Statistical Evaluation

TP - true pos	TN - true neg
FP - false pos	FN - false neg

Sensitivity

$$Sn = \frac{TP}{TP + FN}$$

$0 \le Sn \le 1$

Specificity

$$Sp = \frac{TN}{TN + FP}$$

$0 \le Sp \le 1$

Positive Predictive Value

$$PPV = \frac{TP}{TP + FP}$$

$0 \le PPV \le 1$

Correlation Coefficient

$$cc = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP) * (FP + TN) * (TN + FN) * (FN + TP)}}$$

$-1 \le cc \le 1$

cc:
- 1.0 no FP or FN
- 0.0 when *f* is random with respect to S+ and S-
- -1.0 only FP and FN

[Brazma et.al., 1998]

Compression

(1) *Raw Compression*:

$$C_{raw} = \left(\sum_{i=1}^{n} |S_i| \right) - (n-1) * |P|$$

i.e. SumOfElementsInExamples - (NumberOfExamples - 1) * elements in pattern

(2) *Normalised compression*:

$$C_{norm} = \frac{\left(\sum_{i=1}^{n} |S_i| \right) - C_{raw}}{\left(\sum_{i=1}^{n} |S_i| \right) - \min_{i=1}^{n} \left(|S_i| \right)}$$

This is a goodness measure (1=good, 0=bad).

More compression

(3) Substituting (1) into (2):

$$C_{norm} = \frac{(n-1) * |P|}{\left(\sum_{i=1}^{n} |S_i|\right) - \min_{i=1}^{n}(|S_i|)}$$

(4) Pairwise comparison via compression:

$$Comp(S_1, S_2) = \frac{|P|}{\max\left(|S_1|, |S_2|\right)}$$