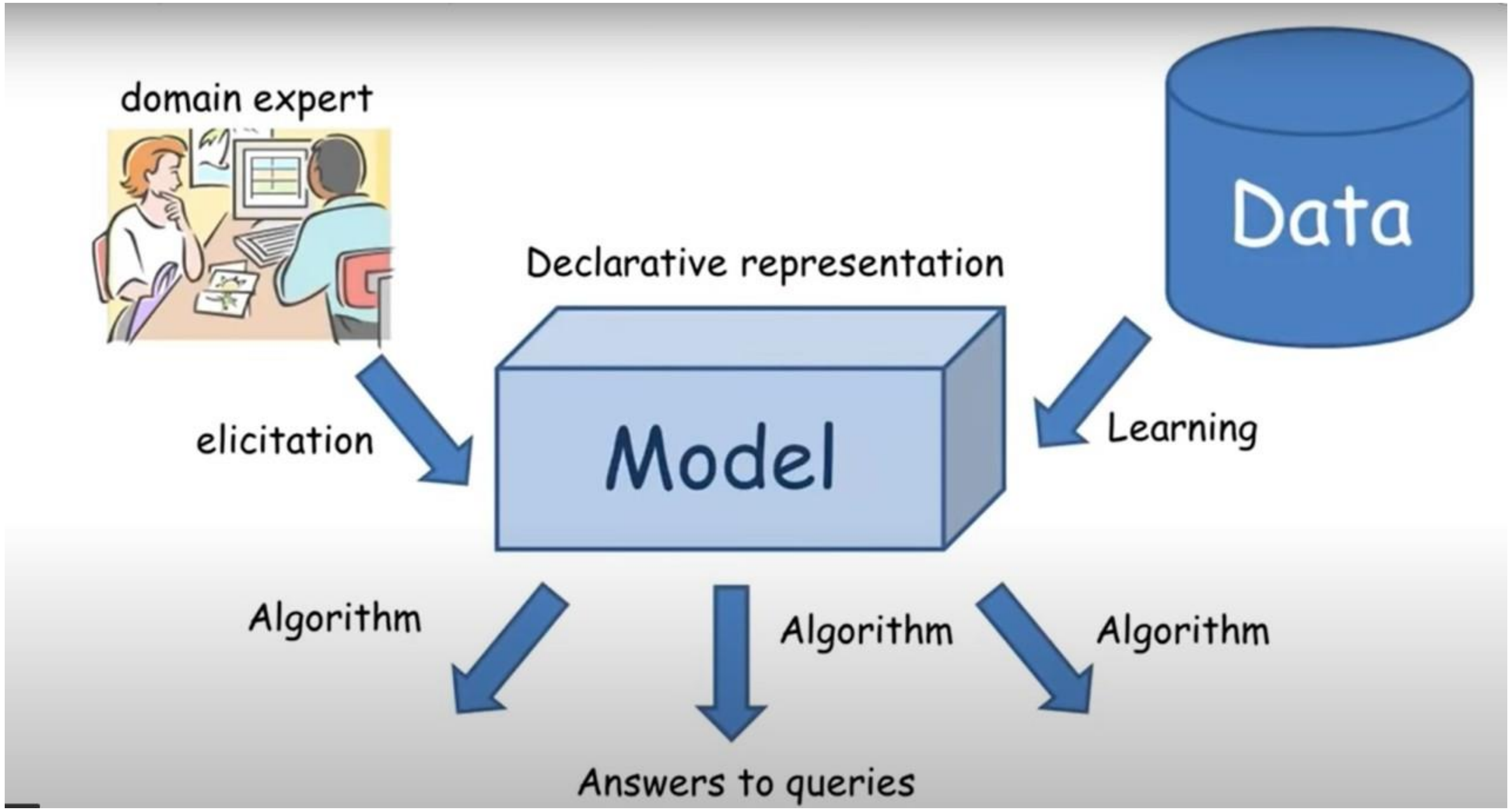
Machine Learning

- 1st Term, 2025/2026
- October 2025

- **Prof. Mohammed A. Al Ghamdi**

Probabilistic Modeling: Review of Statistical Principles



Probabilistic Modeling:

- Suppose that the Learning as a problem of statistical inference.
- Describe the training data as a probability distribution D.
- Learning is then to infer the “best” values of the parameters θ of D given the observed training data.

