Naïve Bayes

1) Given dataset Golf with 4 attributes Outlook, Temp, Humidity, Windy and an attribute Play (class).

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

- How Naïve Bayes predicts the class for 4 examples as follows:

Outlook	**Temp**	**Humidity**	**Windy**	**Play**
Overcast	Cool	High	False	?
Rainy	Cool	High	False	?
Sunny	Hot	Normal	False	?
???	Hot	Normal	False	?

2) Let us consider a data table:

Outlook			Temperature		Humidity		Windy			Play	
	yes	*no*	*yes*	*no*	*yes*	*no*		*yes*	*no*	*yes*	*no*
sunny	2	3	83	85	86	85	false	6	2	9	5
overcast	4	0	70	80	96	90	true	3	3		
rainy	3	2	68	65	80	70					
			64	72	65	95					
			69	71	70	91					
			75		80						
			75		70						
			72		90						
			81		75						

- How Naïve Bayes predicts the class for 4 examples as follows:

Outlook	Temp	Humidity	Windy	Play
Overcast	66	80	False	?
Rainy	73	90	False	?
Sunny	80	85	False	?
???	90	85	???	?

3) Implement the program using **GaussianNB** in **scikit-learn** library. The program requires 2 parameters:

- file name of trainset
- file name of testset

The program reports the classification results (accuracy, confusion matrix) for 5 datasets:

- Iris (.trn: trainset, .tst: testset)
- Optics (.trn: trainset, .tst: testset)
- Letter (.trn: trainset, .tst: testset)
- Leukemia (.trn: trainset, .tst: testset)
- Fp (.trn: trainset, .tst: testset)