1. Consider the following dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| N | Color | Type | Origin | Stolen? |
| 1 | red | sports | domestic | yes |
| 2 | red | sports | domestic | no |
| 3 | Red | sports | domestic | yes |
| 4 | Yellow | sports | domestic | no |
| 5 | Yellow | sports | imported | yes |
| 6 | Yellow | SUV | imported | no |
| 7 | Yellow | SUV | imported | yes |
| 8 | Yellow | SUV | domestic | no |
| 9 | Red | SUV | imported | no |
| 10 | Red | sports | domestic | yes |

***Classify (red, SUV, domestic)*** using Naïve Bayes classifier, justify whether the car might be stolen or not

Solution –

P(Stolen=yes)=1/2 ; P(Stolen=no)=1/2 ; P(red|Stolen=yes)=3/5 ; P(red|Stolen=no)=2/5 ; P(SUV|Stolen=yes)=1/5 ; P(SUV|Stolen=no)=3/5 ; P(domestic|Stolen=yes)=3/5 ; P(domestic|Stolen=no)=3/5

(Note: - Do NOT apply Laplace correction, and do NOT compute actual probabilities. Just give the odds for this car to be stolen vs. safe)

P(Stolen=yes | red, SUV, domestic) =

= α P(red|Stolen=yes) P(SUV|Stolen=yes) P(domestic|Stolen=yes) P(Stolen=yes)

= α \* 3/5\*1/5\*3/5\*1/2

= α \* 9/250

P(Stolen=no | red, SUV, domestic)=α P(red|Stolen=no) P(SUV|Stolen=no)P(domestic|Stolen=no) P(Stolen=no)= α \* 2/5\*3/5\*3/5\*1/2

= α \* 18/250

***Odds 2:1 that this car is safe (will NOT be stolen)***

1. How to Estimate Probabilities from the data set given below,

classify for a**) (Refund = No, Married, 120K)** and b)(Refund = Yes, Divorced, 175K)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Tid* | **Refund** | **Marital Status** | **Taxable Income** | **Evade** |
| 1  2  3  4  5  6  7  8  9  10 | Yes No No Yes No No Yes No No  No | Single  Married  Single  Married  Divorced  Married  Divorced  Single  Married  Single | 125K  100K  70K  120K  95K  60K  220K  85K  75K  90K | **No No No No Yes No No Yes No**  **Yes** |