Review of Some Statistical Principles:

- Bayes Rule.

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

Bayes Rule:

joint distribution

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad \text{Bayes' rule}$$

Thomas Bayes 1701 – 7 April 1761, was an English statistician, philosopher and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: Bayes' theorem.

Example:

Showing a probability distribution for two (or more) random variables (2^3).

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Basic Rule:

$$P(E) = \sum_{\text{rows matching } E} P(\text{row})$$

Find:

P(Female, poor) = ?

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Basic Rule:

$$P(E) = \sum_{\text{rows matching } E} P(\text{row})$$

Find:

P(Female, poor) =
0.253122+0.0421768

P(Male, -40.5) =
0.331313+0.0971295

gender	hours_worked	wealth		
Female	v0:40.5-	poor	0.253122	
		rich	0.0245895	
	v1:40.5+	poor	0.0421768	
		rich	0.0116293	
Male	v0:40.5-	poor	0.331313	
		rich	0.0971295	
	v1:40.5+	poor	0.134106	
		rich	0.105933	

Basic Rule: $$P(E) = \sum_{\text{rows matching } E} P(\text{row})$$

gender	hours_worked	wealth		
Female	v0:40.5-	poor	0.253122	
		rich	0.0245895	
	v1:40.5+	poor	0.0421768	
		rich	0.0116293	
Male	v0:40.5-	poor	0.331313	
		rich	0.0971295	
	v1:40.5+	poor	0.134106	
		rich	0.105933	

Find:

- $P(Male \mid poor) = \dfrac{P(Male, poor)}{P(poor)}$

$$= \frac{(0.331313 + 0.134106)}{(0.253122 + 0.0421768 + 0.331313 + 0.134106)} = \frac{0.465419}{0.7607178} = 0.61181558$$

- $P(Poor, Male) = ?$

- $P(poor \mid Male, -40.5) = = \dfrac{P(poor, (Male, -40.5))}{P(Male, -40.5)} = \dfrac{0.331313}{0.331313+0.0971295} = \dfrac{0.331313}{0.4284425} = 0.7732963.$

Maximum Likelihood Estimation (MLE):

Find out the parameters that maximize the probability of the data.

Maximum Likelihood Estimation (MLE):

- *Given a “biased” coin.*
- *Flipped Multiple times.*
- *Got “α_1” heads and “α_0” tails.*
- *What is probabilities of getting heads ”p(H)” and tails “p(T)”, if we flipped the coin 8 times and getting 5 times heads and 3 times tails ? Simple!.*
- *This is what MLE do.*

Maximum Likelihood Estimation (MLE):

X = 0　　*X = 1*

$P(X=1) = \theta$

Bernoulli

$P(X=0) = 1-\theta$

Each coin flip yields a Boolean value for X

X ~ Bernoulli: $P(X) = \theta^X(1-\theta)^{(1-X)}$

WHEN, X=1, X=0, What is the p(1) and p(0)?

Maximum Likelihood Estimation (MLE):

X ~ Bernoulli: $P(X) = \theta^{X}(1-\theta)^{(1\text{-}X)}$

Now, if we have this sequence (HTT).

What is the Probability of getting it?

