UVA CS 4501: Machine Learning

Lecture 20-Extra: Generative Classifier Vs. Discriminative Classifier

Dr. Yanjun Qi

University of Virginia

Department of
Computer Science

Discriminative vs. Generative

Generative approach

- Model the joint distribution $p(X, C)$ using

  $p(X \mid C = c_k)$ and $p(C = c_k)$

Class prior

Discriminative approach

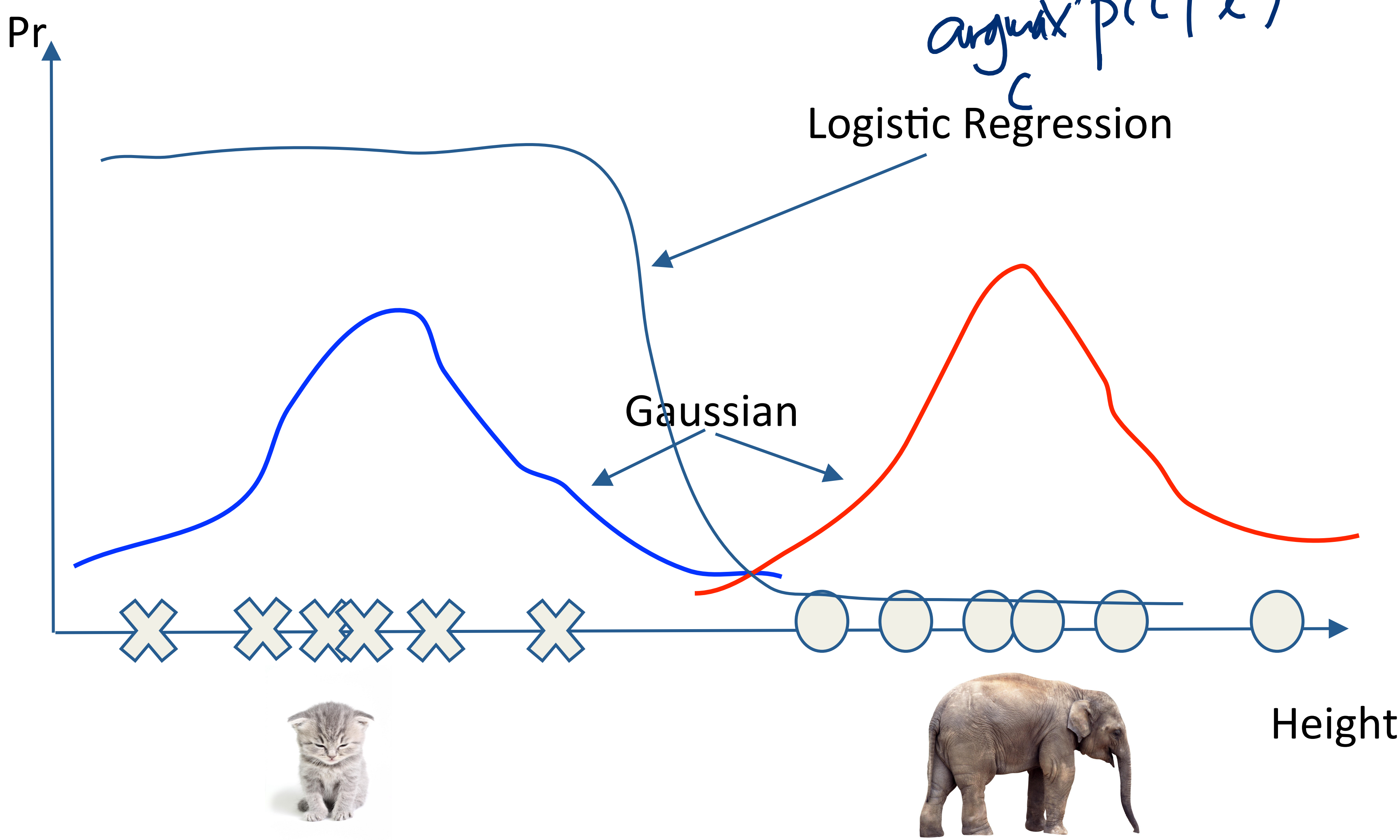
- Model the conditional distribution $p(c \mid X)$ directly

