Session 13

Naive Bayes Classifiers

Applied Mathematics for AI | Hobot Academy | Zahra Amini

Telegram: @hobotacademy & Instagram:@hobotacademy
& LinkedIn: @zahraamini-ai

www.hobotacademy.com

Naive Bayes Classifier:

گروهی از classifier های ساده هستند که پایه ی احتمالات تصمیم می گیرند. ← برخلاف سادگی یک روش آماری توی است که با توجه به داده های ورودی و الگوی رفتاری شان classification را انجام می دهد.

تفاوت مهم Naive Bayes Classifier با سایر classifier هاچیست؟

در Naive Bayes دانش اولیه نیز مؤثر است. یعنی چه؟ یعنی ما از تجربه قبلی متخصص یا دانش اولیه ی فضای مسأله هم در تصمیم گیری استفاده می کنیم. ← به این دانش اولیه Prior knowledge می گوییم.

Prior knowlege: اطلاعاتی است که بدون اندازه گیری و بدون توجه به داده ها بدست می آیند.

اطلاعاتی که به صورت تجربی بدست می آیند.

Classifier with Prior knowledge:

فرض کنید مسأله‌ی ما پرتاب یک سکه است و می خواهیم بدانیم هر بار که سکه را پرتاب می کنیم چقدر احتمال دارد شیر باشد یا خط:

سکه:
- شیر $\xrightarrow{P} \frac{1}{2}$ یا 50%
- خط $\xrightarrow{P} \frac{1}{2}$ یا 50%

$$\sum P = 1$$

تاس:
- $1 \xrightarrow{P} \frac{1}{6}$
- $2 \xrightarrow{P} \frac{1}{6}$
- $3 \xrightarrow{P} \frac{1}{6}$
- $4 \xrightarrow{P} \frac{1}{6}$
- $5 \xrightarrow{P} \frac{1}{6}$
- $6 \xrightarrow{P} \frac{1}{6}$

$$\rightsquigarrow \sum P = 1$$

$$P(c) \rightsquigarrow P(c_1), P(c_2), P(c_3), \dots, P(c_n)$$

فرض کنید یک پزشک متخصص بر اساس تجربه اش می گوید ٪80 بیمارانی که مراجعه می کنند سالم هستند و ٪20 شان مبتلا به سرطان هستند.

$$\begin{cases} 1 \quad P(c_1) \geq P(c_2) \\ 2 \quad P(c_2) > P(c_1) \end{cases}$$

1: $P(c_1) \geq P(c_2)$ ✓ $\rightsquigarrow$ $80\% \geq 20\%$ ✓ $\rightsquigarrow$ ٪80 دقت و ٪20 خطا

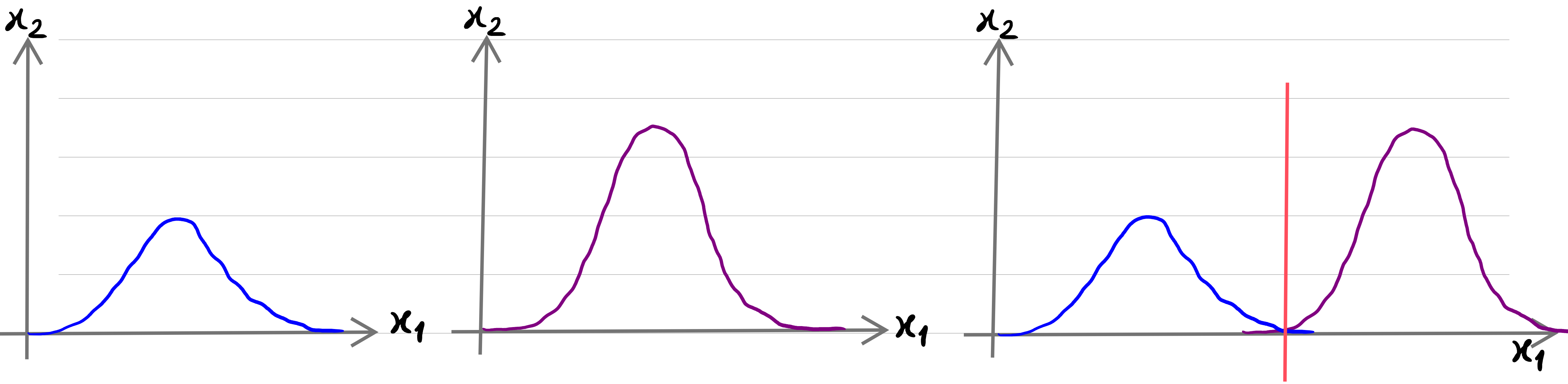
؟: این روش چه زمانی دقت بالایی دارد؟

این روش زمانی با دقت بالا عمل می کند که اختلاف بین $P(C_1)$ و $P(C_2)$ زیاد باشد.

