Lecture 12 Naïve Bayes

* Bayesian Classifier
  + A Bayesian classifier uses Bayes' Theorem to predict the most likely class for an unknown instance, given its attribute values
* Features X = (X1… Xn) set of features
* Label Y = y
* Find the value of y for which the conditional probability P(Y|X) is maximized so we can say that for a particular value of y that the expression becomes maximized, and the class label should be y
  + Denominator stays the same no matter the value of y because the evidence (set of features) always remains the same
* Naïve Bayes
  + Need to assume x1 and x2 are independent of each other
    - Allows us to avoid searching for this in our dataset and consider the probability to be a product
    - A screenshot of a black background with numbers and equations

      Description automatically generated
    - So we can see that the class label one maximizes the probability meaning that it is the more desirable category
  + Continuous Attributes
    - Binning / 2-way or multi-way split
      * Create new attribute for each bin
      * Issue is that these attributes are no longer independent
    - Pdf estimation
      * Assume attribute follows a particular distribution (example: normal)
      * Use data to estimate the parameters of the distribution