**Solution for (a)**

P(Refund=Yes|No) = 3/7 P(Refund=No|No) = 4/7 P(Refund=Yes|Yes) = 0

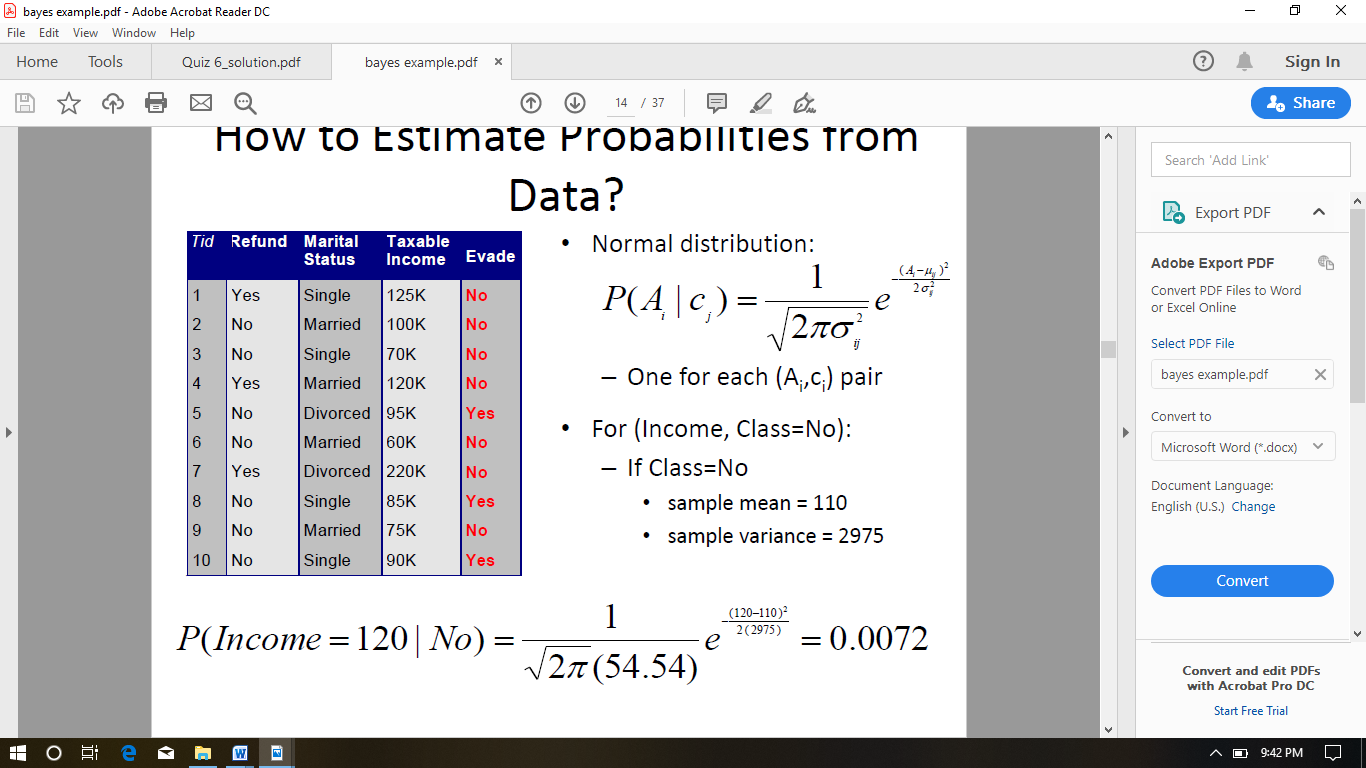
P(Refund=No|Yes) = 1

P(Marital Status=Single|No) = 2/7 P(Marital Status=Divorced|No)=1/7 P(Marital Status=Married|No) = 4/7 P(Marital Status=Single|Yes) = 2/7 P(Marital Status=Divorced|Yes)=1/7 P(Marital Status=Married|Yes) = 0

***For taxable income:***

If class=No: sample mean=110 ; sample variance=2975

If class=Yes: sample mean=90 ; sample variance=25



Using Normal distribution -

**P(X|Class=No)** = P(Refund=No|Class=No) \*P(Married| Class=No) \*P(Income=120K| Class=No)

= 4/7 \* 4/7 \* 0.0072 = 0.0024

**P(X|Class=Yes)**

= P(Refund=No | Class=Yes) \* P(Married | Class=Yes) \* P(Income=120K | Class=Yes)

== 1 \* 0 \* 1.2 \* 10-9 = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes) ; Therefore P(No|X) > P(Yes|X)

**=> Class = No**

1. Find the class to which the below will fall in to, analyses the results

|  |  |  |  |
| --- | --- | --- | --- |
| **Give Birth** | **Can Fly** | **Live in Water** | **Have Legs** |
| yes | no | yes | no |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name** | **Give Birth** | **Can Fly** | **Live in Water** | **Have Legs** | **Class** |
| human | yes | no | no | yes | mammals |
| python | no | no | no | no | non-mammals |
| salmon | no | no | yes | no | non-mammals |
| whale | yes | no | yes | no | mammals |
| frog | no | no | sometimes | yes | non-mammals |
| komodo | no | no | no | yes | non-mammals |
| bat | yes | yes | no | yes | mammals |
| pigeon | no | yes | no | yes | non-mammals |
| cat | yes | no | no | yes | mammals |
| leopard shark | yes | no | yes | no | non-mammals |
| turtle | no | no | sometimes | yes | non-mammals |
| penguin | no | no | sometimes | yes | non-mammals |
| Porcupine | yes | no | no | yes | mammals |
| Eel | no | no | yes | no | non-mammals |
| salamander | no | no | sometimes | yes | non-mammals |
| gila monster | no | no | no | yes | non-mammals |
| Platypus | no | no | no | yes | mammals |
| Owl | no | yes | no | yes | non-mammals |
| Dolphin | yes | no | yes | no | mammals |
| Eagle | no | yes | no | yes | non-mammals |

A: attributes

M: mammals

N: non-mammals

P( A | M ) = 6/7 \* 6/7 \*2/7 \* 2/7 = 0.06

P( A | N ) = 1/13 \*10/13 \* 3/13 \* 4/14 = 0.04

P( A | M )P(M ) = 0.06 \* 7/20=0.021

P( A | N )P(N ) = 0.04 \* 13/20=0.027

**P(A|M)P(M) > P(A|N)P(N)**

* **- Mammals**