محمدرضا عباسی ۹۸۴۲۲۱۰۱ جواب تمرین دوم

جواب سوال۳)

قضیه بیز (Bayes' theorem): روشی برای دسته‌بندی پدیده‌ها، بر پایه احتمال وقوع یا عدم وقوع یک پدیده‌است و در نظریه احتمالات با اهمیت و پرکاربرد است. اگر برای فضای نمونه‌ای مفروضی بتوانیم چنان افرازی انتخاب کنیم که با دانستن اینکه کدامیک از پیشامدهای افراز شده رخ داده‌است، بخش مهمی از عدم قطعیت تقلیل می‌یابد.

این قضیه از آن جهت مفید است که می‌توان از طریق آن، احتمال یک پیشامد را با مشروط کردن نسبت به وقوع یا عدم وقوع یک پیشامد دیگر محاسبه کرد. در بسیاری از حالت‌ها، محاسبه ی احتمال یک پیشامد به صورت مستقیم کاری دشوار است. با استفاده از این قضیه و مشروط کردن پیشامد مورد نظر نسبت به پیشامد دیگر، می‌توان احتمال مورد نظر را محاسبه کرد.

این رابطه به خاطر بزرگداشت توماس بیز فیلسوف انگلیسی به نام فرمول بیز معروف است.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

فرض کنید که فضای فرضیه h و مجموعه مثال‌های آموزش D موجود باشند. مقادیر احتمال زیر را تعریف می‌کنیم:

P(h): احتمال پیشین (prior probablity) که فرضیه h قبل از مشاهده داده آموزشی D داشته‌است. اگر چنین احتمالی موجود نباشد می‌توان به تمامی فرضیه‌ها احتمال یکسانی نسبت داد.

P(D): احتمال مشاهده داده آموزشی D.

P(D|h): درست نمایی (likelihood) یا احتمال مشاهده داده آموزشی D به فرض آنکه فرضیه h صادق باشد.

P(h|D): احتمال پسین (posterior probablity) یا احتمال فرضیه h به شرط مشاهده داده آموزشی D.

توجه شود که احتمال پیشین P(h) مستقل از داده آموزشی است ولی احتمال پسین P(h|D) تأثیر داده آموزشی را منعکس می‌کند.

دسته بند بیز و مدل احتمالاتی

بیز ساده را می‌توان یک مدل برمبنای احتمال شرطی در نظر گرفت. فرض کنید $X = (x_1, \dots, x_n)$ برداری از n ویژگی را بیان کند که به صورت متغیرهای مستقل هستند. به این ترتیب می‌توان احتمال رخداد C_k یعنی $p(C_k \mid x_1, \dots, x_n)$ را به عنوان یکی از حالت‌های کلاس رخدادهای مختلف به ازاء kهای متفاوت، به شکل زیر نمایش داد.

$$p(C_k \mid X) = \frac{p(C_k)\, p(X \mid C_k)}{p(X)}$$

رابطه ۱

همانطور که دیده می شود رابطه بالا همان قضیه بیز است. به عنوان یادآوری قضیه بیز را براساس احتمالات پیشامدهای «پیشین» (Prior)، «پسین» (Posterior)، «درستنمایی» (Likelihood) و «شواهد» (Evidence) در رابطه زیر بازنویسی می‌کنیم.

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

به این ترتیب برای محاسبه احتمال $p(C_k \mid x_1,\ldots,x_n)$ کافی است از «احتمال توام» (Joint Probability) کمک بگیریم و به کمک احتمال شرطی با توجه به استقلال متغیرها، آن را ساده کنیم.

$$\begin{aligned}
&p(C_k, x_1, \ldots, x_n) = p(x_1, \ldots, x_n, C_k) \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k) p(x_2, \ldots, x_n, C_k) \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k) p(x_2 \mid x_3, \ldots, x_n, C_k) p(x_3, \ldots, x_n, C_k)، \\
&= \ldots \\
&= p(x_1 \mid x_2, \ldots, x_n, C_k) p(x_2 \mid x_3, \ldots, x_n, C_k) \ldots p(x_{n-1} \mid x_n, C_k) p(x_n \mid C_k) p(C_k)
\end{aligned}$$

