1. Approximately 1% of women aged between 40 and 50 have breast cancer. 80% of mammogram screening tests detect breast cancer when it is there. 90% of mammograms DO NOT show breast cancer when it's **NOT** there[1]. Based on these information complete the following table.

Cancer	Probability
No	99%
Yes	1%

Cancer	Test	Probability
Yes	Positive	80%
Yes	Negative	?
No	Positive	?
No	Negative	90%

$\rightarrow$ $P(P|C) = 0.8$ TP

$$\begin{cases} P(N|C) = 1 - P(P|C) = 0.2 & \text{FN} \\ P(P|NC) = 1 - P(N|NC) = 0.1 & \text{FP} \end{cases}$$

$P(N|NC) = 0.9$ TN

2. Based on the results in question 1, calculate the **marginal probability** of 'positive' results in a Mammogram Screening Test.

$P(P) = ?$

$\Rightarrow$ Law of total probability

$$P(p) = \sum_{i \in \{C, NC\}} P(P|i)\,P(i) = \sum_{i \in \{C, NC\}} P(p, i)$$

$$= P(p|C)\,P(C) + P(p|NC)\,P(NC)$$

$$= 0.8 \times 0.01 + 0.1 \times 0.99$$

$$= 0.008 + 0.099$$

$$= 0.107$$

3. Based on the results in question 1, calculate *P(Cancer = 'Yes' | Test = 'Positive')*, using the Bayes Rule.

$$P(c|p) = \frac{P(p|c)\,P(c)}{P(p)}$$

$$= \frac{0.8 \times 0.01}{0.107}$$

$$= 0.075 = 7.5\%$$

Positive test result ⇒ 7.5% of having cancer

4. What is optimization? What is a "loss function" ?

optimization in ML: finding the optimal parameters of the model that give us the most accurate results (predictions)

loss function: we define a function that best describe our undesirable outcomes (errors) for each model.
(cost function)

5. For the following set of classification problems, design a Naive Bayes classification model. Answer the following questions for each problem: (1) what are the instances, what are the features (and values)? (2) explain which distributions you would choose to model the observations; and (3) explain the significance of the Naive Bayes assumption.

(i). You want to classify a set of images of animals in to 'cats', 'dogs', and 'others'.

(1) Instances: images

Features: pixels of images

values: $\underbrace{\text{intensity} \quad \text{color} \quad \text{shade}}_{\text{continuous}}$

(2) Gaussian distribution for each feature value: intensity, color, ...

(continuous feature, assume feature values are normally distributed)

(3) NB assumption: Conditional independence assumption (among features)

$$P(\underbrace{t_1, \dots, t_k}_{\text{features}} \mid c) = \prod_{i=1}^{k} (t_k \mid c)$$

class (e.g. cat)

⇒ Not true in the reality: intesity, color, ..., of neighboring pixels depend on one another

⇒ Not independant, but we can still use NB to make predictions (ii). You want to classify whether each customer will purchase a product, given all the products (s)he has bought previously.

(1) Instances : Customer

Features : Products (binary , or # of times the customer bought)

(2) K - dim multinomial distribution : values are the counts of purchases of each product

(or bernoulli dist for each variable)

(3) Under NB assumption : the purchases of products are independant

Not independent ⇒ e.g. (Milk & bread)