e.g.,

$$P(C=1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * X)}}$$

Discriminative vs. Generative

$\arg\max_{c} p(c|x)$



LDA vs. Logistic Regression

$\Rightarrow$ mean $Kp + p^2$ conv

- **LDA (Generative model)** $p(X_{p\times 1} | C_i)$
  - Assumes Gaussian class-conditional densities and a common covariance
  - Model parameters are estimated by maximizing the [full log likelihood,] parameters for each class are estimated independently of other classes, $Kp + \frac{p(p+1)}{2} + (K-1)$ parameters
  - Makes use of marginal density information $\Pr(x)$
  - Easier to train, low variance, more efficient if model is correct
  - Higher asymptotic error, but converges faster

$\Rightarrow (K-1)(p+1)$

- **Logistic Regression (Discriminative model)**
  - Assumes class-conditional densities are members of the (same) exponential family distribution $p(C_i|x)$
  - Model parameters are estimated by maximizing the [conditional log likelihood,] simultaneous consideration of all other classes, $(K-1)(p+1)$parameters
  - Ignores marginal density information $\Pr(x)$
  - Harder to train, robust to uncertainty about the data generation process
  - Lower asymptotic error, but converges more slowly

Discriminative vs. Generative

- Definitions
  - h_{gen} and h_{dis}: generative and discriminative classifiers
  - $h_{gen, inf}$ and $h_{dis, inf}$: same classifiers but trained on the entire population (asymptotic classifiers)
  - n → infinity, h_{gen} → $h_{gen, inf}$ and h_{dis} → $h_{dis, inf}$

Ng, Jordan,. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes." *Advances in neural information processing systems* 14 (2002): 841.

Discriminative vs. Generative

Proposition 1: vs. h_{true}

$$\epsilon\left(h_{dis,\inf}\right) \leq \epsilon\left(h_{gen,\inf}\right)$$

= assymptotic error

Proposition 1 states that aymptotically, the error of the discriminative logistic regression is smaller than that of the generative naive Bayes. This is easily shown

- p : number of dimensions

- n : number of observations

- ϵ : generalization error

Logistic Regression vs. NBC

Discriminative classifier (Logistic Regression)

- Smaller asymptotic error
- Slow convergence ~ O(p)

