GDA

Generative Learning Algorithms

- Gaussian Discriminant Analysis (GDA)
- Generative and discriminative comparison
- Naive Bayes

Discriminative:

learn $p(y|x)$ or $h_\theta(x) = \begin{cases} 1 \\ 0 \end{cases}$ directly.

Generative learning algorithm:

learn $p(x|y)$ and $p(y)$ — class prior;   x — feature, y — class

Bayes rule:

$p(y=1|x) = \frac{p(x|y=1)p(y=1)}{p(x)}$

where $p(x) = p(x|y=1)p(y=1) + p(x|y=0)p(y=0)$

Gaussian Discriminant Analysis (GDA)

Suppose $x \in R^n$ (drop $x_0 = 1$ convention)

Assume $p(x|y)$ is Gaussian (features of each class follow multivariate Gaussian)

$x|y=0 \sim N(\mu, \Sigma); x|y=1 \sim N(\mu, \Sigma)$

GDA model: parameters $\mu_0, \mu_1 \in R^n, \Sigma \in R^{n \times n}, \phi \in R$

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

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

$p(y) = \phi^y (1-\phi)^{1-y}$ or $P(y=1) = \phi$

Training set: $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$

Joint likelihood: $L(\phi, \mu_0, \mu_1, \Sigma) = \Pi_{i=1}^m p(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma)$
$= \Pi_{i=1}^m p(x^{(i)}|y^{(i)}) p(y^{(i)})$

In contrast, Discriminative: maximize

Conditional likelihood: $L(\theta) = \Pi_{i=1}^{m} p(y^{(i)}|x^{(i)}, \theta)$

Maximum likelihood estimation:

$\max_{\phi,\mu_0,\mu_1,\Sigma} L(\phi, \mu_0, \mu_1, \Sigma) \rightarrow$

$\phi = \frac{\Sigma_{i=1}^{m} y^{(i)}}{m} = \frac{\Sigma_{i=1}^{m} L\{y^{(i)}=1\}}{m}$ proportion of $y^{(i)} = 1$

$\mu_0 = \frac{\Sigma_{i=1}^{m} L\{y^{(i)}=0\} x^{(i)}}{\Sigma_{i=1}^{m} L\{y^{(i)}=0\}}$ look at all benign tumors and take averages

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

$\Sigma = \frac{1}{m} \Sigma_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$

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

Compared to Logistics Regression: has stronger assumption; if nearly Gaussian (most cases), performs better, else (e.g. poisson distribution) worse; more efficient