Which are different Bayesian Approaches? Explain in detail.

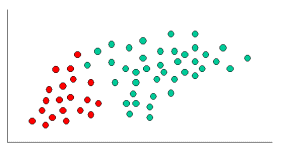
 Bayesian approaches are a fundamentally important DM technique. Given the probability distribution, Bayes classifier can provably achieve the optimal result.     Bayesian method is based on the probability theory. Bayes Rule is applied here to calculate the posterior from the prior and the likelihood, because the later two is generally easier to be calculated from a probability model.

        One limitation that the Bayesian approaches can not cross is the need of the probability estimation from the training dataset. It is noticeable that in some situations, such as the decision is clearly based on certain criteria, or the dataset has high degree of randomality, the Bayesian approaches will not be a good choice.

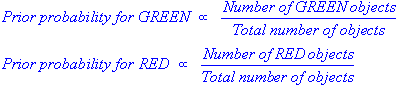
*My introduction slides of Bayesian Approaches.* [[pdf](http://research.cs.queensu.ca/home/xiao/doc/dm/Bayes.pdf)]

**Naïve Bayesian Classifiers**

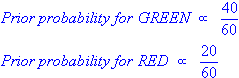
        The Naïve Bayes Classifier technique is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. The following example is a simple demonstration of applying the Naïve Bayes Classifier from [StatSoft](http://www.statsoft.coml/).

        As indicated at Figure 1, the objects can be classified as either GREEN or RED. Our task is to classify new cases as they arrive (i.e., decide to which class label they belong, based on the currently exiting objects).   
            *Figure 1. objects are classified to GREEN or RED.*

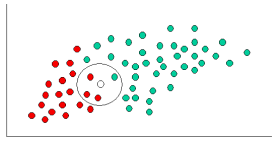
        We can then calculate the priors (i.e. the probability of the object among all objects) based on the previous experience. Thus:



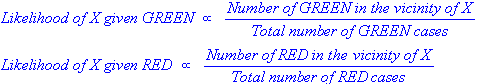
        Since there is a total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class membership are:



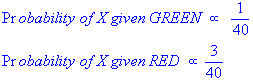
        Having formulated our prior probability, we are now ready to classify a new object (WHITE circle in Figure 2). Since the objects are well clustered, it is reasonable to assume that the more GREEN (or RED) objects in the vicinity of X, the more likely that the new cases belong to that particular color. To measure this likelihood, we draw a circle around X which encompasses a number (to be chosen a priori) of points irrespective of their class labels. Then we calculate the number of points in the circle belonging to each class label.

 *Figure 2. classify the WHITE circle.*

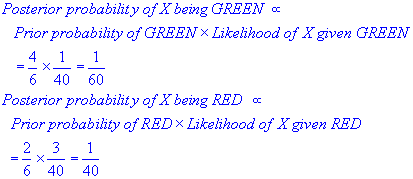
    We can calculate the likelihood:

**

In Figure 2, it is clear that Likelihood of X given RED is larger than Likelihood of X given GREEN, since the circle encompasses 1 GREEN object and 3 RED ones. Thus:



        Although the prior probabilities indicate that X may belong to GREEN (given that there are twice as many GREEN compared to RED) the likelihood indicates otherwise; that the class membership of X is RED (given that there are more RED objects in the vicinity of X than GREEN). In the Bayesian analysis, the final classification is produced by combining both sources of information (i.e. the prior and the likelihood) to form a posterior probability using Bayes Rule.



        Finally, we classify X as RED since its class membership achieves the largest posterior probability.

**Gaussian Bayesian Classifiers**

The problem with the Naïve Bayes Classifier is that it assumes all attributes are independent of each other which in general can not be applied. Gaussian PDF can be plug-in here to estimate the attribute probability density function (PDF). Because the well developed Gaussian PDF theories, we can classify the new object easier through the same Bayes Classifier Model but with certain degree recognition of the covariance. Normally, this gives more accurate classification result.

        I guess one question to be asked is why Gaussian? There are many other PDF's can be applied. But from statistic point of view, many real world distributions are more likely to be estimated by Gaussian PDF than others. If you are familiar with Information Theory, the Gaussian gives the maximum entropy for an unbounded range, which means Gaussian has more ability to estimate the randomality.

*How to apply the Gaussian to the Bayes Classifier?*

        The application here is very intuitive. We assume the Density Estimation follows a Gaussian distribution. Then the prior and the likelihood can be calculated through the Gaussian PDF. The critical thing here is to identify the Gaussian distribution (i.e. find the mean and variance of the Gaussian). The following 5 steps are a general model to initialize the Gaussian distribution to fit our input dataset.

1. Choose a probability estimator form (Gaussian)
2. Choose an initial set of parameters for the estimator (Gaussian mean and variance)
3. Given parameters, compute posterior estimates for hidden variable
4. Given posterior estimates, find distributional parameters that maximize expectation (mean) of joint density for data and hidden variable (Guarantee to also maximize improvement of likelihood)
5. Assess goodness of fit (i.e. log likelihood) If not stopping criterion, return to (3).

        From research perspective, Gaussian may not be the only PDF to be applied to the Bayes Classifier, although it has very strong theoretical support and nice properties. The general model of applying those PDF's should be the same. The estimation results highly depend on whether or how close a PDF can simulate the given dataset.

*Some normal used PDF's are listed below: (just to refresh our statics)*