$$P(C_1) \gg P(C_2)$$

یا

$$P(C_2) \gg P(C_1)$$



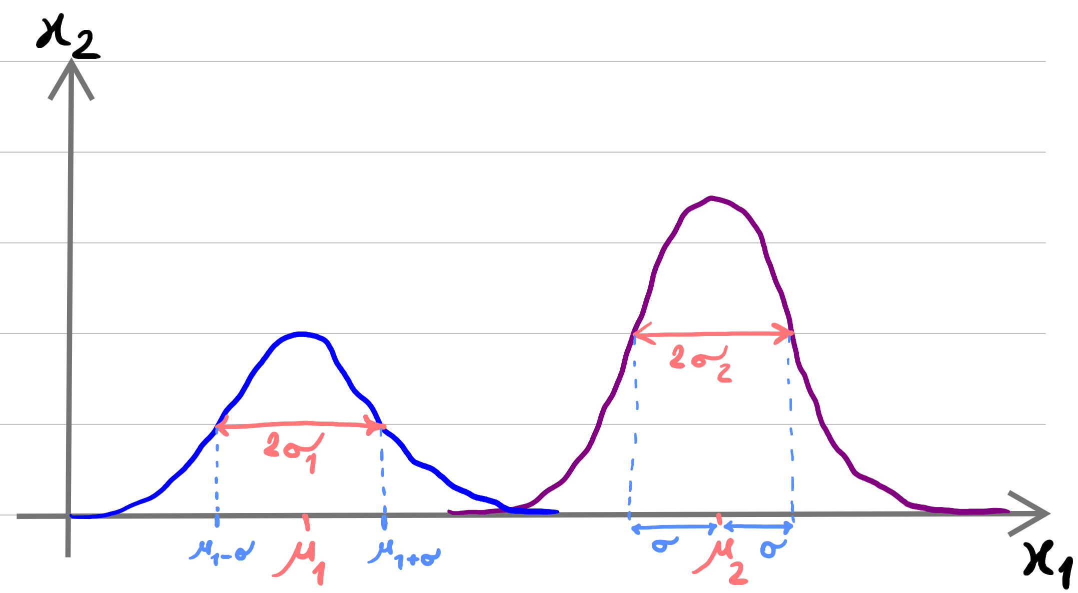
if $P(C_1) = P(C_2) \longrightarrow$ random یا chance

ج: در Naive Bayes Classifier علاوه بر Prior knowledge از Liklihood (شباهت) هم استفاده می‌کنیم.

Liklihood یا شباهت: می‌خواهیم شباهت نمونه‌مان را به هر کلاس محاسبه کنیم.

$$f(x \mid C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right)}$$

$$\begin{cases} 1 & P(C_1) \geq P(C_2) \\ 2 & P(C_2) > P(C_1) \end{cases}$$



؟: حال اگر بخواهیم پارامترهای بهینه ی f را به دست آوریم چه کنیم ؟

$$f(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma_j^2}}\,e^{\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right)}$$

پارامترها μ ، σ

برای به دست آوردن مقدار بهینه ی μ و σ کافیه از تابع هزینه مشتق بگیریم.

$$J=\sum_{i=0}^{m} f(x_i|\mu,\sigma^2)$$

$$\xrightarrow{Ln} \sum_{i=0}^{m} Ln\, f(x_i|\mu,\sigma^2)$$

سپس نسبت به پارامترها مشتق می گیریم و برابر 0 قرار می دهیم:

$$\longrightarrow \frac{\partial}{\partial\mu}J=\frac{\partial}{\partial\mu}\left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\,e^{\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right)}\right)=0 \longrightarrow \hat{\mu}=\frac{1}{m}\sum_{i=0}^{m}x_i$$

$$\longrightarrow \frac{\partial}{\partial\sigma}J=\frac{\partial}{\partial\sigma}\left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\,e^{\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right)}\right)=0 \longrightarrow \hat{\sigma}=\sqrt{\frac{1}{m}\sum_{i=0}^{m}(x_i-\mu)^2}$$

به مجموعه ی مراحلی که طی کردیم برای به دست آوردن پارامترها Maximum Liklihood Estimation (MLE) می گویند.