با فرض استقلال مولفه‌ها یا ویژگی‌های x_iها از یکدیگر می‌توان احتمالات را به شکل ساده‌تری نوشت. کافی است رابطه زیر را در نظر بگیریم.

$$p(x_i \mid x_{i+1}, \ldots, x_n, C_k) \approx p(x_i \mid C_k).$$

به این ترتیب احتمال توام را به صورت حاصلضرب احتمال شرطی می‌توان نوشت.

$$\begin{aligned}
p(C_k \mid x_1, \ldots, x_n) &\propto p(C_k, x_1, \ldots, x_n) \\
&\approx p(C_k)\, p(x_1 \mid C_k)\, p(x_2 \mid C_k)\, p(x_3 \mid C_k)\, \cdots \\
&= p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k),
\end{aligned}$$

رابطه ۲

نکته: در رابطه ۱ مخرج کسر در همه محاسبات یکسان و ثابت است در نتیجه می‌توان احتمال شرطی را متناسب با احتمال توام در نظر گرفت. در رابطه بالا این تناسب را با علامت $\propto$ نشان داده‌ایم.

با توجه به نکته گفته شده، و رابطه ۲ می‌توانیم احتمال شرطی معرفی شده در رابطه ۱ را به صورت زیر بدست آوریم. در نتیجه احتمال تعلق یک مشاهده به دسته یا گروه C_k با توجه به مشاهدات X مطابق با رابطه زیر مشخص خواهد شد.

$$p(C_k \mid x_1, \ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

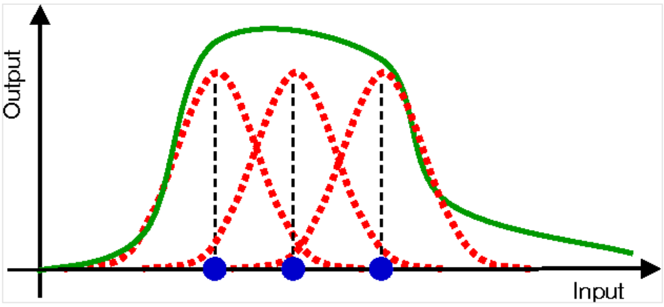
توجه داشته باشید که در اینجا احتمال شواهد (مشاهدات) به صورت $Z = p(\mathrm{x}) = \sum_k p(C_k)\, p(\mathrm{x} \mid C_k)$ در نظر گرفته شده است. واضح است که Z به شواهد و مشاهدات $x_1, \ldots, x_n$ وابسته است.

ساخت دسته‌بند براساس مدل احتمالاتی

در قسمت قبل با نحوه محاسبه مدل احتمالاتی بیز آشنا شدید. اما در این بخش به کمک «قواعد تصمیم» (Decision Rule)، دسته‌بند بیز را ایجاد و کامل می‌کنیم. یکی از اساسی‌ترین قواعد تصمیم، انتخاب فرضیه محتمل‌تر است. به این ترتیب از بین تصمیمات مختلف، آن کاری را انجام می‌دهیم که براساس شواهد جمع‌آوری شده، بیشترین احتمال رخداد را دارد. این قاعده را «حداکثر پسین» (Maximum Posterior) یا به اختصار MP می‌نامند. به این ترتیب دسته بند بیز را می‌توان به صورت تابعی از تصمیمات C_k در نظر گرفت که بوسیله تابع $\hat{y}$ تخمین زده می‌شود. حداکثرسازی این تابع را به صورت زیر نشان می‌دهیم.

$$\hat{y} = \underset{k \in \{1, \ldots, K\}}{\operatorname{argmax}}\ p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k).$$

در نتیجه با توجه به توزیع‌های مختلفی که ممکن است نمونه تصادفی داشته باشد (نمونه از آن جامعه آمده باشد) یعنی $p(x_i|C_k)$ می‌توان پارامترها را محاسبه و یا برآورد کرد.



