|  |  |  |
| --- | --- | --- |
| Naive Bayesian Classifier |  |  |
| The Naive Bayesian classifier is based on Bayes’ theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods. |  |  |
|  |  |  |
| **Algorithm** |  |  |
| Bayes theorem provides a way of calculating the posterior probability, *P*(*c|x*), from *P*(*c*), *P*(*x*), and *P*(*x|c*). Naive Bayes classifier assume that the effect of the value of a predictor (*x*) on a given class (*c*) is independent of the values of other predictors. This assumption is called class conditional independence. |  |  |
| IMG_256 |  |  |
| * *P*(*c|x*) is the posterior probability of *class* (*target*) given *predictor* (*attribute*). * *P*(*c*) is the prior probability of *class*. * *P*(*x|c*) is the likelihood which is the probability of *predictor* given *class*. * *P*(*x*) is the prior probability of *predictor*. |  |  |
| In ZeroR model there is no predictor, in OneR model we try to find the single best predictor, naive Bayesian includes all predictors using Bayes' rule and the independence assumptions between predictors. |  |  |
|  |  |  |
| *Example 1*: |  |  |
| We use the same simple Weather dataset here. |  |  |
| IMG_257 |  |  |
| The posterior probability can be calculated by first, constructing a frequency table for each attribute against the target. Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction. |  |  |
| IMG_258 |  |  |
| The likelihood tables for all four predictors. |  |  |
| IMG_259 |  |  |
|  |  |  |
| *Example 2*: |  |  |
| In this example we have 4 inputs (predictors). The final posterior probabilities can be standardized between 0 and 1. |  |  |
| IMG_260 |  |  |