H=1, T=0:

$P(1) = \theta^{1}(1-\theta)^{(1-1)} = \theta.$

$P(0) = \theta^{0}(1-\theta)^{(1-0)} = 1-\theta.$

So, $P(HTT) = \theta*(1-\theta)*(1-\theta) = \theta(1-\theta)^{2}.$

In other words;

$P(\alpha 1H, \alpha 0T) = \theta^{\alpha 1}(1-\theta)^{\alpha 0}.$

Maximum Likelihood Estimation (MLE):

$X = 0 \qquad X = 1$

$P(X=1) = \theta$

Bernoulli

$P(X=0) = 1 - \theta$

X ~ Bernoulli: $P(X) = \theta^{X}(1-\theta)^{(1-X)}$

Given a data set D of iid flips: $\alpha1$*s and* $\alpha0$*s;*

$P_{\theta}(D) = \theta^{\alpha1}(1-\theta)^{\alpha0}$

Example -> $P(HTHH) = \theta^{3}(1-\theta)^{1}$

Maximum Likelihood Estimation (MLE):

In math, we use the Differentiation “f’(x)” to find out the maximum value of (x) in the function of x ”f(x)”.

$P(HTHH) = \theta^3\,(1-\theta)^1$

Solve it;

$\theta^3\,(1-\theta)^1 = 0$

$\theta^3 * \theta^4 = 0$

Find Max => $(\theta^3 * \theta^4)' = 0$

$3\theta^2 * 4\theta^3 = 0$

$\theta^2(3-4\theta) = 0$

$3-4\theta = 0$

θ = ¾, we have 3H and 1T.

So, $\theta_{MLE} = \mathrm{argmax}_\theta P_\theta(D) = \alpha 1/(\alpha 1 + \alpha 0)$.

Maximum Likelihood Estimation (MLE):

Remember that, we have only two faces in the coin. What is the MLE if we have more than two faces (i.e. a Die).

$P(HTHH) = \theta^3 (1- \theta)^1$

Solve it;

$\theta^3 (1- \theta)^1 = 0$

$\theta^3 - \theta^4 = 0$

$\text{Find Max} \Rightarrow (\theta^3 - \theta^4)' = 0$

$3\theta^2 - 4\theta^3 = 0$

$\theta^2(3 - 4\theta) = 0$

$3 - 4\theta = 0$

$\theta = 3/4$, we have 3H and 1T.

So, $\theta_{MLE} = \text{argmax}_\theta P_\theta(D) = \alpha 1/(\alpha 1 + \alpha 0)$.

Maximum Likelihood Estimation (MLE):

How to model a k-sided die?

$P(X) = \prod_k \theta_k^{1(x=k)}$

Given a data set D of iid rolls, where,

$P\theta(D) = \prod_{i=1}^{k} \theta_i^{xi}$

So, what is P(1,3,1,4,3,1,6)?

