10-601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Learning Theory

Learning Theory Readings:
Murphy --
Bishop --
HTF --
Mitchell 7

Matt Gormley
Lecture 13
March 1, 2016

Reminders

- **Homework 4: Perceptron / Kernels / SVM**
  - **Release: Wed, Feb. 22**
  - **Due: Fri, Mar. 03 at 11:59pm**

9 days
for HW4

- **Midterm Exam (Evening Exam)**
  - **Tue, Mar. 07 at 7:00pm – 9:30pm**
  - **See Piazza for details about location**

Outline

- **Generative vs. Discriminative Modeling**
- **Bayes Optimal Classifier**
  - Loss function, Expected Loss
  - Bayes Error
  - Bayes Optimal
- **Consistency of MLE**
  - Consistency
  - Almost Sure Convergence of MLE

DISCRIMINATIVE AND GENERATIVE CLASSIFIERS

Generative vs. Discriminative

- **Generative Classifiers:**
  - Example: Naïve Bayes
  - Define a joint model of the observations **x** and the labels y: $p(\boldsymbol{x}, y)$
  - Learning maximizes (joint) likelihood
  - Use Bayes' Rule to classify based on the posterior:

$$p(y|\mathbf{x}) = p(\mathbf{x}|y)p(y)/p(\mathbf{x})$$

- **Discriminative Classifiers:**
  - Example: Logistic Regression
  - Directly model the conditional: $p(y|\mathbf{x})$
  - Learning maximizes conditional likelihood

Generative vs. Discriminative

Whiteboard

– Contrast: To model p(x) or not to model p(x)?

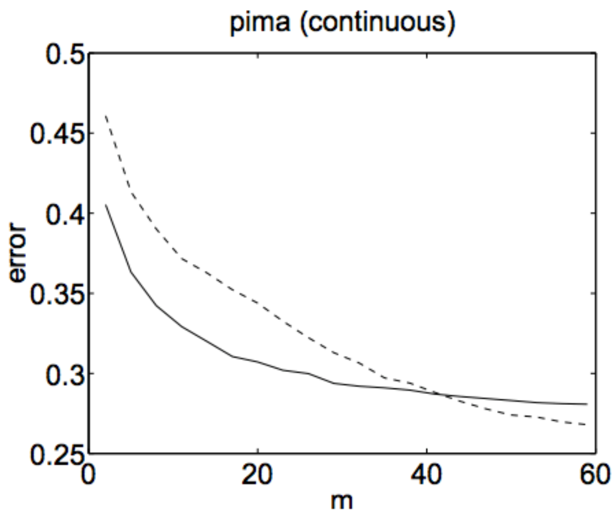
Generative vs. Discriminative

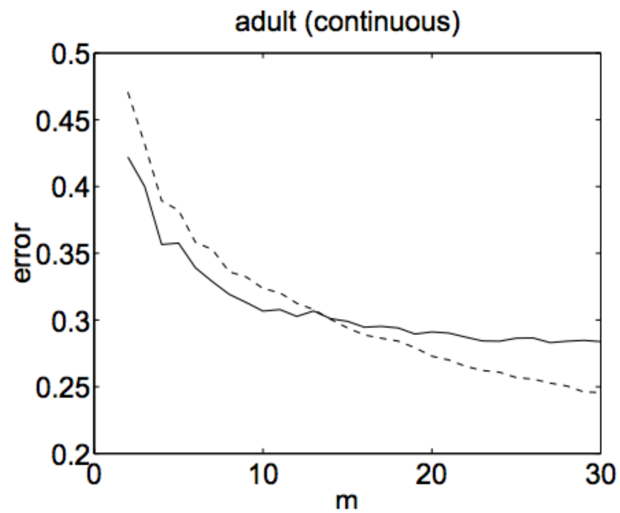
Finite Sample Analysis (Ng & Jordan, 2002)

