## نکاتی درخصوص قضیه آماری بیز

- این قضیه اولین بار توسط توماس بیز نوشته شد و از یک قانون ساده احتمال شرطی استفاده میکند.
- استدلال **بیز** روشی بر پایه احتمالات برای استنتاج کردن است. هدف آمارگران بیزی در آنالیز آزمایشات، بیان احتمالیست که به احتمال واقعی نزدیک تر است.
- اساس این روش بر این اصل استوار است که برای هر کمیتی یک توزیع احتمال وجود دارد که با مشاهده یک داده جدید و استدلال در مورد توزیع احتمال آن میتوان تصمیمات بهینه ای اتخاذ کرد.
- **اهمیت یادگیری بیز**
- **در برخی کاربردها استفاده از روشهای یادگیری بیزی توانسته است راه حل های عملی مفیدی را ارائه کند.**
- **مطالعه یادگیری بیزی به فهم سایر روشهای یادگیری که به طور مستقیم از احتمالات استفاده نمی کنند کمک می کند.**
- **روش بیز برای برآوردهای مشکل مناسب است. خصوصاً آنجایی که از برآوردها برای یک تصمیم اجرایی استفاده میشود.**

**نگرش بیزی**

**نگرش بیزی بصورت زیر است**

۱. **دانش موجود در باره موضوع را بصورت احتمالاتی فرموله میکنیم.**

   - برای اینکار مقادیر کیفی را بصورت توزیع احتمال، فرضیات استقلال و غیره مدل می نمائیم. این مدل دارای پارامترهای ناشناخته ای خواهد بود.
   - برای هر یک از مقادیر ناشناخته، توزیع احتمال اولیه ای در نظر گرفته می شود که بازگو کننده باور ما به محتمل بودن هر یک ازاین مقادیر بدون دیدن داده است.

۲. **داده را جمع آوری می نمائیم.**

۳. **با مشاهده داده ها مقدار توزیع احتمال ثانویه را محاسبه می کنیم.**

۴. **با استفاده از این احتمال ثانویه:**

- به یک نتیجه گیری در مورد عدم قطعیت می رسیم.
- با میانگین گیری روی مقادیر احتمال ثانویه پیش بینی انجام می دهیم.
- برای کاهش خطای ثانویه مورد انتظار تصمیم گیری می کنیم.

## ویژگیهای تئوری بیز

۱. مشاهده هر دسته از داده ها می تواند به صورت جزئی باعث افزایش و یا کاهش احتمال درست بودن یک فرضیه گردد.

۲. برای بدست آوردن احتمال یک فرضیه می توان دانش قبلی را با داده های مشاهده شده ترکیب کرد. این دانش قبلی به دو طریق بدست میاید:

- احتمال قبلی برای هر فرضیه موجود باشد.
- برای داده مشاهده شده توزیع احتمال هر فرضیه ممکن موجود باشد.

۳. روشهای بیزی فرضیه هائی ارائه می دهند که قادر به پیش بینی احتمالی هستند (مثلاً نژاد مورد نظر به احتمال 93% دارای دو قلوزایی بالا می باشد).

۴. مثال های جدید را می توان با ترکیب وزنی چندین فرضیه دسته بندی نمود.

۵. حتی در مواردی که روشهای بیزی قابل محاسبه نباشند، میتوان از آنها به عنوان معیاری برای ارزیابی روشهای دیگر استفاده کرد.

## مشکلات عملی روش بیز

- به دانش اولیه در مورد تعداد زیادی مقادیراحتمال نیاز دارد. وقتی که این اطلاعات موجود نباشند اغلب ناگزیر به تخمین زدن آن هستیم. برای این کار از اطلاعات زمینه، داده هائی که قبلا جمع آوری شده اند، و فرضیاتی در مورد توزیع احتمال استفاده میشود.
- محاسبه فرضیات بهینه بیزی و تجزیه و تحلیل آنها سخت است. زیرا باید مسائل پیچیده را به مجموعه ای از مراحل ساده تفکیک کرد.