① داده های هر کلاس چه رفتاری دارند؟ چه توزیع نرمالی دارند؟ μ و σ هر کلاس را به روش MLE ~~محاسبه می کنیم~~ تخمین می زنیم.

← برای تخمین پارامترها از x_train استفاده می کنیم.

$$\hat{\mu} = \frac{1}{m}\sum_{i=0}^{m} x_i \qquad \hat{\sigma} = \sqrt{\frac{1}{m}\sum_{i=0}^{m}(x_i-\mu)^2}$$

② شباهت داده x را نسبت به هر کلاس محاسبه می کنیم ← محاسبه likelihood نمونه تست x به ازای هر کلاس:

$$f(x \mid C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}}\, e^{\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right)}$$

③ نوبتی هم باشه نوبت Bayes classifier ئه:

کلاسبندی بهینه است که از هر دوی دانش اولیه (Prior) و شباهت (Liklihood) در تصمیم گیری استفاده کند.

$$P(c_j|x) = \frac{P(x|c_j) * P(c_j)}{P(x)}$$

$$P(c_j|x) * P(x) = P(x|c_j) * P(c_j) \Rightarrow$$

شباهت Liklihood

Prior

$$P(c_j|x) = \frac{P(x|c_j) * P(c_j)}{P(x)}$$

باعث می شود حاصل $P(c_j|x)$ بین 0 و 1 باشد

ترم Normalize

اما اگر نباشد هم در تصمیم ما تأثیری ندارد.

$$P(x) = \sum_{i=1}^{c} P(x|c_j) * P(c_j)$$

Naive Bayes Classifier:

① محاسبه پارامترها:

$$\text{Prior} \rightsquigarrow P(C_1),\ P(C_2),\ P(C_3),\ \dots,\ P(C_n)$$

$$\hat{\mu} = \frac{1}{m} \sum_{i=0}^{m} x_i \longrightarrow \mu_1,\ \mu_2,\ \mu_3,\ \dots,\ \mu_c$$

$$\hat{\sigma} = \sqrt{\frac{1}{m} \sum_{i=0}^{m} (x_i - \mu)^2} \longrightarrow \sigma_1,\ \sigma_2,\ \sigma_3,\ \dots,\ \sigma_c$$

② محاسبه Liklihood نمونه تست x به ازای هر کلاس:

$$P(x|C_1),\ P(x|C_2),\ P(x|C_3),\ \dots,\ P(x|C_j)$$

$$f(x|C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}}\, e^{\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right)}$$

③ محاسبه احتمال پسین Postrior نمونه تست به ازای هر کلاس:

$$P(C_j|x) = \frac{P(x|C_j) * P(C_j)}{P(x)}$$

⑤ تصمیم گیری: x به کلاسی تعلق دارد که بیشترین احتمال Postrior را نسبت به آن کلاس داشته باشد.

Gender	height (feet)	Weight (Lbs)	Foot size (inch)
M	6	180	12
M	5.9	190	11
M	5.58	170	12
M	5.92	165	10
F	5	100	6
F	5.5	150	8
F	5.42	130	7
F	5.75	150	9

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x_i$$

$$\mu_{h\,male} = \frac{1}{4} \times (6+5.9+5.58+5.92) = 5.855$$

$$\mu_{h\,female} = \frac{1}{4} \times (5+5.5+5.42+5.75) = 5.4175$$

$$Var(x) = Var(x-\mu) = \sigma^2 = \frac{\sum_{i=1}^{m} (x_i-\mu)^2 - \left(\sum_{i=1}^{m} (x_i-\mu)\right)^2 / m}{m-1}$$

$$\sigma^2 = \frac{\overbrace{\cancel{\left[(6-5.855)^2+(5.92-5.855)^2+(5.58-5.855)^2+(5.92-5.855)^2\right]}}^{0.1051} - \overbrace{\cancel{\left[(6-5.855)+(5.92-5.855)+(5.58-5.855)+(5.92-5.855)\right]}}^{0}{}^2/4}{4-1}$$

$$= 0.1051/3 = 0.035033$$

	height (feet)		Weight (Lbs)		Foot size (inch)	
Gender	mean	Variance	mean	Variance	mean	Variance
M	5.855	3.5033×10^{-2}	176.25	1.2292×10^{2}	11.25	9.1667×10^{-1}
F	5.4175	9.7225×10^{-2}	132.5	5.5833×10^{2}	7.5	1.6667

