Lecture 09: Density Estimation

Course Teacher: Md. Sharifful Islam Bhuyan

Probabilistic Model

- Features/attributes are now random variable
- Learning joint distribution from data
- Simplifying assumption: attributes are independent (Naïve Bayes)
- Why probabilistic?
  - Decision tree: Impure class
  - Dice rolling: always 6

Estimation Techniques

- Prior, likelihood, posterior, partition function

$$p(h_i|\mathbf{d}) = \frac{p(\mathbf{d}|h_i)p(h_i)}{p(\mathbf{d})}$$

$$h_{MLE} = \arg\max_{h_i} p(\mathbf{d}|h_i) = \arg\min_{h_i}(-\log p(\mathbf{d}|h_i))$$

$$h_{MAP} = \arg\max_{h_i} p(\mathbf{d}|h_i)p(h_i) = \arg\min_{h_i}(-\log p(\mathbf{d}|h_i) - \log p(h_i))$$

$$p_{Bayesian}(h_i|\mathbf{d}) = \frac{p(\mathbf{d}|h_i)p(h_i)}{p(\mathbf{d})} = \frac{p(\mathbf{d}|h_i)p(h_i)}{\sum_i p(\mathbf{d}|h_i)p(h_i)}$$

MLE for Univariate Gaussians

Likelihood $l = P(x_1)P(x_2)\ldots P(x_N)$

$$L = \log P(x_1)P(x_2)\ldots P(x_N) = \sum_{j=1}^{N} \log P(x_j) = \sum_{j=1}^{N} \log \frac{1}{\sigma\sqrt{(2\pi)}} e^{\left(-\frac{(x_j-\mu)^2}{2\sigma^2}\right)}$$

$$= \sum_{j=1}^{N} \log \frac{1}{\sigma\sqrt{(2\pi)}} - \sum_{j=1}^{N} \left(\frac{(x_j-\mu)^2}{2\sigma^2}\right) = N(-\log\sqrt{(2\pi)} - \log\sigma) - \sum_{j=1}^{N} \left(\frac{(x_j-\mu)^2}{2\sigma^2}\right)$$

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{N}(x_j-\mu) = 0, \qquad \mu = \frac{\sum_{j=1}^{N} x_j}{N}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{N}(x_j-\mu)^2 = 0, \qquad \sigma = \sqrt{\frac{\sum_{j=1}^{N}(x_j-\mu)^2}{N}}$$

MLE *for* Multivariate Gaussian

$$P(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} e^{\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right)}$$

$$\boldsymbol{\mu} = \frac{\sum_{j=1}^{N} \mathbf{x}_j}{N}$$

$$\boldsymbol{\Sigma} = \frac{\sum_{j=1}^{N} (\mathbf{x}_j - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^T}{N}$$