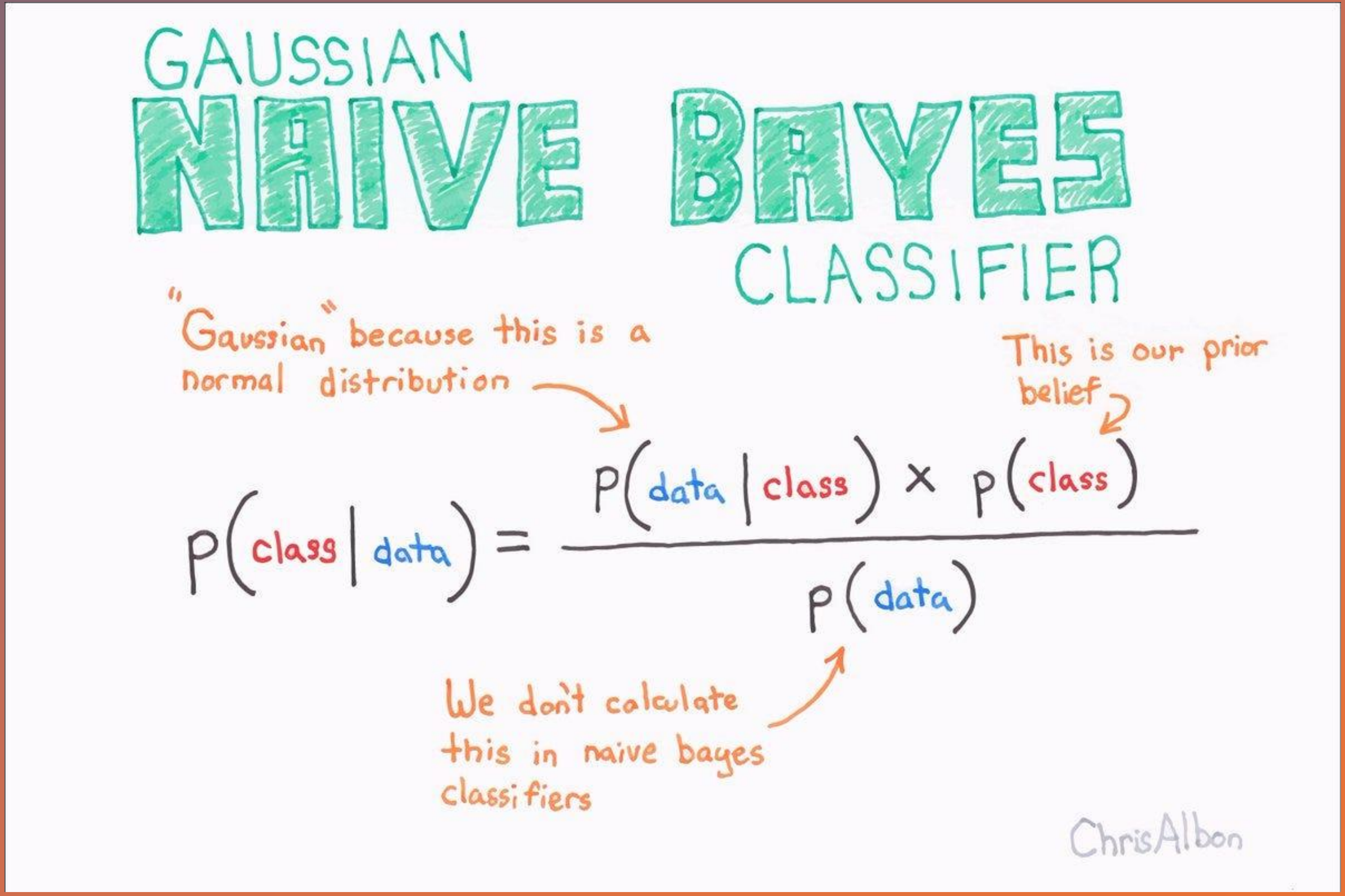
$= \theta_1^3 \, \theta_3^2 \, \theta_4^1 \, \theta_6^1$

So, $\theta_1 = \frac{3}{7}, \theta_2 = \frac{0}{7}, \theta_3 = \frac{2}{7}$

$=> \theta_{I,MLE} = \frac{x_i}{sum(xi)}$

Naïve Bayes:

Bayes' theorem **gives a mathematical rule for inverting conditional probabilities, allowing us to find** the probability of a cause given its effect.



Naïve Bayes:

Conditional Independence Assumption
Features are independent, conditioned on the label.

$$p(x_d \mid y, x_{d'}) = p(x_d \mid y) \quad , \quad \forall d \neq d'$$

$$p_{\boldsymbol{\theta}}((y, \boldsymbol{x})) = p_{\boldsymbol{\theta}}(y) \prod_d p_{\boldsymbol{\theta}}(x_d \mid y, x_1, \ldots, x_{d-1})$$

$$\Downarrow$$

$$= p_{\boldsymbol{\theta}}(y) \prod_d p_{\boldsymbol{\theta}}(x_d \mid y)$$

Naïve Bayes:

Conditional Independence Assumption
Features are independent, conditioned on the label.

