Generative Learning

Tuesday, 20 March 2018 16:42

Descriminitive learning: Learns $p(y|x)$
or learns $h_\theta(x) \in \{0,1\}$ directly.

Generate:

$p(x|y)$

condition class

e.g. Assum $x \in \mathbb{R}^n$, continuous-value.

Gaussian Distribution Algorithm

$p(x|y)$ is Gaussian

// $\vec{z} \sim N(\vec{\mu}, \Sigma)$

$$p(z) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$\vec{\mu}$: mean. Σ covariance matrix. $\Sigma = E[(x-\mu)(x-\mu)^T]$

$$p(y) = \phi^y (1-\phi)^y$$

$$p(x|y=0) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0)\right)$$

$$p(x|y=1) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right)$$

$$l(\phi, \Sigma, \mu_0, \mu_1) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)})$$

Joint likelihood: $= \log \prod_{i=1}^{m} p(x^{(i)} | y^{(i)}) \, p(y^{(i)})$

Logistic Regression: $l(\theta) = \log \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}; \theta)$ // conditional likelihood.

maximize: l with $\phi, \mu_0, \mu_1, \Sigma$

$$\phi = \frac{\sum_i y^{(i)}}{m} = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}{m}$$

$$\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\}}$$ → $\dfrac{\text{sum of } x \text{ which } y^{(i)} = 0}{\text{number of } y^{(i)} = 0.}$

$$\mu_1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}$$

Σ: on the notes.

predict:

$$\arg\max_y p(y|x) = \arg\max_y \frac{p(x|y)p(y)}{p(x)} = \arg\max_y p(x|y)p(y)$$

If $p(y)$ is uniform: $\arg : p(x|y)$

$x|y \sim$ Gaussian $\Rightarrow$ logistic posterior for $p(y=1|x)$

$x|y=1 \sim \text{Poisson}(\lambda_1)$ $x|y=0 \sim \text{Poisson}(\lambda_0)$

⇓

$p(y=1|x)$ is a logistic function.

$x|y \sim$ Gaussian ⇒ generate algorithm will better
but logistic function is a more general way.

$x|y=1$ is Exponential Family(η_1) also $p(y=1|x)$ following logistic function.

Naive Bayes.

assume x_i are conditional independent given y

$P(x_1 \ldots x_m|y) = p(x_1|y)\, p(x_2|y, x_1) \ldots$

with assumption: $= \prod_{i=1}^{n} p(x_i|y)$

Parameters:

$\phi_{i|y=1} = p(x_i = 1 | y = 1)$

$\phi_{i|y=0} = p(x_i = 1 | y = 0)$

$\phi_y = p(y = 1)$.

Joint likelihood

$$\mathcal{L}(\phi_y, \phi_{i|y=0}, \phi_{i|y=1}) = \prod_{i=1}^{n} p(x^{(i)}, y^{(i)})$$

$$\phi_{j|y=1} = \frac{\sum_{i=1}^{m} 1\{x_j^{(i)} = 1, y^{(i)} = 1\}}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}$$

$$\phi_y = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}{m}$$

| $p(y|x)$

| $p(x|y)\ p(y)$

Laplace Smoothing.

if y takes on k possible value.

$$p(y=1) = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = j\} + 1}{m + k}.$$