Notes

Arthur Lui

24 March 2014

1 Sensitivity

$$\frac{\text{Number of true positives}}{\text{Number of true positives+Number of false positives}}$$

2 Specificity

$$\frac{\text{\# of true -ve}}{\text{\# of true -ve + \# of false +ve}}$$

3 1 - Specificity

$$\frac{\text{\# of false +ve}}{\text{\# of true -ve + \# of false +ve}}$$

4 ROC Curve

Plot Sensitivity vs. 1-Specificity
Higher Area Under the Curve (AUC) greater is better
Curve should be above x=y. Otherwise, better to flip a coin.

AUC = Probability that a classifier will rand a randomly chosen positive instance higher than a randomly chosen negative one.

low cutoff $\Rightarrow$ lots of false +ve and few false -ve
high cutoff $\Rightarrow$ lots of false -ve and few false +ve

5 Time-dependent ROC Curves

Classify subjects
Classify based on the risk score
Vary threshold C
Base Sensitivity/Specificity on whether subject jas actually experienced event by time T

D(t) = 1 if subject has experienced event
D(t) = 0 otherwise.

Sensitivity = $P[X > c|D(t) = 1]$
Specificity = $P[X \leq c|D(t) = 0]$
X = risk score = $e^{x'\beta}$
Time-varying component: calculate AUC at all t

Look for:
One AUC curve higher than another to select models.

6 12 March 2014

Sensitivity = $P[e^{x\beta} > c|D(t) = 1]$
Specificity = $P[e^{x\beta} \leq c|D(t) = 0]$

7 Multiple Testing

CI: one CI will likely retain parameter of interest
100 CI's, each with coverage .95: highly likely that at least one will not contain the parameter.
Expected that 5 will not .
P(at least 1 doesnt)=99.4%

8 Hypothesis Testing

One test: 5% chance of incorrectly rejecting H_0
100 test: Expected number of false rejections = 5

9 Familywise Error Rate

$1 - (1 - \alpha)^p = P[Atleast1rejectedH_0|H_0true]$

10 18 March: Confusion Matrix

Draw this table when I have time.

- V = false positive
- T = false negative
- U = true negative
- S = true positive
- M_0 = true null hypothesis
- $M - M_0$ = false null hypothesis
- R = declared significant
- M-R = declared not significant
- M = total counts
- FWER: P$[V \geq 1]$
- FDR: E$[\frac{V}{V+S}]$ = E$[\frac{V}{R}]$ = Expected % of false positives.

- False Positive Rate: $\frac{V}{M_0}$. % of truly null features declared significant.
- False Discovery Rate: $\frac{V}{R}$. % of those declared significant that are non-significant.

11 Code

p.adjust