$$x = [6,\ 130,\ 8]$$

$$\text{Posterior (male)} = \frac{P(\text{male})\, P(\text{height}|\text{male}) * P(\text{weight}|\text{male}) * P(\text{foot}|\text{male})}{P(x)}$$

$$\text{Posterior(female)} = \frac{P(\text{female})\, P(\text{height}|\text{female}) * P(\text{weight}|\text{female}) * P(\text{foot}|\text{female})}{P(x)}$$

$$P(x) = \underbrace{P(\text{male})}_{0.5} * \underbrace{P(\text{height}|\text{male})}_{1.5789} * \underbrace{P(\text{weight}|\text{male})}_{5.9887 \times 10^{-6}} * \underbrace{P(\text{foot}|\text{male})}_{1.3112 \times 10^{-3}}$$

$$* \underbrace{P(\text{female})}_{0.5} * \underbrace{P(\text{height}|\text{female})}_{2.23 \times 10^{-1}} * \underbrace{P(\text{weight}|\text{female})}_{1.6789 \times 10^{-2}} * \underbrace{P(\text{foot}|\text{female})}_{2.8669 \times 10^{-1}} = 3.32 \times 10^{-12}$$

$$P(\text{male}) = 0.5$$

$$P(\text{height} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789$$

$\mu = 5.855$ $\quad$ variance $= \sigma^2 = 3.5033 \times 10^{-2}$

$$P(\text{weight} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(130-\mu)^2}{2\sigma^2}\right) = 5.9881 \times 10^{-6}$$

$$P(\text{foot} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(8-\mu)^2}{2\sigma^2}\right) = 1.3112 \times 10^{-3}$$

$$\prod \xrightarrow[x]{} 6.1984 \times 10^{-9}$$

$P(\text{female}) = 0.5$

$$P(\text{height} \mid \text{female}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 2.23 \times 10^{-1}$$

$\mu = 5.4175$ variance $= \sigma^2 = 9.7225 \times 10^{-2}$

$$P(\text{weight} \mid \text{female}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(130-\mu)^2}{2\sigma^2}\right) = 1.6789 \times 10^{-2}$$

$$P(\text{foot} \mid \text{female}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(8-\mu)^2}{2\sigma^2}\right) = 2.8669 \times 10^{-1}$$

$$\prod_x \to 5.3778 \times 10^{-4}$$

Posterior (male) < Posterior (female) $\rightsquigarrow x \in$ Female

Bernouli Naive Bayes:

برای داده‌های discrete یا گسسته استفاده می‌شود.

$$P(x|C_j) = \begin{cases} 1-P & x=0 \\ P & x=1 \end{cases}$$

$$P(x|C_j) = P(x_i|C_j)\, x_i + (1-P(x_i|C_j))(1-x_i)$$

1. Prior
2. Liklihood
3. Posterior
4. Make a decision

ex: بازی گلف

Weather	temperature	humidity	Windy	Play?
Rainy	Hot	yes	No	No
Rainy	Hot	yes	yes	No
Cloudy	Hot	yes	No	yes
Sunny	Moderate	yes	No	yes
Sunny	Cold	No	No	yes
Sunny	Cold	No	yes	No
Cloudy	Cold	No	yes	yes
Rainy	Moderate	yes	No	No
Rainy	Cold	No	No	yes
Sunny	Moderate	No	No	yes
Rainy	Moderate	No	yes	yes
Cloudy	Moderate	yes	yes	yes
Cloudy	Hot	No	No	yes
Sunny	Moderate	yes	yes	No

↳ class

$$X = \begin{array}{cccc} \text{Weather} & \text{tem} & \text{humidity} & \text{windy} \\ \end{array}$$
$$X = [\text{Rainy}, \text{Cold}, \text{yes}, \text{yes}]$$

1:

$$P(\text{yes}) = \frac{9}{14}$$

$$P(\text{No}) = \frac{5}{14}$$

② $P(\text{Weather}=\text{Rainy} \mid \text{class}=\text{yes}) = \frac{2}{9}$ $P(\text{Weather}=\text{Rainy} \mid \text{class}=\text{No}) = \frac{3}{5}$

$P(\text{tem.}=\text{cold} \mid \text{class}=\text{yes}) = \frac{3}{9}$ $P(\text{tem.}=\text{cold} \mid \text{class}=\text{No}) = \frac{1}{5}$

