Data Mining, CSCI 347, Fall 2019
Simple Probabilistic Modeling and Naïve Bayes, Sept. 9

Naïve Bayes

- Predict based on all attributes being independent and contributing equally
- Bayes Theorem: Pr[H|E] = Pr[E|H]*Pr[H] / Pr[E]
    - H is the hypothesis
    - E is the evidence
    - In the Bayes Theorem Pr[H] is known as the prior probability of the hypothesis H and the Pr[E] in the denominator can be ignore because we'll normalize the probabilities in the end
- Read Pr[X] - the probability that X occurs
- Read Pr[X|Y] – the probability that X occurs given that the condition Y occurred
- When E contains multiple attributes Bayes Theorem becomes:

  $Pr[H|E] = Pr[E_1|H]*Pr[E_2|H]* \ldots *Pr[E_n|H]*Pr[H] \;/\; Pr[E]$
    - Pr[E] in the denominator can be ignored because we'll normalize the probabilities at the end.

Normalization in statistics – shift and scale numbers to make them easier to compare.

We'll turn the numbers into percentages by dividing the 'part' by the 'whole'.

Given values a and b, normalize them to a/(a+b) and b/(a+b)

Laplace estimators:

- Note that in Bayes Theorem if one or more $Pr[E_j|H]$ is 0, the entire product will be 0, giving that "0" veto power over the other probabilities.
- To avoid this, add 1 to the numerator of each $Pr[E_j|H]$ and n to the denominator, where n is the number of values for the attribute. Overall, what is added must sum to 1. This is known as a Laplace estimator.
- Didn't need to add 1, could have added some other number μ, as long as overall, the numbers that are added sum to 1. If μ is large the priors are important compared with the new evidence coming in from the training set. If μ is small the priors have less influence.
- We don't have to make these weights even. Fully Bayesian formulation has prior probabilities assigned to everything. Knowing what weights to assign is difficult.