Solution 1

a) 1->3 = sqrt(8), 2->3 = sqrt(7), Document 2 would be returned
b) Document 1: 0.405, Document 2: 0, Document 3: 0.405. It's a tie between 1 and 3!
c) Document 1: 0.405, Document 2: 0.81, Document 3: 1.22, Document 3 is returned!

Solution 2

a) $y = \frac{(1*3) + (1*4) + (1*6)}{3} = 4.33$

b) $y = \frac{exp(-(1^2/3))*3 + exp(-(0/3)*4 + exp(-(3^2)/3))*6 + exp(-(5^2)/3)*7 + exp(-(7^2)/3)*10}{exp(-(1^2/3)) + exp(-(0)) + exp(-(3^2)/3) + exp(-(5^2)/3) + exp(-(7^2)/3)} = 3.65$

Solution 3

I. a. The precision metric captures the ratio between the number of correctly identified labels to all positive labels. It is useful to tell us how "good" the positive labeled items are. High precision indicates the results of a query are relevant (the positives are actually positive). Low precision indicates many of the positive results aren't relevant.

  b. The recall metric captures the ratio of correctly identified positive labels to all positive labels. It is useful to tell us how "complete" the model is -- how much of the positive label space was our model actually able to predict correctly. High recall indicates our model has labeled most of the positive labels correctly, low indicating many of the actual positive labels were missed.

  c. The precision-recall tradeoff relates how well the model was able to correctly identify positive labels, vs the willingness of the model to attribute a positive label. The "perfect" model would have precision 1.0 and recall 1.0, having no false positive and no false negatives.

II.
  a. 73 / (73+39) = 0.65
  b. 73 / (73+22) = 0.76

Solution 4

(a) TP =8, FP =2
    FN =3, TN = 17

Precision = $\frac{TP}{TP+FP}$ = 0.8
Recall = $\frac{TP}{TP+FN}$ = 0.73

(b) Precision = 0.88, Recall = 0.64
(c) Precision = 0.75, Recall = 0.82