یادگیری ماشین

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

مدرس: دکتر سید ابوالفضل مطهری

دانشکده‌ی مهندسی کامپیوتر

درس‌نامه هشتم

می‌دانیم که در رابطه بیز به

$$p_k(x) = \mathbb{P}[Y = k | X = x]$$

نیاز داریم. با استفاده از قانون بیز داریم:

$$p_k(x) = \frac{\mathbb{P}[X = x|Y = k]\mathbb{P}[Y = k]}{\mathbb{P}[X = x]} = \frac{f_k(x)\pi_k}{\mathbb{P}[X = x]}$$

عملا به تخمین $\pi_k, f_k(x)$ نیاز داریم. در روش Bayes Naive یک فرض ساده کننده قرار می‌دهیم و آن این است که

$$\mathbb{P}[X = x|Y = k] = \mathbb{P}[X_۱ = x_۱|Y = k]\mathbb{P}[X_۲ = x_۲|Y = k] \dots \mathbb{P}[X_p = x_p|Y = k]$$

و یا به عبارت دیگر

$$f_k(x) = f_{k_۱}(x_۱) f_{k_۲}(x_۲) \dots f_{k_p}(x_p)$$

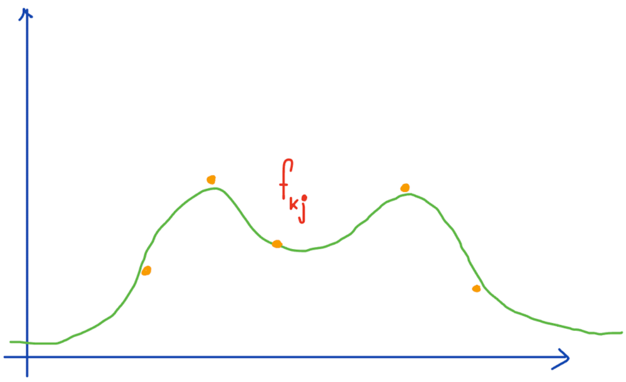
حال برای تخمین f_{k_i} ها، می‌توانیم در هر بعد یک تخمینگر مناسب داشته باشیم. بنابراین کافیست داده‌های مربوط به یک دسته را مورد استفاده قرار دهیم و

$$\hat{\pi_k} = \frac{n_k}{n}$$

که در آن n_k تعداد نمونه‌های دسته k می‌باشد. و برای $f_{k_j}(x_j)$ در ابتدا یک توزیع برای آن درنظر می‌گیریم پارامترهای آن را روی داده‌ها به‌دست می‌آوریم، به عنوان مثال:

$$X_j|Y = k \sim N(\mu_{j_k}, \sigma^۲_{j_k})$$

یک راه دیگر یک تخمین از توزیع به صورت غیر پارامتری بزنیم، به‌عنوان مثال از Kernel density estimator استفاده نماییم.



اگر ورودی x_j رقمی نبود، آنگاه می‌توانیم تعداد هر کدام از حالت‌ها را بشماریم و از این طریق تخمینی از توزیع داشته باشیم. مثلا اگر

$$X_j \in \{۱, ۲, ۳\}$$

آنگاه

$$X_j|Y = k \sim \begin{cases} \frac{n_{j_۱}}{n_j} & X_j = ۱ \\ \frac{n_{j_۲}}{n_j} & X_j = ۲ \\ \frac{n_{j_۳}}{n_j} & X_j = ۳ \end{cases}$$

که در آن n_{j_l} تعداد نمونه‌های مشاهده شده در دسته که مقداری برابر l دارند می‌باشد.