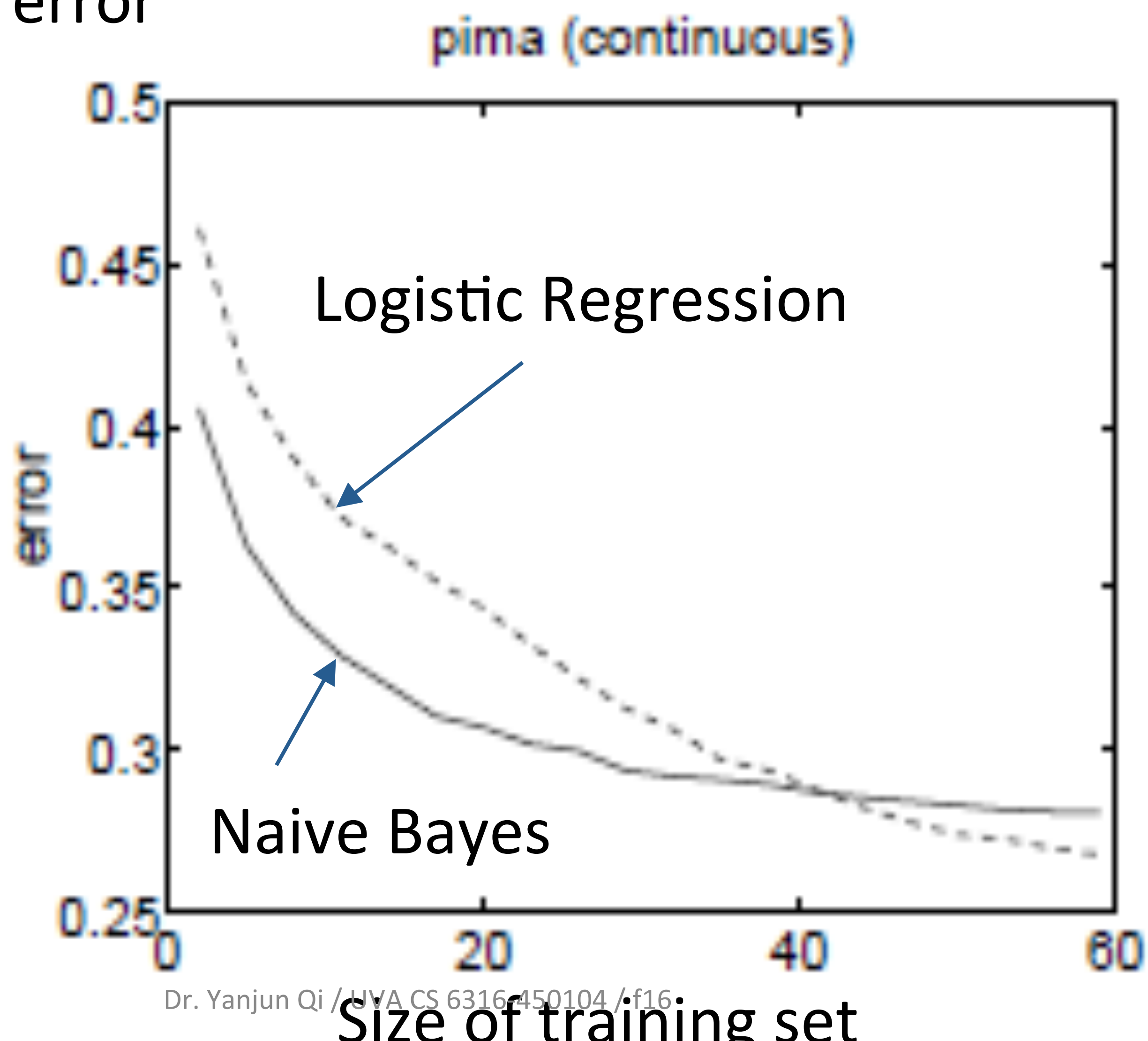
Generative classifier (Naive Bayes)

- Larger asymptotic error
- Can handle missing data (EM)
- Fast convergence ~ O(lg(p))

