M-Estimate of Probabilities for Naive Bayes

- Let n be the number of training examples with class c. For attribute a, let n_v be the number of training examples with class c and $a = v$.

- We defined

$$P(a = v \mid \text{class} = c) \approx \frac{n_v}{n_c} \qquad (1)$$

- Define ***m*-estimate of probability** as:

$$P(a = v \mid \text{class} = c) \approx \frac{n_v + mp}{n_c + m} \qquad (2)$$

  where p = prior estimate of $P(a = v \mid \text{class} = c)$ and m is the weight given to p.

- The idea here is to pretend that we have an extra m training examples with class = c, with $p \times m$ of them having $a = v$.

- Usually, unless you have good prior information, set p = 1/(*number of values of attribute a*). m is usually set to 2.

- For each attribute value, use equation (2) instead of equation (1) for Naive Bayes.

Binning for Naive Bayes

Method 1: For each attribute, put probability values (m-estimates) in k approximately equal-sized bins.

On Homework: We are using 17 bins (one for each attribute value). We may experiment with using fewer bins.

Method 2: Let n = number of training examples. For each attribute, create $\approx\sqrt{n}$ bins. Sort values of attribute in ascending order, and put $\approx\sqrt{n}$ of them in each bin. (We won't use this in the homework.)