$P(\text{hum.}=\text{yes} \mid \text{class}=\text{yes}) = \frac{3}{9}$ $P(\text{hum.}=\text{yes} \mid \text{class}=\text{No}) = \frac{4}{5}$

$P(\text{windy}=\text{yes} \mid \text{class}=\text{yes}) = \frac{3}{9}$ $P(\text{windy}=\text{yes} \mid \text{class}=\text{No}) = \frac{3}{5}$

③ Postrior

$$P(C_j \mid x) = \frac{P(x \mid C_j) * P(C_j)}{P(x)}$$

$$P(\text{class}=\text{yes} \mid x) = \frac{P(x \mid \text{class}=\text{yes}) * P(\text{yes})}{P(x)} = \frac{\overset{0.005}{\cancel{\frac{2}{9} * \frac{3}{9} * \frac{3}{9} * \frac{3}{9} * \frac{9}{14}}}}{\underset{0.021}{}} = 0.238$$

$$P(x) = P(\text{weather}=\text{Rainy}) * P(\text{tem}=\text{cold}) * P(\text{hum.}=\text{yes}) * P(\text{windy}=\text{yes}) =$$

$$\frac{5}{14} * \frac{4}{14} * \frac{7}{14} * \frac{6}{14} = 0.021$$

$$P(class=No \mid x) = \frac{P(x \mid class=No) * P(No)}{P(x)} = \frac{\frac{3}{5} * \frac{1}{5} * \frac{4}{5} * \frac{3}{5} * \frac{5}{14}}{0.021} = 0.952$$

$$P(class=yes \mid x) < P(class=No \mid x) \Rightarrow x \in No$$

ex: نمونه زیر جزو کدام کلاس است؟

Confident	Studied	Sick	Result
yes	No	No	Fail
yes	No	yes	Pass
No	yes	yes	Fail
No	yes	No	Pass
yes	yes	yes	Pass

$$x = \begin{array}{ccc} \text{confident} & \text{studied} & \text{sick} \\ \end{array}$$
$$x = [\text{yes}, \text{yes}, \text{No}]$$

ex: نمونه زیر جزو کدام کلاس است؟

Confident	Studied	Sick	Result
yes	No	No	Fail
yes	No	yes	Pass
No	yes	yes	Fail
No	yes	No	Pass
yes	yes	yes	Pass

$$X = \begin{array}{ccc} \text{confident} & \text{studied} & \text{sick} \\ \end{array}$$
$$X = [\text{yes}, \text{yes}, \text{No}]$$

① $P(\text{Pass}) = ?\ \frac{3}{5}$ $\qquad$ $P(\text{Fail}) = ?\ \frac{2}{5}$

② $P(\text{confident}=\text{yes} \mid \text{Result}=\text{Pass}) = ?\ \frac{2}{3}$ $\qquad$ $P(\text{confident}=\text{yes} \mid \text{Result}=\text{Fail}) = ?\ \frac{1}{2}$

$P(\text{Studied}=\text{yes} \mid \text{Result}=\text{Pass}) = ?\ \frac{2}{3}$ $\qquad$ $P(\text{Studied}=\text{yes} \mid \text{Result}=\text{Fail}) = ?\ \frac{1}{2}$

$P(\text{Sick}=\text{No} \mid \text{Result}=\text{Pass}) = ?\ \frac{1}{3}$ $\qquad$ $P(\text{Sick}=\text{No} \mid \text{Result}=\text{Fail}) = ?\ \frac{1}{2}$

③ Postrior:

$$P(C_j|x) = \frac{P(x|C_j) * P(C_j)}{P(x)}$$

$$P(Result = Pass \mid x) = \frac{P(x \mid Result = Pass) * P(Result = Pass)}{P(x)} = \frac{\frac{2}{3} * \frac{2}{3} * \frac{1}{3} * \frac{3}{5}}{\frac{18}{125}}$$

$$P(x) = P(confident = yes) * P(studied = yes) * P(Sick = No) \qquad = 0.611$$

$$\hookrightarrow P(x) = \frac{3}{5} * \frac{3}{5} * \frac{2}{5} = \frac{18}{125}$$

$$P(Result = Fail \mid x) = \frac{P(x \mid Result = Fail) * P(Result = Fail)}{P(x)} = \frac{\frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{2}{5}}{\frac{18}{125}}$$

$$= 0.34$$

$$\Rightarrow 0.611 > 0.34 \rightsquigarrow x \in Pass$$