DATA MINING B(2)

10-0. Answer for 1st quiz

Q1

- Please calculate the probability of play=yes/play=no under the following condition.

Outlook	Temperature	Humidity	Windy	Play
Sunny	66	90	true	?

- Answer
  - Temperature=66 when play=yes: Average=73, deviation=6.2
    - $f(temp = 66|yes) = \frac{1}{\sqrt{2\pi}\times 6.2} e^{-\frac{(73-66)^2}{2\times 6.2^2}}$=0.0340
  - Temperature=66 when play=no: Average=74.6, deviation=7.9
    - $f(temp = 66|no) = \frac{1}{\sqrt{2\pi}\times 7.9} e^{-\frac{(74.6-66)^2}{2\times 7.9^2}}$=0.279
  - Humidity=66 when play=yes: Average=79.1, deviation=10.2
    - $f(hum = 90|yes) = \frac{1}{\sqrt{2\pi}\times 10.2} e^{-\frac{(79.1-90)^2}{2\times 10.2^2}}$=0.0221
  - humidity=66 when play=no: Average=86.2, deviation=9.7
    - $f(hum = 90|no) = \frac{1}{\sqrt{2\pi}\times 9.7} e^{-\frac{(86.2-90)^2}{2\times 9.7^2}}$=0.0381

Q1

- Please calculate the probability of play=yes/play=no under the following condition.

Outlook	Temperature	Humidity	Windy	Play
Sunny	66	90	true	?

- Answer(cont.)
  - Yes cases are 9 of 14, No cases are 5 of 14.
  - P(outlook=sunny|yes): “yes” cases are 9, and outlook=sunny is 2, so p=2/9.
  - P(outlook=sunny|no): “no” cases are 5, and outlook=sunny is 3, so p=3/5.
  - P(windy=true|yes): “yes” cases are 9, and outlook=sunny is 3, so p=3/9.
  - P(windy=true|yes): “no” cases are 5, and outlook=sunny is 3, so p=3/5.

Q1

- Please calculate the probability of play=yes/play=no under the following condition.

Outlook	Temperature	Humidity	Windy	Play
Sunny	66	90	true	?

- Answer(cont.)
  - Likelihood of yes = p(sunny|yes) * p(temp=66|yes) * p(hum=90|yes) * p(windy=true|yes) * p(play=yes)
    = 2/9 * 0.0340 * 0.221 * 3/9 * 9/14 = 0.000036
  - Likelihood of no = p(sunny|no) * p(temp=66|no) * p(hum=90|no) * p(windy=true|no) * p(play=no)
    = 3/5 * 0.0279 * 0.381 * 3/5 * 5/14 = 0.000137

  Prob of yes = 0.000036 / (0.000036 + 0.000137) = 20.8%

  Prob of no = 0.0000137 / (0.000036 + 0.000137) = 79.2%

Supplement: Naïve Bayes Document Classification

- Please calculate a probability of document {blue, yellow, blue,} for class H'.

- Answer
  - Document class H'
    - Pr[yellow|H']=10%, Pr[blue|H']=90%
  - The document we want to classify is {blue, yellow, blue}
  - $P[\{\text{yellow, blue, blue}\}|H'] = 3! \cdot \frac{\left(\frac{1}{10}\right)^1}{1!} \cdot \frac{\left(\frac{9}{10}\right)^2}{2!} = \frac{3\times2\times1\times1\times9\times9}{2\times10\times10^2} = \frac{273}{1000} = 27.3\%$

For class H.

P[[{yellow, blue, blue}|H] = 14%