Bayesian classifier

3 Nov 2018
Bayesian classifier.docx

Definitions

a_j = character value for character *j*, forming the vector **a**
T_i = taxon *i*
n = number of taxa
q = number of characters

Data

$P(T_i | a_j)$ = probability of taxon *i* having a particular specimen with character value a_j for character *j.* This forms the matrix **X** with *n* rows (taxa) and *q* columns (characters).
$P(T_i)$ = prior probability of a random specimen belonging to taxon *i.* This forms the vector **T** with *n* elements.

Calculated Quantities

$P(a_j | T_i)$ = probability of a particular specimen with character value a_j in taxon *i*
$P(a_j)$ = probability of observing character state *a* for character *j* .

For one character *j*, by Bayes theorem:

$$P(T_i|a_j) = \frac{P(a_j|T_i)\,P(T_i)}{P(a_j)} \text{ and } P(a_j) = \sum_{i=1}^{n} P(a_j|T_i)P(T_i)$$

or

$$P(T_i|a_j) = \frac{P(a_j|T_i)\,P(T_i)}{\sum_{i=1}^{n} P(a_j|T_i)P(T_i)}$$

Note that the numerator is an unscaled quantity that is proportional to $P(T_i|a_j)$ and that dividing by the sum transforms the numerator into a probability. The resulting probabilities sum to 1 across all taxa:

$$\sum_{i=1}^{n} P(T_i|a_j) = 1 \text{ and } 0 \le P \le 1$$

For multiple characters j, and assuming independence of characters:

$$P(T_i|\mathbf{a}) = \frac{\prod_{j=1}^{q} P(a_j|T_i)\,P(T_i)}{\sum_{i=1}^{n} \prod_{j=1}^{q} P(a_j|T_i)P(T_i)}$$

For large data sets the products of probabilities can become very small, so that calculations need to be done as logarithms:

$$P(T_i|\mathbf{a}) = = \left[\frac{\log^{-1}\left[\sum_{j=1}^{q} \log P(a_j|T_i)\ \log P(T_i)\right]}{\sum_{i=1}^{n} \log^{-1}\left[\sum_{j=1}^{q} \log P(a_j|T_i)\ \log P(T_i)\right]}\right]$$

Verification checks:

$$\sum_{i=1}^{n} P(T_i|\mathbf{a}) = 1 \text{ and } 0 \le P \le 1$$