NLP

Introduction to NLP

Bayes' Theorem

Bayes' Theorem

- Formula for joint probability
  - $p(A,B) = p(B|A)p(A)$
  - $p(A,B) = p(A|B)p(B)$
- Therefore
  - $p(B|A)=p(A|B)p(B)/p(A)$
- Bayes' theorem is used to calculate P(A|B) given P(B|A)

Example

- Diagnostic test
- Test accuracy
  - P(positive | ¬disease) = 0.05 – false positive
  - P(negative | disease) = 0.05 – false negative
  - so p(positive | disease) = 1–0.05 = 0.95

Example

- Diagnostic test with errors

P(A|B)		A=TEST	
		Positive	Negative
B=DISEASE	Yes	0.95	0.05
	No	0.05	0.95

Example

- What is p(disease | positive)?
  - P(disease|positive) = P(positive|disease)*P(disease)/P(positive)
  - P(¬disease|positive) = P(positive| ¬disease)*P(¬disease)/ P(positive)
  - P(disease|positive)/P(¬disease|positive) = ?
- We don't really care about p(positive)
  - as long as it is not zero, we can divide by it on both sides

Example

- P(disease|positive)/P(¬disease|positive) = (P(positive|disease) x P(disease))/(P(positive|¬disease) x P(¬disease))
- Suppose P(disease) = 0.001
  - so P(¬disease) = 0.999
- P(disease|positive)/P(¬disease|positive) = (0.95 x 0.001)/(0.05 x 0.999) =0.019
- P(disease|positive)+P(¬disease|positive) = 1
- P(disease|positive) ≈ 0.02
- P(disease) is called the prior probability
- P(disease|positive) is called the posterior probability
- In this example the posterior is 20 times larger than the prior

NLP