Thứ
Ngày . . No.

~~Bài~~ Question 11.

Denote: L : Risk Level = Low
H : Risk Level = High.
A : Age
C : CreditScore
E : Education

$P(L) = 0,5$, $P(H) = 0.5$

Since Age, CreditScore, Education are continuous, we use Gaussian distribution for each class.

$$P(x|Class) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

where μ is the mean and σ is the standard deviation for the feature in class.

For Age:
Low: 35, 45, 52, 42

$$\mu = \frac{(35+45+52+42)}{4} = 43,5$$

$$\sigma^2 = \frac{(35-43,5)^2 + (45-43,5)^2 + (52-43,5)^2 + (42-43,5)^2}{4} = 75,75$$

$\sigma = 8,7034.$

High:
$\mu = 30,25$
$\sigma^2 = 14,75$
$\sigma = 3,8406$

Thứ
Ngày
No

T1: Age = 37.

$$P(A=37|L) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}$$

$$= \frac{1}{\sqrt{2\pi \cdot 75{,}75}} e^{\left(\frac{-(37-49{,}5)^2}{2 \cdot 75{,}75}\right)} = 0{,}0347$$

$P(A=37|H) = 0{,}0248$

Credit Score.

Low:

$\mu = 740$

$\sigma^2 = 750$, $\sigma = 27{,}3861$

High:

$\mu = 670$

$\sigma^2 = 350 \Rightarrow \sigma = 18{,}7082$

T1: Credit Score = 705

$P(C=705|L) = 0{,}0064$

$P(C=705|H) = 0{,}0000069$

Education

Low:

$\mu = 16{,}67$

$\sigma^2 = 0{,}8889$ $\sigma = 0{,}9428$

High:

$\mu = 13$

$\sigma^2 = 4$ $\sigma = 2$

$P(E=16|L) = 0{,}3287$

$P(E=16|H) = 0{,}0648$

probability :

$P(L \mid T_1) = p(L) \cdot P(A=37 \mid L) \cdot P(C=705 \mid L) \cdot P(E=16 \mid L)$
$= 0,5 \cdot 0,0347 \cdot 0,0064 \cdot 0,3287 = 0,0000365$

$P(H \mid T_1) = P(H) \cdot P(A=37 \mid H) \cdot P(C=705 \mid H) \cdot P(E=16 \mid H)$
$= 0,5 \cdot 0,0248 \cdot 0,0000069 \cdot 0,0648 = 5,54 \cdot 10^{-9}$

Then Naive Bayes classifier T_1 as Low Risk

Non ~~Ba~~-naive Bayesian approach does not assume independence between features.
Key difference is that non-naive Bayes go along with feature dependencies, while Naive Bayes assumes independences