20180118 COGS 118a Lecture Notes

Cabinet COGS118a Lecture Notes

Modeling

Y are labels
X are observations

Storing all features for all observations is infeasible (combinatoric explosion).

Conditional Probabilities

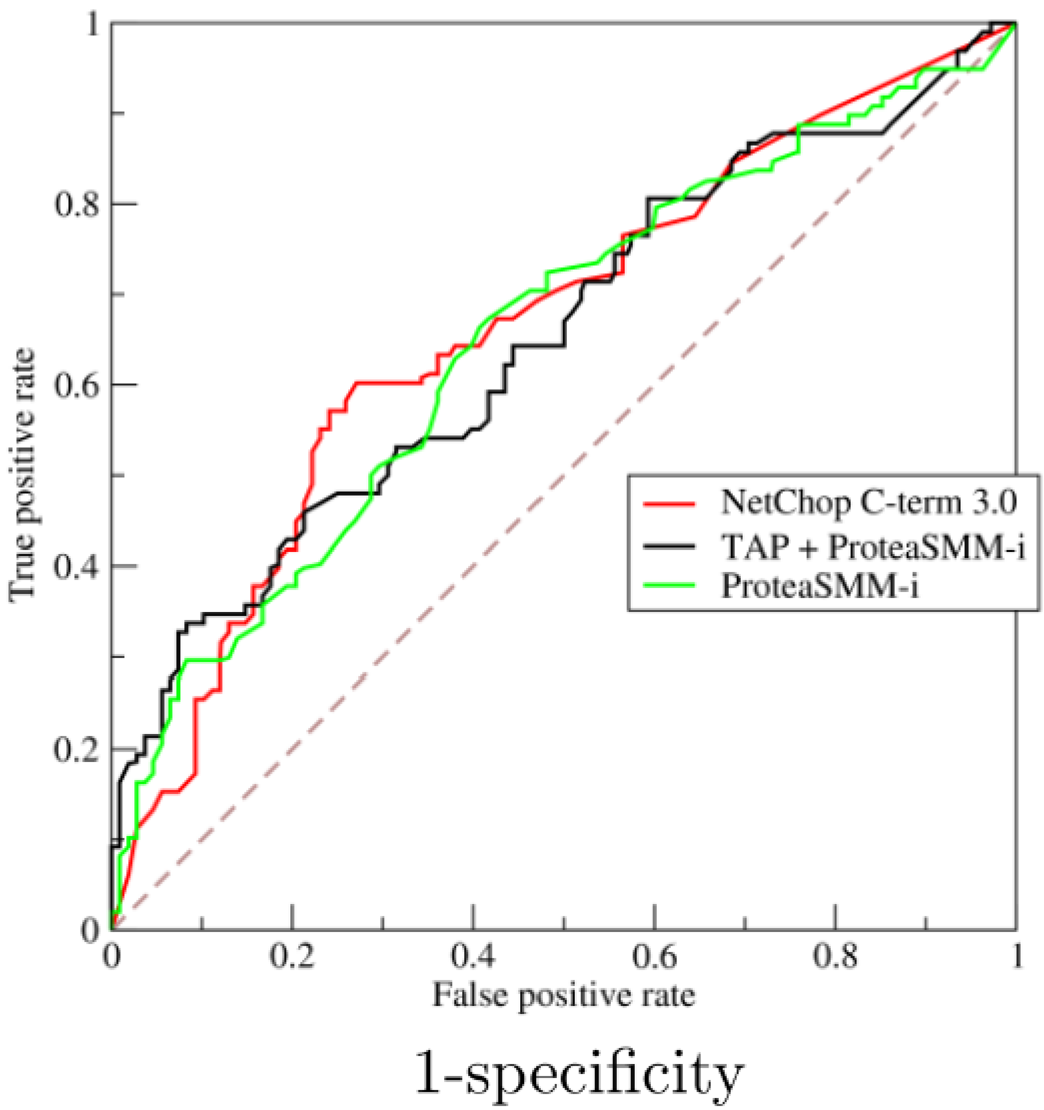
What is the probability that a person has a disease given that they tested positive?

Error Measures and Metrics

True Positive Rate (correctly identified): $P(test+|sick+)$ a.k.a $sensitivity$ a.k.a $recall$
False Positive Rate (incorrectly identified): $P(test+|sick-)$
True Negative Rate (correctly rejected): $P(test-|sick-)$ a.k.a. $specificity$
False Negative Rate (incorrectly rejected): $P(test-|sick+)$

Typically we want high rates for True Positive and True Negatives, and low values for False Positives and False Negatives

Reciever Operating Characteristic (ROC)



Notice, as we increase sensitivity (True positivity rate) we also increase 1 - specificity (False positivity rate). At the extreme, we classify everything as positive (which gives a 1 for true positive rate, but also a 1 for false positive rate).

The ideal location on the curve is top-left (high sensitivity, high specificity).

> This curve is biased by the ratio between “sick” and “healthy”

Precision and Recall

Precision: $\frac{P(sick+ \ and \ test+)}{P(total test+)}$

Recall(sensitivity): $\frac{P(sick+\ and\ test+)}{P(totalsick+)}$

These two are inversely related: as one increases the other decreases.

F-Score

$$\frac{2 \times (Precsision \times Recall)}{Precision + Recall}$$

This is a single metric for quantifying the performance of the classsifier.

Error

A measure of how many times we mispredicted a classification.

$$e_{training} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(y_i \neq f(\mathbf{x}_i)) \qquad \mathbf{1}(z) = \begin{cases} 1 & if\ z = TRUE \\ 0 & otherwise \end{cases}$$

$$e_{testing} = \frac{1}{q} \sum_{i=1}^{q} \mathbf{1}(y_i \neq f(\mathbf{x}_i))$$

> The ability to overfit data is important, its a sanity check for our inputs, algorithms, etc..

The worst possible error value is $.5$. Anything greater, we can flip the classification and get a “better” error rate.

Decision Stump

Finding the best feature with the best threshold to minimize the error rate.