NaiveBayes

19 July 2018 15:47

Given a training dataset which are sampled from a hidden $P(x, y)$ distribution

$D = \{ (x^1, y_1), \ldots (x^N, y_N) \}$

Goal: (during the training process) to estimate $\hat{p}(x|y)$, $\hat{P}(y)$

(1) How to estimate $\hat{p}(y)$ $\quad P(x,y) = p(y) P(x|y)$

$D \Rightarrow \{N_1, N_2 \ldots N_k\}$ $\quad \sum_j N_j = N$

# of examples in each class

$\hat{p}(y) = \boxed{\frac{N_1}{N}, \frac{N_2}{N}, \ldots \frac{N_k}{N}}$

$\hat{p}(y=j) = \frac{N_j}{N}$

Eg: students

2	3	4
0.2	0.7	0.1

(2) How to estimate $\hat{p}(x|y) \equiv \hat{P}(x_1, \ldots x_d | y)$?

eg: student : $x_1 \equiv$ age

$\hat{p}(x_1 | y=2) \quad \hat{p}(x_1 | y=3) \quad \hat{p}(x_1 | y=4)$

$p(x_1 | y) \sim N(\mu_y, \sigma_y)$

$\hat{p}(x_1 | y=2) \sim N(\mu_2 = 18.7, \sigma_2 = 0.5)$

$\hat{p}(x_1 | y=3) \sim N(\mu_3 = 19.6, \sigma_3 = 0.4)$

$\hat{p}(x_1 | y=4) \sim N(\mu_4 = 20.5, \sigma_4 = 0.45)$

Testing! $x : x_1 = 19$

$$p(y|x) \propto \hat{p}(x|y)\hat{p}(y)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_y} e^{\frac{-(x_1 - \mu_y)^2}{2\sigma_y^2}} \hat{p}(y)$$

$$p(y=2 | x : x_1 = 19) \propto \frac{1}{0.5} e^{\frac{-(19-18.7)^2}{2 * 0.5^2}} * 0.2$$

$$p(y=3 | x_1 = 19) \propto = \frac{1}{0.4} e^{\frac{-(19-19.6)^2}{2 * 0.4^2}} * 0.7$$

$$p(y=4 | x_1 = 19) \propto \frac{1}{\ldots} e^{\frac{-(20.5-19)^2}{\ldots^2}} * 0.1$$

$0.45 < 2 * 0.45$

$P(y|x) \leftarrow$ posterior probability of y

$d > 1$

$P(x_1, x_2 \ldots x_d | y)$

age

late or not

# of questions answered

row in class

$\mathcal{E}v$

Conditional independence assumption
Example from slides