Week 5  
**Activity**

You can use Weka, R, Matlab, Excel, etc. You have the following table with some data from a bank, where each row is a person.

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

Suppose you are working for the Internal Revenue Service (SII in Chile) as Data Scientist and your first job is to make a model to know if a person is evading taxes.

Apply Naive Bayes to determine if the IRS should put attention in the following people.  
1) X=(Refund = no, Married, Income = 105 K)   
2) X=(Refund = Yes, Divorced, Income = 127 K)

You answer must include: all probabilities distributions necessaries to calculate the final probabilities (3 and 1 points for questions 1 and 2), the final probabilities (0.5 points per question), and your conclusions/assumptions (0.5 points per question).