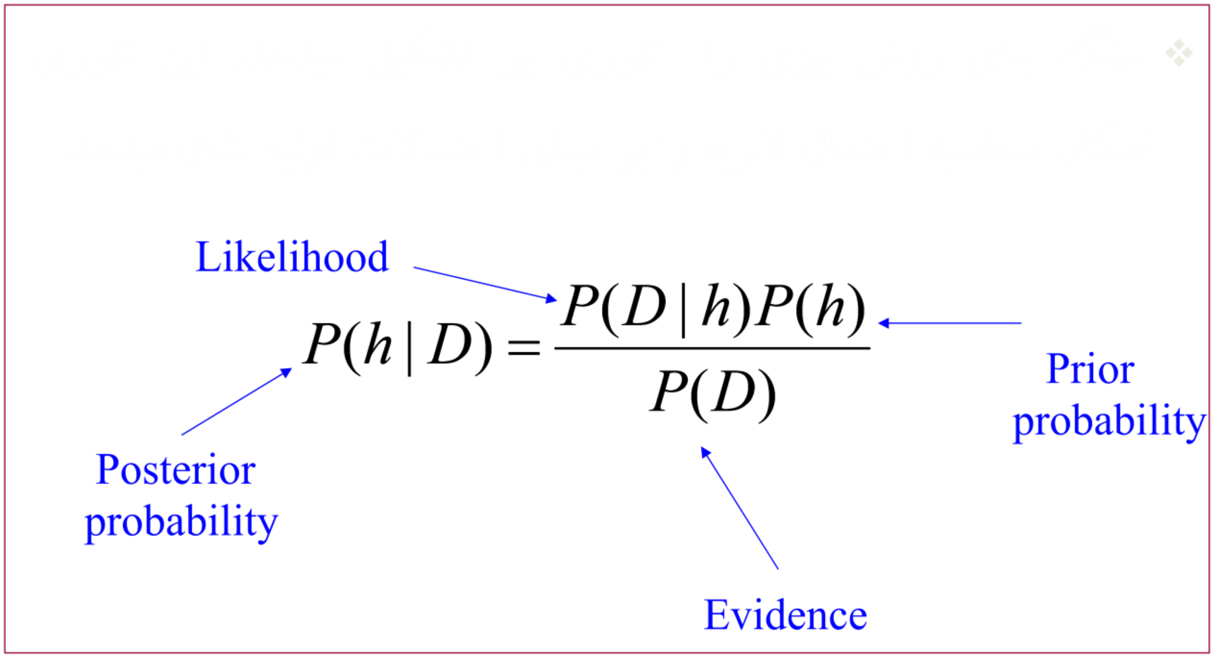
**تئوری بیز**

- در یادگیری بیزی معمولاً در فضای فرضیه H بدنبال **بهترین فرضیه ای** هستیم که درمورد داده های آزمایشی صدق کند. یک راه تعیین بهترین فرضیه، این است که بدنبال **محتمل ترین فرضیه ای** باشیم که با داشتن داده های آزمایشی و احتمال قبلی در مورد فرضیه های مختلف میتوان انتظار داشت.
- تئوری بیز چنین راه حلی را ارائه میدهد. این روش راه حل مستقیمی است که نیازی به جستجو ندارد.

**سنگ بنای روش بیز**

سنگ بنای روش بیزی را تئوری بیز تشکیل میدهد. این تئوری امکان محاسبه احتمال ثانویه را بر مبنای احتمالات اولیه نشان میدهد

# سنگ بنای روش بیز




**تئوری بیز**

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

❖ همان طور که مشاهده میشود با افزایش P(D) مقدار P(h|D) کاهش می یابد. زیرا هر چه احتمال مشاهده D مستقل از h بیشتر باشد به این معنا خواهد بود که D شواهد کمتری در حمایت از h در بر دارد.

**Maximum A Posteriori (MAP) hypothesis**

❖ در مسایلی که مجموعه ای از فرضیه های H وجود داشته و بخواهیم محتمل ترین فرضیه را از میان آنان انتخاب بکنیم، فرضیه با حداکثر احتمال Maximum A Posteriori (MAP) hypothesis نامیده میشود و از رابطه زیر بدست می آید.

$$\begin{aligned} h_{MAP} &\equiv \arg\max_{h\in H} P(h|D) \\ &= \arg\max_{h\in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg\max_{h\in H} P(D|h)P(h) \end{aligned}$$

