



بخش 1:

اگر فضای نمونه ای ما توسط $B_1, B_2, B_3, \dots, B_n$ افراز شده باشد بطوریکه احتمال هر کدام از آنها بزرگتر از 0 باشد آنگاه برای هر پیشامد A می توانیم بنویسیم:

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)}$$

زیرا بر اساس رابطه احتمال شرطی می دانیم:

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)}$$

که نام این قضیه، قضیه بیز می باشد.

: Naive Bayes Classifiers

بیز ساده را می توان 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 می توان احتمال رخداد C_k یعنی $p(C_k|x_1, \dots, x_n)$ را به عنوان یکی از حالت های کلاس رخداد های مختلف به ازاء k های متفاوت، به شکل زیر نمایش داد:

$$p(C_k | X) = \frac{p(C_k) p(X | C_k)}{p(X)}$$

$$P(X|C_k) = \prod_{i=1}^n p(x_i | C_k)$$

چون مولفه های X از یکدیگر مستقل هستند داریم که:



در نتیجه احتمال تعلق یک مشاهده به دسته یا گروه C_k با توجه به مشاهدات X مطابق با رابطه زیر مشخص خواهد شد.

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

$$Z = p(x) = \sum_k p(C_k) p(x | C_k) \quad \text{که}$$

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 k 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 k ام یعنی C_k باشد. همچنین v 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:

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

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 $X=(x_1, \dots, x_n)$ با احتمالات (p_1, \dots, 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.

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

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 کلاس‌های مختلف C_k به شکل زیر نوشته می‌شود.

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

کاربردهای الگوریتم دسته بندی بیز ساده Naive Bayes:

از الگوریتم دسته بندی بیز ساده Naive Bayes می‌توان استفاده های بی شماری کرد. در اینجا برخی از برنامه های رایج Naive Bayes برای کارهای روزمره وجود دارد:

طبقه بندی اسناد:

این الگوریتم می‌تواند به شما کمک کند تا تعیین کنید که یک سند مشخص به کدام دسته تعلق دارد. می‌تواند برای طبقه بندی متون به زبان‌ها، ژانرها یا موضوعات مختلف (از طریق کلمات کلیدی) مورد استفاده قرار گیرد. فیلتر هرزنامه Naive Bayes به راحتی هرزنامه ها را با استفاده از کلمات کلیدی مرتب می‌کند به همین



دلیل از آنها برای حل مسائلی مانند spam-filtering استفاده می شود. الگوریتم باید برای تشخیص چنین احتمالاتی آموزش ببیند و سپس بتواند آنها را به طور موثر دسته بندی کند.

تحلیل احساسات:

بر اساس آنچه احساسات کلمات در متن بیان می شود، Naive Bayes می تواند احتمال مثبت یا منفی بودن آن را محاسبه کند. به عنوان مثال در بررسی مشتری، "خوب" یا "ارزان" معمولاً به معنای رضایت مشتری است. با این حال، Naive Bayes به کنایه و کلماتی که به معنای متفاوتی از خودشان به کار می رود حساس نیست.

طبقه بندی تصویر:

طبقه بندی تصویر برای اهداف شخصی و تحقیقاتی، ساخت طبقه بند Naive Bayesian آسان است، می توان آن را آموزش داد تا از طریق یادگیری ماشین تحت نظارت، ارقام دست نوشته را تشخیص دهد یا تصاویر را در دسته ها قرار دهد.

پیش بینی در لحظه:

Naive Bayes یک طبقه بندی یادگیری مطمئن و سریع است. بنابراین، می توان از آن برای پیش بینی در زمان واقعی استفاده کرد. این کاربرد بیشتر مربوط به طبقه بیز ساده گوسی می باشد.

پیش بینی چند کلاسه:

این الگوریتم به دلیل ویژگی پیش بینی چند کلاس نیز شناخته شده است. در اینجا می توانیم احتمال چندین کلاس متغیر هدف را پیش بینی کنیم.



نام و نام خانوادگی: محمدرضا ضیالاری

شماره دانشجویی: 97222057

شماره تمرین: بخش دوم تمرین شماره 2 درس داده کاوی

بخش 3, 2 :

در این بخش پس از بررسی غیر تهی بودن داده ها توابع برای محاسبه دسته بند بیزی ساده گوسی را پیاده سازی کردیم و به بررسی معیار های سنجش دقت پرداختیم . نتایج زیر حاصل شد :

```
f1_score 0.7516960651289011
precision_score 0.756043956043956
recall_score 0.7510775862068966
```

بخش 4 و 5 :

در ادامه با استفاده از پکیج ها دسته بند بیزی ساده گوسی را به کار بردیم و نتایج مشابه حاصل شد . که این اتفاق منطقی است زیرا داده های آموزشی و تست و فرمول های محاسبه یکسان هستند .

```
f1_score 0.7516960651289011
precision_score 0.756043956043956
recall_score 0.7510775862068966
```