**Naïve Bayes Classifiers:**

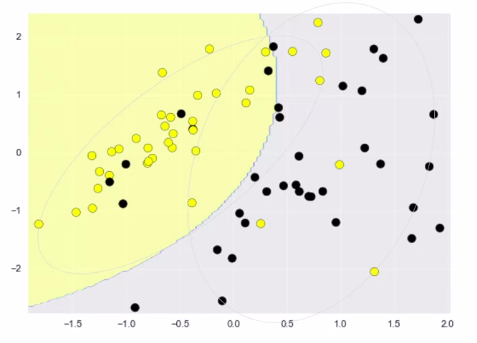
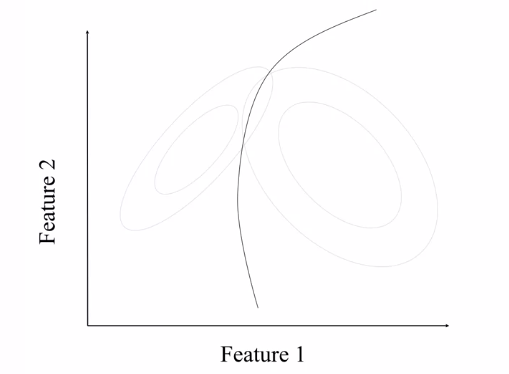
This method is based on a **simple probabilistic** model of how the data in the classes might have been generated. The method is called naïve because the model assumes that **each feature is independent of all the others**. E.g. some feature that might be correlated are the size of the bedroom area and the land area. This method can be highly **efficient at learning and predictions** but can tend to **generalize poorly** than more sophisticated learning methods. However, this method can be competitive for some tasks.

**Naïve Bayes Types:**

* **Bernoulli**: This method uses a set of binary occurrence features. This is **useful in classification of text** documents as it will return a binary value depending on if the word is present or not (but not how often).
* **Multinomial**: This model uses a count base feature which does account for how many times a particular feature such as a word is observed in training. **Useful for text classification**.
* **Gaussian**: Useful for **continuous/real-valued** features. During training this model estimates the mean and standard deviation for each feature of each class. Once calculated the model tries to match the predicted statistics to the best class that matches these values. This model assumes that the data for each class was generated by a simple class specific Gaussian distribution. Predicting a class from a new data point is determined by the probability that that gaussian distribution of the class was most likely to have generated the new data point.

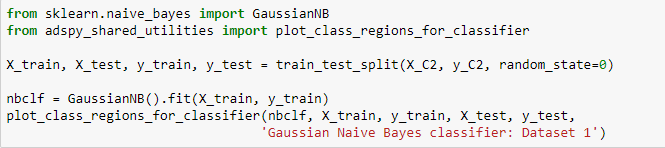
**Gaussian Naïve Bayes Classifier:**

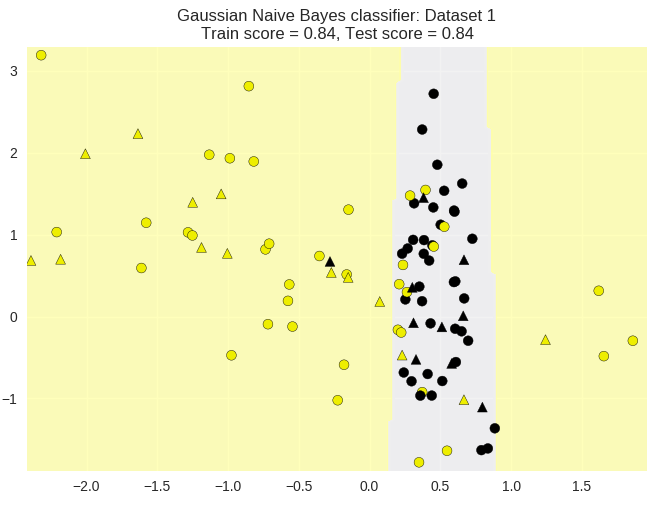
The decision boundary between two classes in this model can be thought of as being a parabolic curve. For the special case where the variance of these features is the same for both classes, the decision boundary will be linear.



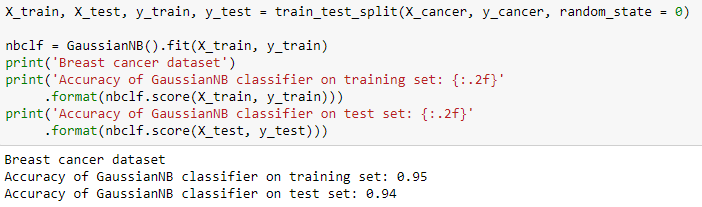
The grey ellipses give an example of the **distribution of the data for each class**. Where the centres of the distributions correspond to the **mean values** for each class. The outer grey line shows the distribution for a limit up to **2 standard deviations from the mean value**, and finally the separation line is the **decision boundary**.

**Example:**





We can see that the Gaussian classification model performs quite well on this simple data set. We can see from using real world data (cancer data) that the model can perform very well:



**The Gaussian method performs well on data sets that has a very high number of features. E.g. 100 features. Likewise, the Bernoulli and Multinomial are used for text classification where there are a very large number of distinct words as features and where the feature vectors are sparse.**

