Bayes' Theorem

Bayes' Theorem

- p(A | B) = p(B | A) * p(A) / p(B)

- Easy to check by removing syntactic sugar
- Use 1: Converts p(B | A) to p(A | B)
- Use 2: Updates p(A) to p(A | B)

- Stare at it so you'll recognize it later

Language ID

- Given a sentence x, I suggested comparing its prob in different languages:
  - p(SENT=x | LANG=english) (i.e., $p_{english}$(SENT=x))
  - p(SENT=x | LANG=polish) (i.e., p_{polish}(SENT=x))
  - p(SENT=x | LANG=xhosa) (i.e., p_{xhosa}(SENT=x))
- But surely for language ID we should compare
  - p(LANG=english | SENT=x)
  - p(LANG=polish | SENT=x)
  - p(LANG=xhosa | SENT=x)

Language ID

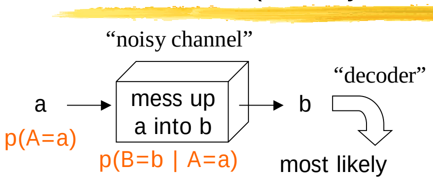
- For language ID we should compare
  - p(LANG=english | SENT=x)
  - p(LANG=polish | SENT=x)
  - p(LANG=xhosa | SENT=x)

  a posteriori
- For ease, multiply by p(SENT=x) and compare
  - p(LANG=english, SENT=x)
  - p(LANG=polish, SENT=x)
  - p(LANG=xhosa, SENT=x)

  sum of these is a way to find p(SENT=x); can divide back by that to get posterior probs
- Must know prior probabilities; then rewrite as
  - p(LANG=english) * p(SENT=x | LANG=english)
  - p(LANG=polish) * p(SENT=x | LANG=polish)
  - p(LANG=xhosa) * p(SENT=x | LANG=xhosa)

  a priori — likelihood (what we had before)

General Case ("noisy channel")



most likely reconstruction of a

language	→ text
text	→ speech
spelled	→ misspelled
English	→ French

maximize p(A=a | B=b)

= p(A=a) p(B=b | A=a) / (B=b)

= p(A=a) p(B=b | A=a)

/ Σ p(A=a') p(B=b | A=a')

Language ID

- For language ID we should compare
  - p(LANG=english | SENT=x)
  - p(LANG=polish | SENT=x)
  - p(LANG=xhosa | SENT=x)

  a posteriori
- For ease, multiply by p(SENT=x) and compare
  - p(LANG=english, SENT=x)
  - p(LANG=polish, SENT=x)
  - p(LANG=xhosa, SENT=x)
- Must know prior probabilities; then rewrite as
  - p(LANG=english) * p(SENT=x | LANG=english)
  - p(LANG=polish) * p(SENT=x | LANG=polish)
  - p(LANG=xhosa) * p(SENT=x | LANG=xhosa)

  a priori — likelihood

General Case ("noisy channel")

- Want most likely A to have generated evidence B
  - p(A = a1 | B = b)
  - p(A = a2 | B = b) *a posteriori*
  - p(A = a3 | B = b)
- For ease, multiply by p(SOUND=x) and compare
  - p(A = a1, B = b)
  - p(A = a2, B = b)
  - p(A = a3, B = b)
- Must know prior probabilities; then rewrite as
  - p(A = a1) * p(B = b | A = a1)
  - p(A = a2) * p(B = b | A = a2)
  - p(A = a3) * p(B = b | A = a3)
  - *a priori* likelihood

Speech Recognition

- For baby speech recognition we should compare
  - p(MEANING=gimme | SOUND=uhh)
  - p(MEANING=changeme | SOUND=uhh) *a posteriori*
  - p(MEANING=loveme | SOUND=uhh)
- For ease, multiply by p(SOUND=uhh) & compare
  - p(MEANING=gimme, SOUND=uhh)
  - p(MEANING=changeme, SOUND=uhh)
  - p(MEANING=loveme, SOUND=uhh)
- Must know prior probabilities; then rewrite as
  - p(MEAN=gimme) * p(SOUND=uhh | MEAN=gimme)
  - p(MEAN=changeme) * p(SOUND=uhh | MEAN=changeme)
  - p(MEAN=loveme) * p(SOUND=uhh | MEAN=loveme)
  - *a priori* likelihood

Life or Death!

- p(hoof) = 0.001 so p(¬hoof) = 0.999
- p(positive test | ¬hoof) = 0.05 "false pos"
- p(negative test | hoof) = x ≈ 0 "false neg"
  so p(positive test | hoof) = 1-x ≈ 1
- What is p(hoof | positive test)?
  - don't panic - still very small! < 1/51 for any x