School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 1, 2020)
Sample Solutions: Week 3

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 is **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%

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

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

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

5. For the following set of classification problems, we want to design a Naive Bayes classification model.

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

   (ii). You want to classify whether each customer will purchase a product, given all the products (s)he has bought previously.

   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.

---

[1] Remember these numbers are not accurate and simplified to ease the calculations in this question.