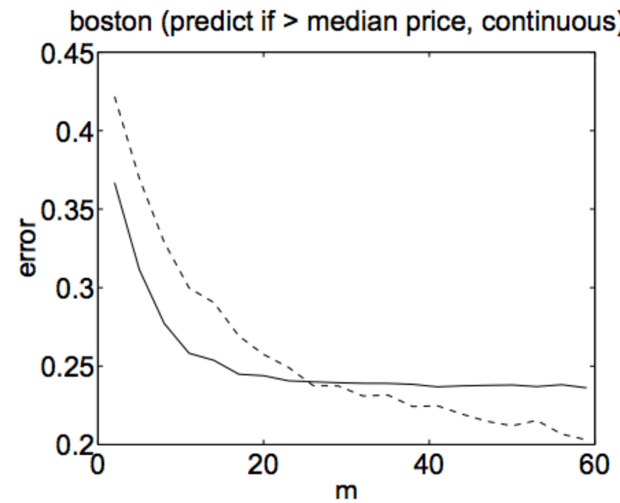
[Assume that we are learning from a finite training dataset]

If model assumptions are correct: Naive Bayes is a more efficient learner (requires fewer samples) than Logistic Regression

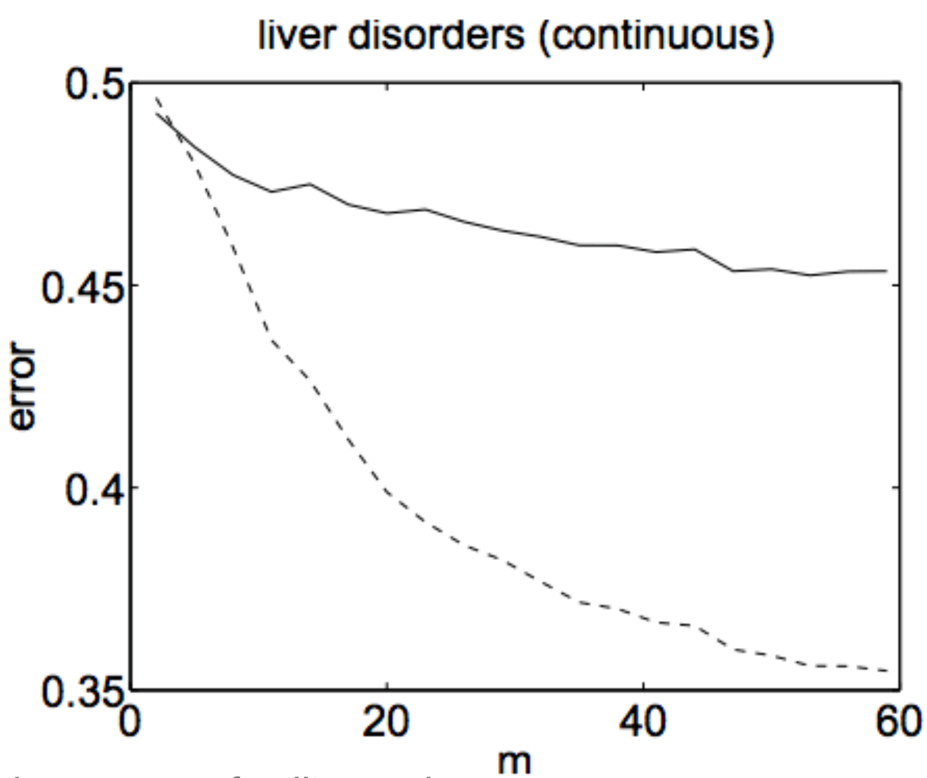
If model assumptions are incorrect: Logistic Regression has lower asymtotic error, and does better than Naïve Bayes

