Generative Models Revisited

Lyle Ungar

Penn
University of Pennsylvania

Which are generative models?

$p(y|x; \theta)$

a) Yes
b) No

$p(x; \theta)$

$p(x,y; \theta)$

Generative Models

hidden	observed
$s \sim p(k)$	$p(\boldsymbol{x}\|s=k) = N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
$s \sim p(y)$	$p(\boldsymbol{x}\|s=y) = N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y) = \Pi_j N(\mu_{jy}, \sigma_{jy}^2)$
$s \sim N(0,I)$	$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \varepsilon$
$s \sim N(0,I)$	$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{s} + \varepsilon$ $\quad$ $\boldsymbol{y} = \boldsymbol{B}\,\boldsymbol{s} + \varepsilon$

Dimensions:

$\boldsymbol{s}$ is $k \times 1$ $\quad$ $\boldsymbol{x}$ is $p \times 1$ $\quad$ $\boldsymbol{y}$ is $q \times 1$

A) PCA

B) CCA

C) GMM

D) Naïve Bayes

Generative models for classification

- Naïve Bayes
- K-Means/GMM

Generative models for classification

- **K-Means for prediction**
  - Cluster training points using k-means
  - Find the most frequent label (or average of real-valued labels) for the points in each cluster

    $\hat{y}_k = \Sigma_i \Sigma_k \, I(x_i \text{ in cluster } k) \, y_i$
  - To predict at a new x
    - —find which cluster centroid μ_k the new x is closest to
    - —look up the label or $\hat{y}_k$ for that cluster k

Generative models for classification

- **GMM for prediction**
  - **Estimate GMM**
    - —estimate π_k μ_k, Σ_k for each cluster
  - **Find the expected value of the label for each cluster**
    - —weighting by the degree of membership of x_i in the cluster k

      $\hat{y}_k = \Sigma_i \Sigma_k \, p(\text{cluster}=k \mid x=x_i) \, y_i$
  - **To predict at a new x**
    - —find the probability of x belonging to each cluster

      $p(\text{cluster}=k \mid x=x_i)$
    - —look up the $\hat{y}_k$ for that cluster
    - —the prediction is a weighted combination of those cluster labels

      $\Sigma_k \, p(\text{cluster}=k \mid x=x_i) \, \hat{y}_k$

This is not common, so I just made it up. Is there a better way?

LDA – Review

- **Which symbol corresponds to each of these**
  - P(topic=k)
  - P(topic=k) in document d
  - P(word= w_j | topic=k)

A) β_{jk}

B) θ_k

C) α_k

D) **z**