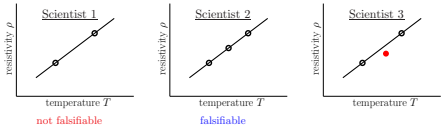
Learning From Data
Lecture 15
Reflecting on Our Path - Epilogue to Part I

M. Magdon-Ismail
CSCI 4100/6100

RECAP: ## Three Learning Principles

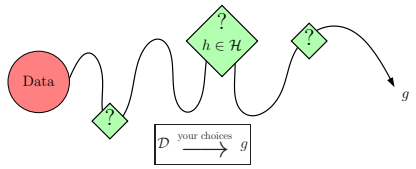
Occam's razor: simpler is better; falsifiable.



Sampling bias: ensure that training and test distributions are the same, or else acknowledge/account for it. You cannot sample from one bin and use your estimates for another bin.

Data snooping: you are charged for *every choice* influenced by $\mathcal{D}$. Choose the learning process (usually $\mathcal{H}$) before looking at $\mathcal{D}$.

We know the price of choosing g from $\mathcal{H}$.



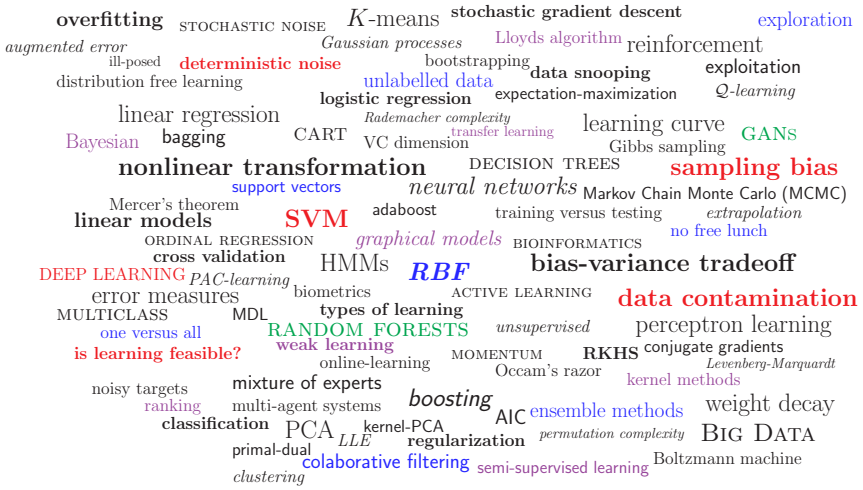
Zen Moment

Our Plan

1. What is Learning?
   Output $g \approx f$ after looking at data $(\mathbf{x}_n, y_n)$.
2. Can We do it?
   $E_{\text{in}} \approx E_{\text{out}}$ simple $\mathcal{H}$, finite d_{VC}, large N
   $E_{\text{in}} \approx 0$ good $\mathcal{H}$, algorithms
3. How to do it?
   Linear models, nonlinear transforms
   Algorithms: PLA, pseudoinverse, gradient descent
4. How to do it well?
   Overfitting: stochastic & deterministic noise
   Cures: regularization, validation.
5. General principles?
   Occams razor, sampling bias, data snooping
6. Advanced techniques.
7. Other Learning Paradigms.

concepts
theory
practice

Learning From Data: It's A Jungle Out There



Navigating the Jungle: Theory

THEORY

VC-analysis
bias-variance
complexity
Bayesian
Rademacher
SRM
⋮

Navigating the Jungle: Techniques

THEORY

VC-analysis
bias-variance
complexity
Bayesian
Rademacher
SRM
⋮

TECHNIQUES

Models

Methods

Navigating the Jungle: Models

THEORY

VC-analysis
bias-variance
complexity
Bayesian
Rademacher
SRM
⋮

TECHNIQUES

Models

linear
neural networks
SVM
similarity
Gaussian processes
graphical models
bilinear/SVD
⋮

Methods

Navigating the Jungle: Methods

THEORY

VC-analysis
bias-variance
complexity
Bayesian
Rademacher
SRM
⋮

TECHNIQUES

Models

linear
neural networks
SVM
similarity
Gaussian processes
graphical models
bilinear/SVD
⋮

Methods

regularization
validation
aggregation
preprocessing
⋮

Navigating the Jungle: Paradigms

THEORY

VC-analysis
bias-variance
complexity
Bayesian
Rademacher
SRM
⋮

TECHNIQUES

Models

linear
neural networks
SVM
similarity
Gaussian processes
graphical models
bilinear/SVD
⋮

Methods

regularization
validation
aggregation
preprocessing
⋮

PARADIGMS

supervised
unsupervised
reinforcement
active
online
unlabeled
transfer learning
big data
⋮

Moving Forward

1. What is Learning?
   Output $g \approx f$ after looking at data $(\mathbf{x}_n, y_n)$.

2. Can We do it?
   $E_{\text{in}} \approx E_{\text{out}}$ simple $\mathcal{H}$, finite d_{VC}, large N
   $E_{\text{in}} \approx 0$ good $\mathcal{H}$, algorithms

3. How to do it?
   Linear models, nonlinear transforms
   Algorithms: PLA, pseudoinverse, gradient descent

4. How to do it well?
   Overfitting: stochastic & deterministic noise
   Cures: regularization, validation.

5. General principles?
   Occams razor, sampling bias, data snooping

6. Advanced techniques.
   Similarity, neural networks, SVMs, preprocessing & aggregation

7. Other Learning Paradigms.
   Unsupervised, reinforcement

concepts
theory
practice