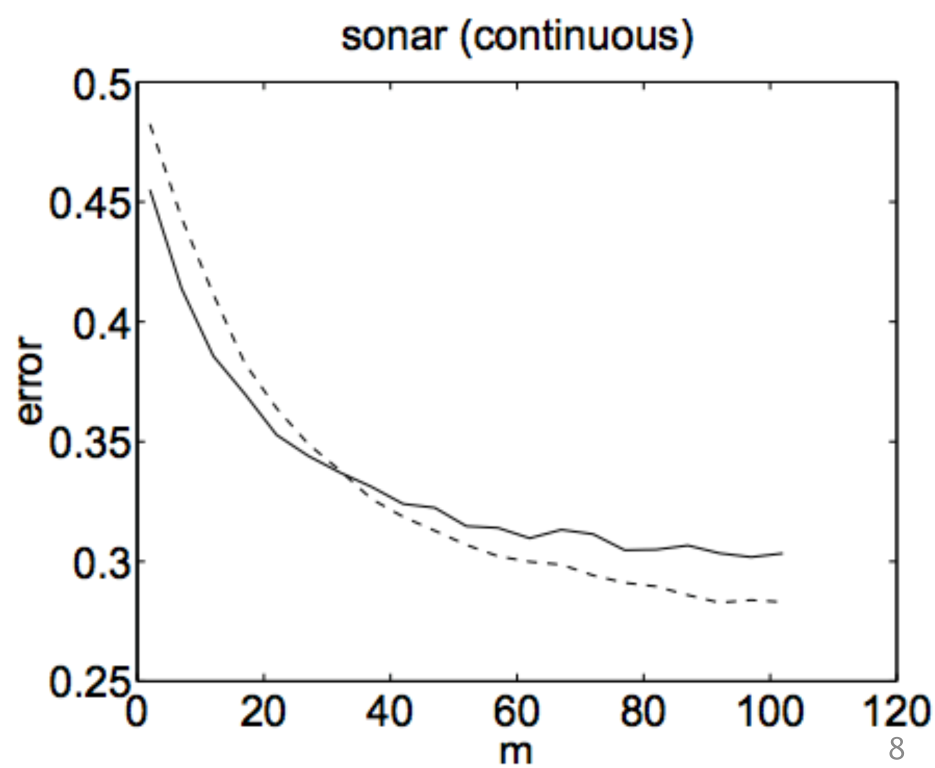






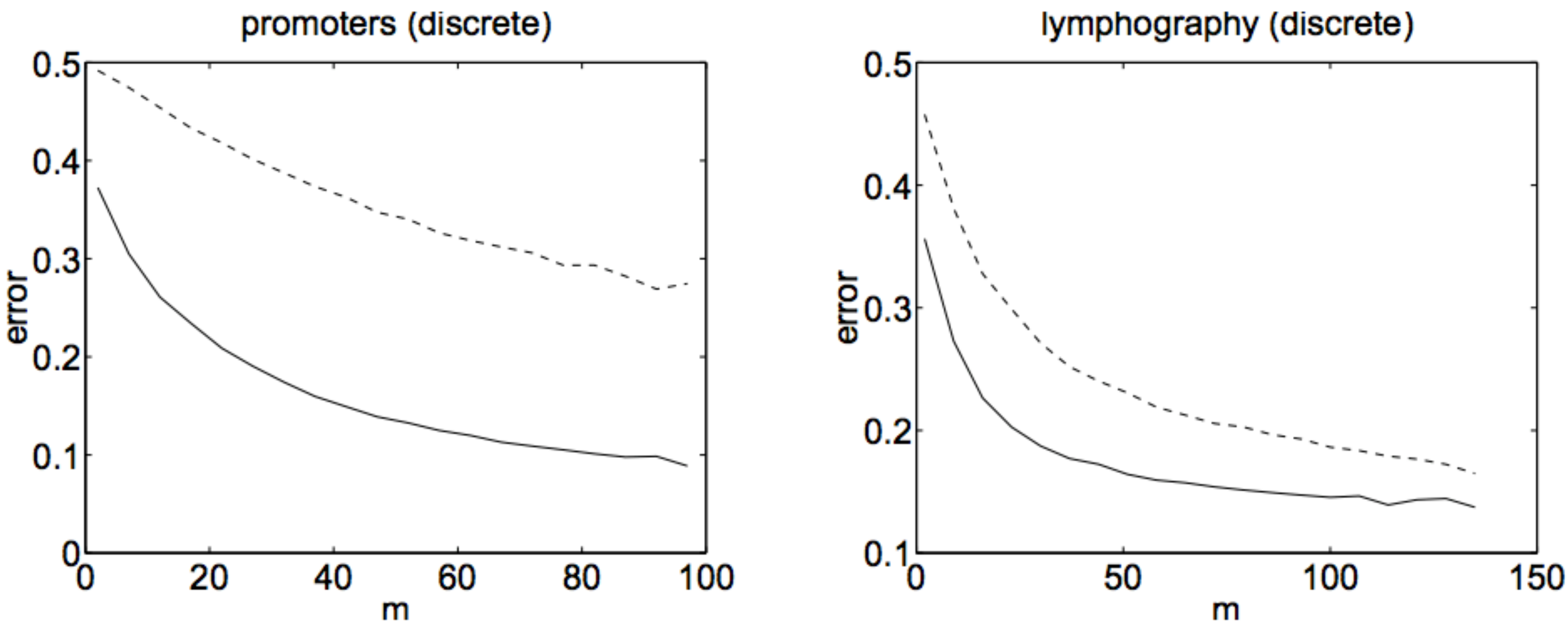
solid: NB dashed: LR





Slide courtesy of William Cohen

solid: NB dashed: LR



Naïve Bayes makes stronger assumptions about the data but needs fewer examples to estimate the parameters

"On Discriminative vs Generative Classifiers: ...." Andrew Ng and Michael Jordan, NIPS 2001.

Slide courtesy of William Cohen

Generative vs. Discriminative

Learning (Parameter Estimation)

Naïve Bayes:
Parameters are decoupled → Closed form solution for MLE

Logistic Regression:
Parameters are coupled → No closed form solution – must use iterative optimization techniques instead

Naïve Bayes vs. Logistic Reg.

Learning (MAP Estimation of Parameters)

Bernoulli Naïve Bayes:
Parameters are probabilities → Beta prior (usually) pushes probabilities away from zero / one extremes

Logistic Regression:
Parameters are not probabilities → Gaussian prior encourages parameters to be close to zero

(effectively pushes the probabilities away from zero / one extremes)

Naïve Bayes vs. Logistic Reg.

Features

Naïve Bayes:
Features $\boldsymbol{x}$ are assumed to be conditionally independent given y. (i.e. Naïve Bayes Assumption)

Logistic Regression:
No assumptions are made about the form of the features $\boldsymbol{x}$. They can be dependent and correlated in any fashion.

LEARNING THEORY

Outline

- Generative vs. Discriminative Modeling
- **Bayes Optimal Classifier**
  - Loss function, Expected Loss
  - Bayes Error
  - Bayes Optimal
- **Consistency of MLE**
  - Consistency
  - Almost Sure Convergence of MLE

Questions For Today

1. Given a probability distribution, how do we construct the optimal classifier?
   (Bayes Optimal Classifier)

2. Given data, what guarantees do we have about the MLE parameters?
   (Consistency of MLE)

Bayes Optimal Classifier

