Information Retrieval Homework 2

Chengjun Yuan

October 19, 2015

1 Joint, marginal and conditional probabilities

1.1

$$P(X \leq 2, Y > 1) = \frac{1}{12}$$

1.2

$$P(X = 1) = \frac{5}{12},\ P(X = 2) = \frac{1}{6},\ P(X = 4) = \frac{5}{12}$$
$$P(Y = 1) = \frac{7}{12},\ P(Y = 2) = \frac{5}{12}$$

1.3

$$P(Y = 2|X = 1) = \frac{1}{5}$$

1.4

No, they are not independent. Because, for example,

$$P(X = 1, Y = 1) = \frac{1}{3} \neq P(X = 1)P(Y = 1) = \frac{35}{144}$$

1.5

$$P(Z = 0) = P(X = 2, Y = 1) + P(X = 4, Y = 2) = \frac{1}{2}$$
$$P(X = 2|Z = 0) = \frac{P(X=2,Z=0)}{P(Z=0)} = \frac{P(X=2,(X=2,Y=1))+P(X=2,(X=4,Y=2))}{P(Z=0)}$$
$$\Rightarrow P(X = 2|Z = 0) = \frac{1}{3}$$

1.6

$$E[X|Y=1] = \sum_x xP(X=x|Y=1) = \sum_x x\frac{P(X=x,Y=1)}{P(Y=1)}$$
$$\Rightarrow E[X|Y=1] = \tfrac{4}{7} + \tfrac{4}{7} + \tfrac{4}{7} = \tfrac{12}{7}$$

1.7

$$Var[X|Y=2] = E[(X - E[X|Y=2])^2|Y=2]$$
$$\text{and } E[X|Y=2] = \sum_x xP(X=x|P=2) = \tfrac{17}{5}$$
$$\Rightarrow Var[X|Y=2] = E[(X-\tfrac{17}{5})^2|Y=2] = \sum_x (x-\tfrac{17}{5})^2 P(X=x|P=2)$$
$$\Rightarrow Var[X|Y=2] = \tfrac{36}{25}$$

2 Proof of probabilistic ranking principle

Here, we define two types of loss:

$$a_1 = \text{loss of retrieval of non-relevant result.}$$
$$a_2 = \text{loss of non-retrieval of relevant result.}$$

$\phi(d_i, q)$ denotes the probability of d_i being relevant to q. When the retrieval results are ranked by probabilistic ranking principle, we have :

$$\phi_i = \phi(d_i, q) \geq \phi_j = \phi(d_j, q) \text{ when } i < j. \text{ } (i, j \text{ is the ranking order})$$

When the user just takes the first k items of the ranked results, the overall loss expectation is:

$$f(k) = \sum_{i=1}^{k} (1-\phi_i)\cdot a_1 + \sum_{i=k+1}^{\infty} \phi_i \cdot a_2$$

If we exchange the ranking position of the m^{th} and the n^{th} results in the original ranked retrieval ($m < n$), then we have $\phi_m < \phi_n$ and :

$$f'(k) = \sum_{i=1}^{k} (1-\phi_i)\cdot a_1 + \sum_{i=k+1}^{\infty} \phi_i \cdot a_2$$

If $m < n \leq k$ or $k < m < n$, then $f'(k) = f(k)$. However, if $m \leq k < n$, $f'(k) \neq f(k)$ and:

$$f(k) - f'(k) = (1 - \phi(d_m, q) - 1 + \phi(d_n, q))\cdot a_1 + (\phi(d_n, q) - \phi(d_m, q))\cdot a_2$$
$$\Rightarrow f(k) - f'(k) = (\phi(d_n, q) - \phi(d_m, q))\cdot(a_1 + a_2) \leq 0$$

Therefore, the loss of probabilistic ranking principle is always less than or equal to the loss of any other non-ranked or non-entire ranked retrieval results, which means that the probabilistic ranking principle will minimize the loss or risk.