$$p_\theta((y, \boldsymbol{x})) = p_\theta(y) \prod_d p_\theta(x_d \mid y)$$

e.g., Bernoulli distribution

$$= \left(\theta_0^{[y=+1]}(1-\theta_0)^{[y=-1]}\right) \prod_d \theta_{(y),d}^{[x_d=1]}(1-\theta_{(y),d})^{[x_d=0]}$$

$$\hat{\theta}_0 = \frac{1}{N}\sum_n [y_n = +1]$$

$$\hat{\theta}_{(+1),d} = \frac{\sum_n [y_n = +1 \wedge x_{n,d} = 1]}{\sum_n [y_n = +1]}$$

$$\hat{\theta}_{(-1),d} = \frac{\sum_n [y_n = -1 \wedge x_{n,d} = 1]}{\sum_n [y_n = -1]}$$

Naïve Bayes:

Steps of Calculating Naïve Bayes Classifier:

Naive Bayes classifier calculates the probability of an event in the following steps:

- Step 1: Calculate the prior probability for given class labels.
- Step 2: Find Likelihood probability with each attribute for each class.
- Step 3: Put these value in Bayes Formula and calculate posterior probability.
- Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

Naïve Bayes:

Example:

Naive Bayes classifier calculates the probability of an event in the following steps:

- Step 1: Calculate the prior probability for given class labels.
- Step 2: Find Likelihood probability with each attribute for each class.
- Step 3: Put these value in Bayes Formula and calculate posterior probability.
- Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

Naïve Bayes:

Example:

X1	X2	X3	…	Xn

Y
yes
no
no
…

Bayes Theorem vs Pythagorean theorem.

General Rule: P(A|B) = $\frac{P(B|A)*P(A)}{P(B)}$

Where;

P(A) is known as prior probability.

P(A|B) is known as posterior probability.

So;

P(Y=yes | X) = $\frac{P(X|Y)*P(Y)}{P(X)}$, where x including x1,x2,x3 … ,xn.

$$= \frac{P(X1|Y)*P(X2|Y) \dots P(Xn|Y)*P(Y)}{P(X1)*P(X2) \dots P(Xn)}$$

Feature independence !!.

Naïve Bayes:

Real Example:

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Find out the probability of playing when the outlook is Sunny, temperature is Cool, humidity is High, and Wind is Strong?

Naïve Bayes:

Real Example:

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Find out the probability of playing when the outlook is Sunny, temperature is Cool, humidity is High, and Wind is Strong?

1) Find the probabilities of the output classes "P(Play=Yes) & P(Play=No)".

Naïve Bayes:

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Real Example:

Find out the probability of playing when the outlook is Sunny, temperature is Cool, humidity is High, and Wind is Strong?

1) **Find the probabilities of the output classes "P(Play=Yes) & P(Play=No)".**

- **P(Play=Yes) = 9/14.**
- **P(Play=No) = 5/14.**

Naïve Bayes:

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Real Example:

Find out the probability of playing when the outlook is Sunny, temperature is Cool, humidity is High, and Wind is Strong?

2) Create the Learning Phase.

Outlook	Play=Yes	Play=No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temp	Play=Yes	Play=No
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=Yes	Play=No
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=Yes	Play=No
Strong	3/9	3/5
Weak	6/9	2/5

Naïve Bayes:

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Prof. Mohammed A. Al Ghamdi, Machine Learning Course

Real Example:

Find out the probability of playing when the outlook is Sunny, temperature is Cool, humidity is High, and Wind is Strong?

2) Start the Execution.

x = (Outlook=Sunny, Temp=Cool, Humidity=High, Wind=Strong).