Whiteboard:

- Loss function, Expected Loss
- Bayes Error
- Bayes Optimal

Questions For Today

1. Given a probability distribution, how do we construct the optimal classifier?
   (Bayes Optimal Classifier)

2. Given data, what guarantees do we have about the MLE parameters?
   (Consistency of MLE)

Consistency of MLE

Whiteboard:

- Consistency
- Almost Sure Convergence of MLE

MIDTERM EXAM LOGISTICS

Midterm Exam

- **Logistics**
  - **Evening Exam**
    Tue, Mar. 07 at 7:00pm – 9:30pm
  - 8-9 Sections
  - Format of questions:
    - Multiple choice
    - True / False (with justification)
    - Derivations
    - Short answers
    - Interpreting figures
  - No electronic devices
  - You are allowed to **bring** one 8½ x 11 sheet of notes (front and back)

Midterm Exam

- **How to Prepare**
  - Attend the midterm review session:
    Thu, March 2 at 6:30pm (PH 100)
  - Attend the midterm review lecture
    Mon, March 6 (in-class)
  - Review prior year’s exam and solutions
    (we’ll post them)
  - Review this year’s homework problems

Midterm Exam

- **Advice (for during the exam)**
  - Solve the easy problems first
    (e.g. multiple choice before derivations)
    - if a problem seems extremely complicated you're likely missing something
  - Don't leave any answer blank!
  - If you make an assumption, write it down
  - If you look at a question and don't know the answer:
    - we probably haven't told you the answer
    - but we've told you enough to work it out
    - imagine arguing for some answer and see if you like it

Topics for Midterm

- Foundations
  - Probability
  - MLE, MAP
  - Optimization
- Classifiers
  - KNN
  - Naïve Bayes
  - Logistic Regression
  - Perceptron
  - SVM
- Regression
  - Linear Regression
- Important Concepts
  - Kernels
  - Regularization and Overfitting
  - Experimental Design