3 Maximum likelihood estimation

For Gaussian distribution, the log likelihood is

$$L(\mu, \sigma) = -\frac{1}{2}N log(2\pi\sigma^2) - \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2\sigma^2}$$

We can then find the values of μ and σ^2 that maximize the log likelihood by taking derivative with respect to the desired variable and solving the equation obtained. By doing so, we find that the **maximum likelihood estimation of mean** is

$$\frac{\partial L(\mu,\sigma)}{\partial \mu} = \sum_{n=1}^{N} \frac{x_n-\mu}{\sigma^2} = 0$$
$$\Rightarrow \mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$
$$\frac{\partial L(\mu,\sigma)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2\sigma^4} = 0$$
$$\Rightarrow \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$$

4 Cosine v.s. Euclidean distance

The Cosine similarity and Euclidean distance of **X** and **Y** is respectively:

$$cos(X, Y) = \frac{X \cdot Y}{||X|| \cdot ||Y||} = \sum_{i=1}^{n} x_i \cdot y_i, \text{ since } ||X|| = ||Y|| = 1.$$
$$d(X, Y) = ||X - Y|| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}.$$

Then, we can have:

$$(d(X, Y))^2 = ||X - Y||^2 = (X - Y)^T(X - Y) = ||X||^2 + ||Y||^2 - 2X^TY.$$
$$\Rightarrow d(X, Y) = \sqrt{2(1 - \sum_{i=1}^{n} x_i \cdot y_i)} = \sqrt{2 - 2cos(X, Y)}$$
$$\Rightarrow cos(X, Y) = 1 - \frac{1}{2}(d(X, Y))^2$$

5 α_d in language model ranking modules

For dirichlet prior smoothing methods, we have :

$$p_s(w|d) = \frac{c(w;d) + \mu p(w|C)}{|d| + \mu}.$$

Then we substitute it into the generic ranking function:

$$logP(q|d) = \sum_{w \in d \cap q}(log\frac{p_s(w;d)}{\alpha_d p(w|C)}) + |q| log\alpha_d$$

We can get:

$$logP(q|d) = \sum_{w \in d \cap q}(log(\frac{c(w;d)}{\alpha_d(|d|+\mu)p(w|C)} + \frac{\mu}{\alpha_d(|d|+\mu)})) + |q|log\alpha_d$$

To set $\frac{\mu}{\alpha_d(|d|+\mu)} = 1$, then we obtain:

$$\alpha_d = \frac{\mu}{|d|+\mu}$$

Or we can also calculate α_d through the basic idea of smoothing:

$$p(w|d) = \begin{cases} p_s(w|d) & \text{if } w \in d \\ \alpha_d p(w|C) & \text{if } w \notin d \end{cases}$$

Since $\sum_{w \in d} p_s(w|d) + \sum_{w \notin d} \alpha_d p(w|C) = 1$, so we substitute $p_s(w|d)$ into it, then get:

$$\frac{|d|}{|d|+\mu} + \frac{\mu \sum_{w \in d} p(w|C)}{|d|+\mu} + \alpha_d \sum_{w \notin d} p(w|C) = 1$$

$$\text{given } \sum_{w \in q} p(w|C) = 1$$

$$\Rightarrow \alpha_d = \frac{\mu}{|d|+\mu}$$

6 Bonus questions: α_d in language model ranking modules

For **linear interpolation smoothing:**

$$p_s(w|d) = (1-\lambda)\frac{c(w;d)}{|d|} + \lambda p(w|C).$$

Using the same method from the Problem 5, we can have:

$$\alpha_d = \frac{1-\sum_{w \in d} p_s(w|d)}{\sum_{w \notin d} p(w|C)} = \lambda$$

For **additive smoothing:**

$$p(w|d) = \begin{cases} p_s(w|d) = \frac{c(w,d)+\delta}{|d|+\delta|V|} & \text{if } w \in d \\ \alpha_d p(w|C) = \frac{\delta}{|d|+\delta|V|} & \text{if } w \notin d \end{cases}$$

So we have $p(w|C) = \frac{1}{\alpha_d}\frac{\delta}{|d|+\delta|V|}$. According to the definition of α_d to make sure that all probabilities sum to one:

$$\alpha_d = \frac{1-\sum_{w \in V \cap D} p_s(w|d)}{1-\sum_{w \in V \cap D} p(w|C)}$$

$$\Rightarrow 1 - \frac{|d|+\delta|w \in V \cap D|}{|d|+\delta|V|} = \alpha_d - \frac{\delta|w \in V \cap D|}{|d|+\delta|V|}$$

$$\Rightarrow \alpha_d = \frac{\delta|V|}{|d|+\delta|V|}$$