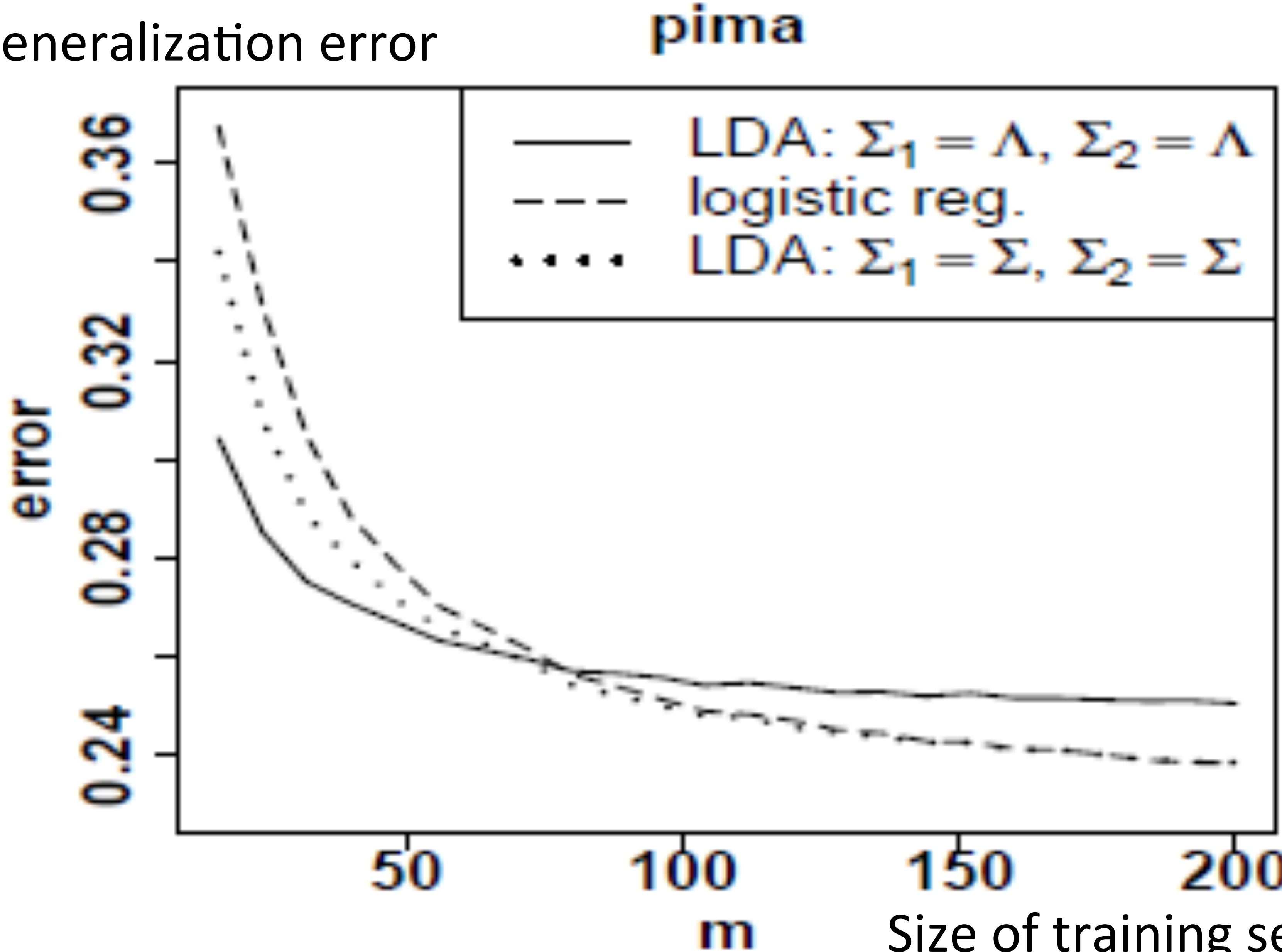
In numerical analysis, the speed at which a convergent sequence approaches its limit is called the rate of convergence.

Ng, Jordan,. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes." *Advances in neural information processing systems* 14 (2002): 841.

generalization error



generalization error



Size of training set

Xue, Jing-Hao, and D. Michael Titterington. "Comment on “On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes”."*Neural processing letters* 28.3 (2008): 169-187.

Discriminative vs. Generative

- Empirically, generative classifiers approach their asymptotic error faster than discriminative ones
  - Good for small training set
  - Handle missing data well (EM)
- Empirically, discriminative classifiers have lower asymptotic error than generative ones
  - Good for larger training set

References

- Prof. Tan, Steinbach, Kumar's "Introduction to Data Mining" slide
- Prof. Andrew Moore's slides
- Prof. Eric Xing's slides
- Prof. Ke Chen NB slides
- Hastie, Trevor, et al. *The elements of statistical learning*. Vol. 2. No. 1. New York: Springer, 2009.