Generative v.s. Discriminative Classifiers

INSTRUCTOR: HONGJIE CHEN

JUNE 2ND 2022

Generative v.s. Discriminative Classifiers

- Training classifiers: $f : X \rightarrow Y$ or $P(Y|X)$
- Generative classifiers (e.g., Naïve Bayes)
  - Assumption on $P(X|Y), P(Y)$
  - Estimates parameters of $P(X|Y), P(Y)$ from training data
  - Use Bayes rule to calculate $P(Y|X)$

- Discriminative classifers (e.g., Logistic Regression)
  - Assumption on $P(Y|X)$
  - Estimates parameters of $P(Y|X)$ directly from training data

Which one to use

- Restrictiveness of modeling assumption
  - NB assumes conditional independence

- Learning curve
  - Rate of convergence (in amount of training data) toward asympototic (infinite data) hypothesis
    - NB can converge faster if assumption is correct

Gaussian Naïve Bayes v.s. Logistic Regression

- Assume Boolean Y, continuous X_i, n features for X
- Number of parameters to estimate, or model size
  - Gaussian Naïve Bayes

$$P(X_i \mid Y = y_k) \sim \mathcal{N}(\mu_{ik}, \sigma_{ik})$$

  - Gaussian Naïve Bayes, assume independent of Y

$$P(X_i \mid Y = y_k) \sim \mathcal{N}(\mu_{ik}, \sigma_i)$$

  - Logistic Regression

$$P(Y = 1 \mid X, W) = \frac{1}{1 + exp(w_0 + \sum_{i=1}^{n} w_i x_i)}$$

Gaussian Naïve Bayes v.s. Logistic Regression

- Assume Boolean Y, continuous X_i, n features for X
- Number of parameters to estimate, or model size
  - Gaussian Naïve Bayes

    $$P(X_i \mid Y = y_k) \sim \mathcal{N}(\mu_{ik}, \sigma_{ik}),\ 1 + 2n + 2n = 4n + 1$$

  - Gaussian Naïve Bayes, assume independent of Y

    $$P(X_i \mid Y = y_k) \sim \mathcal{N}(\mu_{ik}, \sigma_i),\ 1 + 2n + n = 3n + 1$$

  - Logistic Regression

    $$P(Y = 1 \mid X, W) = \frac{1}{1 + exp(w_0 + \sum_{i=1}^{n} w_i x_i)},\ n + 1$$

Asymptotic Comparison

Have infinite training samples

- When the conditional independence assumption is correct
  - Identical classifiers
- When the assumption is incorrect
  - Logistic Regression is less biased

Convergence Rate

- Let $\epsilon_{Algo,m}$ refer to expected error of learning algorithm $Algo$ after m steps
- Let n be the number of features for X

$$\epsilon_{LR,m} = \epsilon_{LR,\infty} + O(\sqrt{\frac{n}{m}})$$

$$\epsilon_{GNB,m} = \epsilon_{GNB,\infty} + O(\sqrt{\frac{\log n}{m}})$$