(5)

Gaussian discriminant analysis
& Naive Bayes

Generative Learning algorithms

⇒ So far we've mainly been talking about learning algorithms that models

$$P(y|x;\theta)$$

↳ Such algorithms are called discriminative learning algorithm.

⇒ Here, we'll talk about algorithms that instead try to model $P(x|y)$ (and $P(y)$)

↳ These algorithms are called generative learning algorithms.

⇒ After modeling $P(y)$ (called the class priors) and $P(x|y)$, our algorithm can then use Bayes rule to derive the posterior distribution on y given x

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$$\rightarrow \sum P(x|y)P(y)$$

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

$$= \arg\max_y P(x|y)P(y)$$

Date

Page

* Gaussian discriminant analysis (GDA)

⇒ In this model, we'll assume that $P(x|y)$ is distributed according to a multivariate normal distribution.

$$y \sim \text{Bernoulli}(\phi)$$
$$x|y=0 \sim N(\mu_0, \Sigma)$$
$$x|y=1 \sim N(\mu_1, \Sigma)$$

⇒ The log-likelihood of the data is given by:

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

⇒ By maximizing l with respect to the parameters, we find the maximum likelihood estimate of the parameters:

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

$$\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\}}$$

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

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

⇒ If $P(x|y)$ is multivariate gaussian then $P(y|x)$ necessarily follows a logistic function.

⇒ The Converse, however is not true; i.e. $P(y|x)$ being a logistic function does not imply $P(x|y)$ is multivariate gaussian.

⇒ This shows that GDA makes stronger modeling assumptions about the data than does logistic regression.

⇒ When modeling assumptions are Correct, then GDA will find better fits to the data, and is a better model.

⇒ In Contrast, by making significantly weaker assumptions, logistic regression is more robust and less sensitive to incorrect modeling assumptions.