Title

David Rosenberg

January 14, 2018

Introduction

What we'll be thinking about for each ML Algorithm

- what's the hypothesis space?
- what's our objective function?
- what's our loss function?
- what's the learning algorithm? (i.e. how use $\mathcal{D}$ to choose f from $\mathcal{F}$?)
    - most often framed as an optimization problem, but not always
- how does it compare to and relate to other ML algorithms

Generalization

- Algorithm works well on new data... data not from training data

Sources of Prediction Error

- Irreducible error (in the true distribution)
  - error of the Bayes prediction funtion
  - labeling error (can think of as part of the response distribution)
  - (even Bayes prediction function will have error)
- Hypothesis space deficiency
  - Bayes prediction function not contained in hypothesis space
  - missing a feature / variable
- Estimation Error
- Search / Optimization error

Evaluation Metrics

Eval Matric

- give example of evaluation challenges
- spam classification / medical diagnosis
- ranking challenge
- Precision / Recall
- F1 - Harmonic mean
- ROC - AUC
- Lift Curves

Confidence Intervals and Test Set Size

- asdf a

Model Selection

Comparing models: Error bars and Statistical Testing

- asdf

Issue of Multiple Hypothesis Testing

- Reference to jupyter notebook / writeup
- give 1 sd heuristic (simplest within 1 SD)

Hypothesis Spaces

Goal is to Learn the Building Blocks

- Large majority of our machine learning algorithms take the following form
- FIND f in HYPOTHESIS SPACE $\mathcal{H}$, subject to constraint $\psi(f) \leqslant \lambda$, for REGULARIZATION PARAMETER λ, that minimizes the average training LOSS
- To solve the resulting minimization problem, need an OPTIMIZATION METHOD
- objective
- We'll consider each of these pieces:
- Week 2: regularization (l1, l2, elastic net)
- Week 3: loss functions for classification and regression (l1, l2, huber, hinge, logistic, etc)
- Week 4: hypothesis space: kernel methods (kind of nonlinear)
- Week 5: hypothesis space: neural networks (nonlinear)
- Week 6: hypothesis space: trees and ensembles (nonlinear)
- Week 7: loss functions / hypothesis space: multiclass & intro to structured prediction
- Week 8-14: Probabilistic modeling (negative log-likelihood loss function)

Example deconstructions into building blocks

Name	Loss	Regularization	Hypothesis space	Op
Linear regression	Square loss	None	Linear	
Ridge regression	Square	ℓ_2	Linear	
Lasso regression	Square	ℓ_1	Linear	
Logistic regression	Logistic	None	Linear	
SVM classification	Hinge	ℓ_2	Linear	
Adaboost	Exponential		Span(Base hypothesis space)	Fo
Neural network (MLP)	Any differentiable	Any	Neural network	