Play = Yes	Play = No
P(Outlook=Sunny | Play=Yes) = 2/9	P(Outlook=Sunny | Play=No) = 3/5
P(Temp=Cool | Play=Yes) = 3/9	P(Outlook=Cool | Play=No) = 1/5
P(Humidity=High | Play=Yes) = 3/9	P(Humidity=High | Play=No) = 4/5
P(Wind=Strong | Play=Yes) = 3/9	P(Wind=Strong | Play=No) = 3/5

P(Play=Yes | x)*P(Play=Yes) = (2/9*3/9*3/9*3/9)*9/14 = 0.0053
P(Play=No | x)*P(Play=No) = (3/5*1/5*4/5*3/5)*5/14 = 0.0206

Given the fact, P(Play=No | x) > P(Play=Yes | x) => label is “NO”

Example:

sex	age	income	insurance	illness	actdays	hscore	chcond	doctorco	nondocco	hospadmi	hospdays	medecine	prescrib	nonpresc
1	0.19	0.55	levyplus	1	4	1	np	1	0	0	0	1	1	0
1	0.19	0.45	levyplus	1	2	1	np	1	0	0	0	2	1	1
0	0.19	0.9	medlevy	3	0	0	np	1	0	1	4	2	1	1
0	0.19	0.15	medlevy	1	0	0	np	1	0	0	0	0	0	0
0	0.19	0.45	medlevy	2	5	1	la	1	0	0	0	3	1	2
1	0.19	0.35	medlevy	5	1	9	la	1	0	0	0	1	1	0
1	0.19	0.55	medlevy	4	0	2	np	1	0	0	0	0	0	0
1	0.19	0.15	medlevy	3	0	6	np	1	0	0	0	1	1	0
1	0.19	0.65	levyplus	2	0	5	np	1	0	0	0	1	0	1
0	0.19	0.15	levyplus	1	0	0	np	1	0	0	0	1	1	0
0	0.19	0.45	medlevy	1	0	0	np	1	0	0	0	1	1	0
0	0.19	0.25	freerepa	2	0	2	np	1	0	1	80	1	1	0
0	0.19	0.55	medlevy	3	13	1	la	2	0	0	0	0	0	0
0	0.19	0.45	medlevy	4	7	6	la	1	0	0	0	0	0	0
0	0.19	0.25	levyplus	3	1	0	la	1	0	0	0	2	2	0
0	0.19	0.55	medlevy	2	0	7	np	1	0	0	0	3	2	1
0	0.19	0.45	levyplus	1	0	5	np	2	0	0	0	1	1	0
1	0.19	0.45	medlevy	1	1	0	la	1	0	0	0	1	1	0
1	0.19	0.45	levyplus	1	0	0	np	2	0	0	0	1	1	0
1	0.19	0.35	levyplus	1	0	0	np	1	0	0	0	0	0	0
1	0.19	0.45	levyplus	1	3	0	np	1	0	0	0	0	0	0
1	0.19	0.35	levyplus	1	0	1	np	1	0	0	0	2	1	1
0	0.19	0.45	levyplus	2	2	0	np	1	0	0	0	0	0	0
0	0.19	0.55	medlevy	2	14	2	np	1	0	0	0	1	1	0
1	0.19	0.25	freerepa	2	14	11	nla	1	0	1	11	5	5	0

Example:

```
import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

df = pd.read_csv('DoctorAUS.csv')
X=df[['age','income','sex','illness','actdays','hscore','doctorco','nondocco','hospadmi','hospdays','medecine','prescrib']]
y=df['insurance']

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=.3,random_state=0)
clf=GaussianNB()
clf.fit(X_train,y_train)

y_pred=clf.predict(X_test)
print(classification_report(y_test,y_pred))
```

Example:

	precision	recall	f1-score	support
freepor	0.29	0.12	0.16	69
freerepa	0.63	0.52	0.57	321
levyplus	0.50	0.31	0.39	670
medlevy	0.48	0.81	0.60	497
accuracy			0.51	1557
macro avg	0.47	0.44	0.43	1557
weighted avg	0.51	0.51	0.48	1557

Thanks