در این رابطه مقدار $P(D)$ مستقل از $h$بوده و حذف میشود

**Maximum likelihood (ML) hypothesis**

- ❖ در مواقعی که هیچ اطلاعی در مورد $P(h)$ وجود نداشته باشد میتوان فرض کرد که تمام فرضیه های H دارای احتمال اولیه یکسانی هستند. در این صورت برای محاسبه فرضیه با حداکثر احتمال میتوان فقط مقدار $P(D|h)$ را در نظر گرفت. این مقدارlikelihood داده D با فرض h نامیده میشود و هر فرضیه ای که مقدار آن را ماکزیمم کند، فرضیه maximum liklihood (ML) نامیده میشود:

$$h_{ML}=\arg\max_{h\in H} P(D|h)$$

**مثال: تشخیص بیماری**

- ❖ **در یک مسئله تشخیص بیماری با دو فرصیه روبرو هستیم:**

**1- بیمار دارای سرطان است 2- بیمار سالم است**

- ❖ **داداهای آزمایشگاهی نشان میدهد که 0.008 جمعیت دارای این بیماری هستند.**
- ❖ **بعلت نادقیق بودن تست های آزمایشگاهی نتایج آن بصورت زیر است:**
  - ❖ **در 98% مواقعی که شخص واقعا بیمار است نتیجه صحیح مثبت حاصل میشود.**
  - ❖ **در 97% مواقعی که بیمار سالم است نتیجه صحیح منفی حاصل میشود.**
- ❖ **P(cancer)=0.008, P(+|cancer)=0.98, P(+|~cancer)=0.03,**
- ❖ **P(~cancer)=0.992, P(-|cancer)=0.02, P(-|~cancer)=0.97**

❖ حال اگر بیمار جدیدی مشاهده شود که جواب آزمایشگاه مثبت باشد، آیا باید بیمار را مبتلا به سرطان بدانیم؟

❖ احتمال ابتلای بیمار به سرطان عبارت است از:

$$P(cancer|+) = P(+|cancer)\,P(cancer) / P(+) = (0.98)(0.008) / P(+) = 0.0078 / P(+)$$

❖ احتمال نداشتن سرطان عبارت است از:

$$P(\sim cancer|+) = P(+|\sim cancer)\,P(\sim cancer) / P(+) = (0.03)(0.992) / P(+) = 0.0298 / P(+)$$

❖ لذا فرضیه MAP عبارت خواهد بود از:

$h_{map} = \sim cancer$

**Brute-force MAP Learning**

❖ میتوان با استفاده از تئوری بیز، الگوریتمی برای یادگیری مفهوم ارائه نمود که بتواند فرضیه با بیشترین احتمال را بدست دهد:

**Brute-force MAP Learning Algorithm**

❖ برای هر فرضیه h موجود در H مقدار احتمال ثانویه را حساب میکنیم.

❖ فرضیه $h_{MAP}$ را که بیشترین احتمال ثانویه را دارد مشخص میکنیم.

**دسته بندی کننده بیزی بهینه**

Bayes Optimal Classifier

- الگوریتم Brute-Force MAP learning در پی پاسخگوئی به این سوال است: محتمل ترین فرضیه برای مجموعه داده آموزشی چیست؟

- در حالیکه اغلب دنبال یافتن پاسخ این سوال هستیم: محتمل ترین دسته بندی یک نمونه مشاهده شده چیست؟

- اگر چه به نظر میرسد که پاسخ سوال دوم را میتوان با اعمال فرضیه MAP به نمونه مورد نظر بدست آورد، روش بهتری برای اینکار وجود دارد:

در عمل محتمل ترین دسته بندی برای یک نمونه جدید از ترکیب پیش بینی تمامی فرضیه ها بدست می آید. مقدار پیش بینی هر فرضیه در احتمال ثانویه آن ضرب شده و حاصل آنها با هم ترکیب میشود.

- فرض کنید 3 فرضیه h1,h2,h3 برای داده های آزمایشی دارای احتمال ثانویه زیر باشند:

P(h1|D) = 0.4, P(h2|D) = 0.3, P(h3|D) = 0.3

در نتیجه h1 فرضیه MAP میباشد.

- اگربه نمونه جدیدی مثل x بربخوریم که

P(h1) = +, P(h2) = - and P(h3) = -

- در اینصورت احتمال مثبت بودن x برابر با 0.4 و احتمال منفی بودن آن 0.6 است در اینصورت دسته بندی x چیست؟

دسته بندی کننده بیزی بهینه

Bayes Optimal Classifier

- در عمل محتمل ترین دسته بندی برای یک نمونه جدید از ترکیب وزنی پیش بینی تمامی فرضیه ها بدست میاید. اگر دسته بندی مثال جدید بتواند هر مقدار $v_j$ از مجموعه V را

$$P(v_j \mid D) = \sum_{h_i \in H} P(v_j \mid h_i) P(h_i \mid D)$$

**داشته باشد در اینصورت احتمال اینکه مثال جدید دسته بندی $V_j$ را داشته باشد برابر است با:**

- **مقدار ماکزیمم رابطه فوق دسته بندی بهینه این نمونه را مشخص خواهد نمود:**

$$\arg\max_{v_j \in V} \sum_{h_i \in H} P(v_j \mid h_i) P(h_i \mid D)$$

**دسته بندی بهینه**

**Optimal Classification**

- برای مثال فوق دسته بندی بهینه بیزی بصورت زیر خواهد بود.

