Bioinformatics III

Third Assignment

Thibault Schowing (2571837)
Wiebke Schmitt (2543675)

April 27, 2018

Exercise 3.1:

(a) *Given the states of the features, you want to infer if two proteins are likely to physically interact. In practice, log-likelihood ratios are used in binary classification:*

$$log\frac{P(C|S)}{P(\bar{C}|S)}$$

Derive a term that uses observable probabilities such as $P(S_i|C)$ to calculate the loglikelihood ratio from training data. How does the actual classification work?

First we have:

$$P(S_i|C) = \frac{P(C|S_i)P(S_i)}{P(C)}$$

And:

$$P(S_i|\bar{C}) = \frac{P(\bar{C}|S_i)P(S_i)}{P(\bar{C})}$$

Then we develop the desired final output

$$\frac{P(S_i|C)}{P(S_i|\bar{C})} \Longleftrightarrow \frac{P(S_i|C)P(C)}{P(S_i|\bar{C})P(\bar{C})} = \frac{P(C|S_i)P(S_i)}{P(\bar{C}|S_i)P(S_i)} = \frac{P(C|S_i)}{P(\bar{C}|S_i)}$$

$$log\frac{P(C|S)}{P(\bar{C}|S)} = log\prod_i^n \frac{P(S_i|C)P(C)}{P(S_i|\bar{C})P(\bar{C})} = \sum_i^n log\frac{P(S_i|C)P(C)}{P(S_i|\bar{C})P(\bar{C})} = \bigwedge(C|S)$$

(b) *bla instructions*

(c)

Exercise 3.2: Classify real-world network examples

(a)

(b)

(c)