دسته بند بیز ساده گاوسی (Gaussian Naive Bayes)

اگر مشاهدات و داده‌ها از نوع پیوسته باشند، از مدل احتمالی با توزیع گاوسی یا نرمال برای متغیرهای مربوط به شواهد می‌توانید استفاده کنید. در این حالت هر دسته یا گروه دارای توزیع گاوسی است. به این ترتیب اگر k دسته یا کلاس داشته باشیم می‌توانیم برای هر دسته میانگین و واریانس را محاسبه کرده و پارامترهای توزیع نرمال را برای آن‌ها برآورد کنیم. فرض کنید که μ_k میانگین و σ_k^2 واریانس دسته kام یعنی C_k باشد. همچنین ν را مشاهدات حاصل از متغیرهای تصادفی X در نظر بگیرید. از آنجایی که توزیع X در هر دسته گاوسی (نرمال) فرض شده است، خواهیم داشت:

$$p(x=v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\, e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

دسته بند بیز ساده چندجمله‌ای (Multinomial Naive Bayes)

بیز ساده چندجمله‌ای، به عنوان یک دسته‌بند متنی بسیار به کار می‌آید. در این حالت برحسب مدل احتمالی یا توزیع چند جمله‌ای، برداری از n ویژگی برای یک مشاهده به صورت $X=(x_1,\ldots,x_n)$ با احتمالات $(p_1,\ldots,p_n)$ در نظر گرفته می‌شود. مشخص است که در این حالت بردار X بیانگر تعداد مشاهداتی است که ویژگی خاصی را دارا هستند. به این ترتیب تابع درستنمایی در چنین مدلی به شکل زیر نوشته می‌شود.

$$p(\mathrm{x} \mid C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!}\prod_i p_{ki}{}^{x_i}$$

اگر مدل بیز ساده را براساس لگاریتم تابع درستنمایی بنویسیم، به صورت یک دسته‌بند خطی درخواهد آمد.

$$\begin{aligned}\log p(C_k \mid \mathrm{x}) &\propto \log\left(p(C_k)\prod_{i=1}^{n} p_{ki}{}^{x_i}\right)\\ &= \log p(C_k) + \sum_{i=1}^{n} x_i \cdot \log p_{ki}\\ &= b + \mathrm{w}_k^\top \mathrm{x}\end{aligned}$$

واضح است که در این حالت $b=\log p(C_k)$ و $w_{ki}=\log p_{ki}$ است.

دسته بند بیز ساده برنولی (Bernoulli Naive Bayes)

در این قسمت به بررسی توزیع برنولی و دسته‌بندی بیز خواهیم پرداخت. به شکلی این نوع از دسته‌بند بیز بیشترین کاربرد را در دسته‌بندی متن‌های کوتاه داشته، به همین دلیل محبوبیت بیشتری نیز دارد. در این مدل در حالت چند متغیره، فرض بر این است که وجود یا ناموجود بودن یک ویژگی در نظر گرفته شود. برای مثال با توجه به یک لغتنامه مربوط به اصطلاحات ورزشی، متن دلخواهی مورد تجزیه و تحلیل قرار می‌گیرد و بررسی می‌شود که آیا کلمات مربوط به لغتنامه ورزشی در متن وجود دارند یا خیر. به این ترتیب مدل تابع درستنمایی متن براساس کلاس های مختلف C_k به شکل زیر نوشته می‌شود.

$$p(\mathsf{x} \mid C_k) = \prod_{i=1}^{n} p_{ki}^{x_i}(1 - p_{ki})^{(1-x_i)}$$

مشخص است که منظور از p_{ki} احتمال تولید مشاهده x_i از کلاس C_k است.

نکته: با توجه به استقلال مشاهدات، تابع درستنمایی به صورت حاصلضرب نوشته شده است.