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 $p(C = c_k)$

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

\[ p(c=1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \]
Discriminative vs. Generative

Logistic Regression

Gaussian

\[ \arg\max_c p(c|x) \]
LDA vs. Logistic Regression

- **LDA (Generative model)**
  - 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

- **Logistic Regression (Discriminative model)**
  - Assumes class-conditional densities are members of the (same) exponential family distribution \( \Pr(x|\theta) \)
  - 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
  - $h_{\text{gen}}$ and $h_{\text{dis}}$: generative and discriminative classifiers
  - $h_{\text{gen, inf}}$ and $h_{\text{dis, inf}}$: same classifiers but trained on the entire population (asymptotic classifiers)
  - $n \to \infty$, $h_{\text{gen}} \to h_{\text{gen, inf}}$ and $h_{\text{dis}} \to h_{\text{dis, inf}}$

Discriminative vs. Generative

Proposition 1: \( \frac{h_{true}}{vs.} \)

\[
\epsilon(h_{dis,\inf}) \leq \epsilon(h_{gen,\inf})
\]

= asymptotic error

Proposition 1 states that asymptotically, 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
- \( \epsilon \): generalization error
Logistic Regression vs. NBC

**Discriminative** classifier (Logistic Regression)
- Smaller asymptotic error
- Slow convergence $\sim O(p)$

**Generative** classifier (Naive Bayes)
- Larger asymptotic error
- Can handle missing data (EM)
- Fast convergence $\sim O(lg(p))$

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

generalization error

![Graph showing comparison between Logistic Regression and Naive Bayes in terms of error rate vs. size of training set.]
generalization error

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