General notes on machine learning

Training and testing data

- Training and predictions should be done on different sets of data. If training and predictions are done on same type of data, then there will be overfitting of the data
- Generalizing to new data is not possible if the training and prediction are both on the same dataset
- Save 10% of the dataset to be used as

Unpacking NB using Bayes Rule

Quiz : Prior and Posterior

- Bayes rule
  - prior probability + test evidence = posterior probability **semantically Bayes incorporates some evidence from the test into prior probability to arrive at a posterior probability.**

  1. PRIOR
     - P(C) = 0.01 = 1% **prior probability of cancer**
     - P(Pos|C) = 0.9 = 90% **prior probability of the test being positive and the persion having cancer**
     - P(Neg|ˆC) = 0.9 = 90% **prior probability of the test being negative and the person not having cancer**
     - P(ˆC) = 0.99 = 99% **prior probability of not having cancer**
     - P(Pos|ˆC) = 0.1 **prior probability of the test being positive and the person not having cancer**
  2. POSTERIOR
     - P(C | Pos) = P(C) X P(Pos | C) = 0.009 = 0.9%
       - **P(C|Pos) is posterior of the probability of cancer given that the test says positive**
       - ___P(C) is prior probability of cancer
       - **P(Pos | C) is probability of positive result given that a person has cancer. This is the test sensitivity.**
     - P(ˆC | Pos) = P(ˆC) X P(Pos | ˆC) = 0.099 = 9.9%
       - **P(ˆC|Pos) is posterior of the probability of not having cancer given that the test says positive**
       - ___P(ˆC) is prior probability of not having cancer
       - **P(Pos | ˆC) is probability of positive result given that a person does not have cancer.**

[fig1] [fig1]: nb1.png

```
* But probabilities do not add upto 1
```