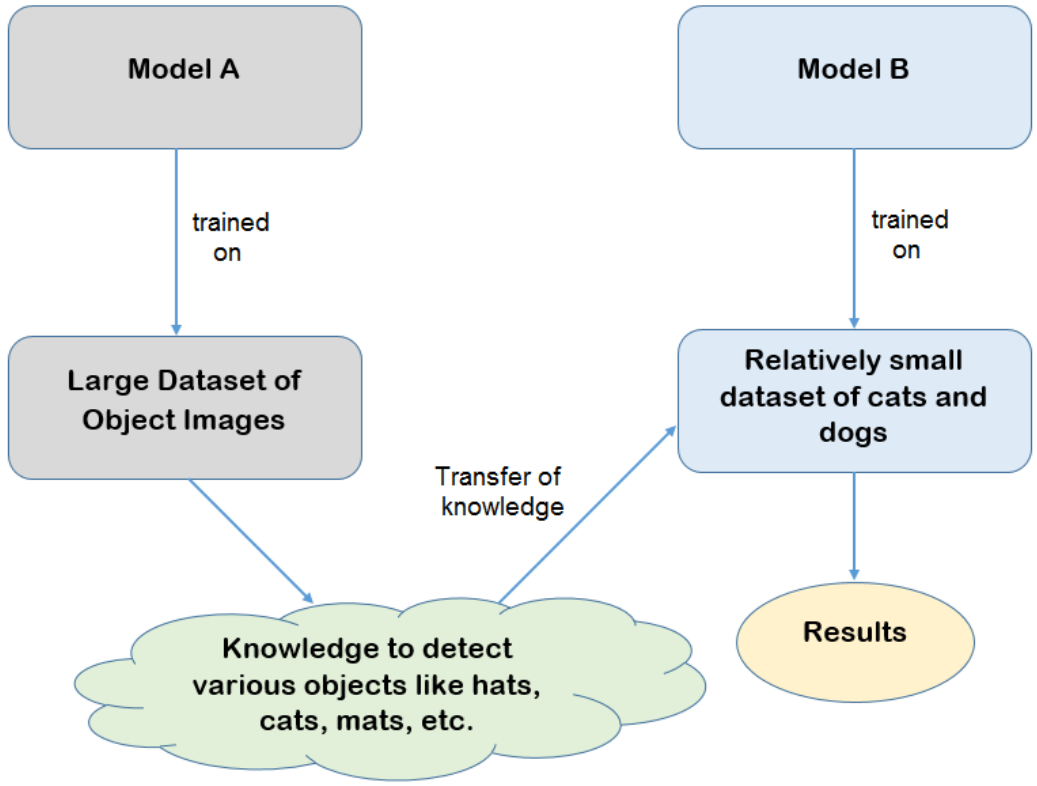
More training examples	high variance
Adding more polynomial features	high bias
Decreasing regularization	high bias
Try a smaller set of features	high variance
Increasing regularization	high variance
Try getting additional features	high bias

Dealing with small data

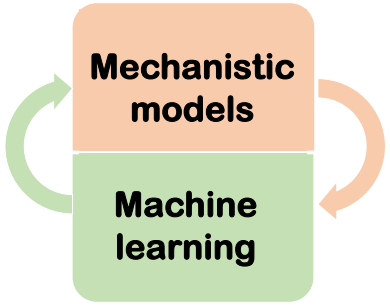
Data augmentation



Transfer learning



Integrating mechanistic information



Pawara, Pornntiwa, et al. "Data augmentation for plant classification." International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, Cham, 2017.
R Machine Learning Projects