$P(h1|D) = 0.4 \quad P(-|h1) = 0 \quad P(+|h1) = 1$

$P(h2|D) = 0.3 \quad P(-|h2) = 1 \quad P(+|h2) = 0$

$P(h3|D) = 0.3 \quad P(-|h3) = 1 \quad P(+|h3) = 0$

- لذا

$\Sigma i \; P( + \mid hi ) \; P (hi \mid D) = 0.4$ and

$\Sigma i \; P( - \mid hi ) \; P (hi \mid D) = 0.6$

- در نتیجه این نمونه بصورت منفی دسته بندی خواهد شد.

- ستفاده از این روش برای فضاهای فرضیه های  بزرگ غیرعملی است

**Naive Bayes Classifier**

- ❖ یک روش یادگیری بسیار عملی روش Naive Bayes learner است. در کاربردهائی نظیر دسته بندی متن و تشخیص پزشکی این روش کارائی قابل مقایسه ای دارد.
- ❖ این روش در مسایلی کاربرد دارد که:
  - ❖ نمونه x توسط ترکیب عطفی ویژگیها قابل توصیف بوده و
  - ❖ این ویژگیها بصورت شرطی مستقل از یکدیگر باشند.
  - ❖ تابع هدف f(x) بتواند هر مقداری را از مجموعه محدود v داشته باشد.
  - ❖ مجموعه داده های آزمایشی نسبتاً زیادی در دست باشد .

**Naive Bayes Classifier**

- ❖ تابع هدف زیر را در نظر بگیرید $f : X \rightarrow V$ که در آن هر نمونه x توسط ویژگی زیر مشخص میشود $(a_1, \ldots a_n)$.
- ❖ صورت مسئله: برای یک نمونه مشاهده شده مقدار تابع هدف یا بعبارت دیگر دسته بندی آنرا مشخص کنید.
- ❖ در روش بیزی برای حل مسئله محتمل ترین مقدار هدف $v_{map}$محاسبه میشود:

$$v_{MAP} = \arg\max_{v_j \in V} P(v_j \mid a_1, \ldots, a_n)$$

$$
\begin{aligned}
v_{MAP} &= \underset{v_j \in V}{\arg\max} \frac{P(a_1,\ldots,a_n \mid v_j)P(v_j)}{P(a_1,\ldots,a_n)} \\
&= \underset{v_j \in V}{\arg\max}\, P(a_1,\ldots,a_n \mid v_j)P(v_j)
\end{aligned}
$$

این رابطه با استفاده از تئوری بیز بصورت زیر نوشته میشود :

**Naive Bayes Classifier**

$$v_{MAP} = \underset{v_j \in V}{\arg\max}\, P(a_1,\ldots,a_n \mid v_j)P(v_j)$$

- در رابطه فوق مقدار $P(v_j)$ با شمارش دفعاتی که $v_j$ در مثالهای آموزشی مشاهده شده است محاسبه میشود.
- اما محاسبه $P(a_1,\ldots a_n \mid v_j)$ چندان عملی نیست مگر اینکه مجموعه داده آموزشی بسیار بزرگی در دست باشد.
- روش یادگیری Naive Bayes Classifier بر پایه این فرض ساده (Naive) عمل میکند که:

**مقادیر ویژگیها بصورت شرطی مستقل هستند**

- در اینصورت برای یک مقدار هدف مشخص احتمال مشاهده ترکیب عطفی $(a_1,\ldots a_n)$ برابر است با حاصلضرب احتمال تک تک ویژگیها. در اینصورت رابطه فوق بصورت زیر در می آید:

$$v_{NB} = \underset{v_j \in V}{\arg\max}\, P(v_j)\prod_{i=1}^{n} P(a_i \mid v_j)$$

**Naive Bayes Classifier**

خلاصه:

- در روش یادگیری Naive Bayes Classifier مقادیر مختلف $P(v_j)$ و $P(a_i \mid v_j)$ با استفاده از دفعات تکرار آنها تخمین زده میشود.
- مجموعه این تخمین ها فرضیه ای را تشکیل میدهد که با استفاده از رابطه زیر برای دسته بندی داده جدید بکار میرود:

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_{i=1}^{n} P(a_i \mid v_j)$$

- در این روش هیچگونه عمل جستجوی آشکاری وجود ندارد.
- 