I239 – Assignment 4: Naïve bayes

- Question 1: Giving the training data in the table below, predict the class of following new example Using naïve bayes classification

Age > 40, income = medium, student = yes, credit-rating = fair

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

HINTS:

We assume that :

- N = Age > 40, income=medium, student=yes, credit-rating = fair
- N1 = age > 40
- N2 = (income=medium)
- N3 = (student=yes)
- N4 = (credit-rating=fair)

So we need to compare P (Yes | N) and P (No | N)

P(yes | N) = $\frac{P\,(N1 \mid yes)P(yes)\ldots}{P(N)}$

- Question 2: Giving the training data in the table below, predict the class of the following new example using Naïve Bayes classification:

Outlook = overcast, temperature = 60, humidity = 62, windy = yes

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

HINTS:

- We have 2 numeric columns are temperature and humidity, then calculate the probability by 3 formulars below:

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n-1}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Second, the probability for windy and outlook .
- Applying Laplace Smoothing if there is 0 probability.