

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی برق و کامپیوتر

بازیابی هوشمند اطلاعات - تمرین پنجم

سید مهدی رضوی

استاد : خانم دکتر شاکری

دی ماه ۱۴۰۲

فهرست مطالب

| | |
|---|-------------|
| ۳ | ۱ تمرین اول |
| ۴ | ۲ تمرین دوم |
| ۵ | ۳ تمرین سوم |

فهرست تصاویر



۱ تمرین اول



۲ تمرین دوم

۳ تمرین سوم

$$P(\text{Minus}|\text{Doc1}) < P(\text{Plus}|\text{Doc1})$$

$$P(\text{Minus}|\text{Doc2}) > P(\text{Plus}|\text{Doc2})$$

$$P(\text{Minus}|\text{Doc3}) < P(\text{Plus}|\text{Doc3})$$

$$P(\text{Minus}|\text{Doc4}) > P(\text{Plus}|\text{Doc4})$$

$$P(\text{Minus}|\text{Doc5}) < P(\text{Plus}|\text{Doc5})$$

با توجه به نتایج بالا ، هر یک از اسناد برجسب متناظر خود را دریافت خواهند کرد.
بر اساس منطق این رده‌بند ، هر کدام که از این اسناد که احتمال شرطی بالاتری برای هر کدام از کلاس‌ها داشته‌باشد ، به آن کلاس مربوط می‌شود.
باید به مقایسه دو عبارت زیر به ازای سند ۶ و ۷ بپردازیم و سپس برجسب متناظر با آن داده را بر روی سند مدنظر بزنیم.

$$P(\text{plus}) * P(W_1|\text{plus}) * P(W_2|\text{plus}) * P(W_3|\text{plus}) * \dots * P(W_n|\text{plus})$$

$$P(\text{minus}) * P(W_1|\text{minus}) * P(W_2|\text{minus}) * P(W_3|\text{minus}) * \dots * P(W_n|\text{minus})$$

با توجه به معادلات جایگاه نامعادلات بالا ، به داده‌های زیر خواهیم رسید :

$$P(\text{plus}) = \frac{3}{5}$$

$$P(\text{minus}) = \frac{2}{5}$$

$$P(\text{love}|\text{plus}) = \frac{1+1}{11+6} = \frac{2}{17}$$

$$P(\text{movie}|\text{plus}) = \frac{4+1}{11+6} = \frac{5}{17}$$

$$P(\text{great}|\text{plus}) = \frac{2+1}{11+6} = \frac{3}{17}$$

$$P(\text{good}|\text{plus}) = \frac{2+1}{11+6} = \frac{3}{17}$$

$$P(\text{acting}|\text{plus}) = \frac{1+1}{11+6} = \frac{2}{17}$$

$$P(I|\text{plus}) = \frac{1+1}{11+6} = \frac{2}{17}$$

$$P(\text{hated}|\text{minus}) = \frac{1+1}{5+5} = \frac{2}{10}$$

$$P(I|\text{minus}) = \frac{1+1}{5+5} = \frac{2}{10}$$

$$P(\text{movie}|\text{minus}) = \frac{1+1}{5+5} = \frac{2}{10}$$

$$P(\text{poor}|\text{minus}) = \frac{1+1}{5+5} = \frac{2}{10}$$

$$P(\text{acting}|\text{minus}) = \frac{1+1}{5+5} = \frac{2}{10}$$

Document6 : I loved the poor play.

$$P(plus|Document6) = P(plus) * \frac{P(love|plus) * P(I|plus) * P(poor|plus) * P(play|plus)}{P(Document6)}$$

$$P(minus|Document6) = P(minus) * \frac{P(love|minus) * P(I|minus) * P(poor|minus) * P(play|minus)}{P(Document6)}$$

با توجه به این که مقدار احتمال سند ششم برای هر دو ترم با هم برابر است ، به صورت زیر تخمین خواهیم زد :

$$P(plus|Document6) = \frac{3}{5} * \frac{2}{17} * \frac{2}{17} * \frac{1}{17} * \frac{1}{17} = 0.01936$$

$$P(minus|Document6) = \frac{2}{5} * \frac{1}{10} * \frac{2}{10} * \frac{2}{10} * \frac{1}{10} = 0.00016$$

برچسب سند ششم مثبت خواهد بود.

Document7 : I hated the play movie.

همانند جایگاه معادلات بالا برای سند ۶ ، مقادیر احتمالاتی را برای کلاس‌های ممکن برای سند ۷ را محاسبه خواهیم کرد :

$$P(plus|Document7) = \frac{3}{5} * \frac{2}{17} * \frac{1}{17} * \frac{1}{17} * \frac{5}{17} = 0.00007183822$$

$$P(minus|Document7) = \frac{2}{5} * \frac{2}{10} * \frac{2}{10} * \frac{2}{10} * \frac{1}{10} = 0.00032$$

برچسب سند هفتم منفی خواهد بود.

با توجه به نتایج بالا و وزن کلمه مثبتی همچون love و همچنین وزن کلمه منفی مانند hate همچنین نتایجی قابل پیش‌بینی بود. نتایج به دست‌آمده منطقی هستند.