CSM146 Homework 3

Zipeng Fu

March 16, 2018

1 Naive Bayes

(a) The information about the relationship between the current word and the word before it and after it is lost.

(b)

$$\log Pr(D_i, y_i) = \log Pr(D_i|y_i) + \log Pr(y_i)$$

$$= \log\left(\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i(1-y_i)}\alpha_1^{a_iy_i}\beta_0^{b_i(1-y_i)}\beta_1^{b_iy_i}\gamma_0^{c_i(1-y_i)}\gamma_1^{c_iy_i}\right) + \log\theta^{y_i}(1-\theta)^{1-y_i}$$

(c)

$$\frac{\partial\log\left(\prod_{i=1}^{m} P(D_i, y_i)\right)}{\partial\alpha_1} = 0$$

$$\frac{\partial\sum_{i=1}^{m}\left(a_iy_i\log\alpha_1 + c_iy_i\log(-\alpha_1 - \beta_1 + 1)\right)}{\partial\alpha_1} = 0$$

$$\frac{\sum_{i=1}^{m} a_iy_i}{\alpha_1} - \frac{\sum_{i=1}^{m} c_iy_i}{-\alpha_1 - \beta_1 + 1} = 0$$

$$\alpha_1(\sum a_iy_i + \sum c_iy_i) + (\beta_1 - 1)\sum a_iy_i = 0$$

Similarly, take derivative with respect to β_1 will get

$$\beta_1(\sum b_iy_i + \sum c_iy_i) + (\alpha_1 - 1)\sum b_iy_i = 0.$$

By solving the above two equations simultaneously, we will get,

$$\alpha_1 = \frac{\sum a_iy_i}{\sum a_iy_i + \sum b_iy_i + \sum c_iy_i},$$

$$\beta_1 = \frac{\sum b_iy_i}{\sum a_iy_i + \sum b_iy_i + \sum c_iy_i}.$$

By the same approach, we will get,

$$\gamma_1 = \frac{\sum c_iy_i}{\sum a_iy_i + \sum b_iy_i + \sum c_iy_i},$$

and also,

$$\alpha_0 = \frac{\sum a_i(-y_i + 1)}{\sum a_i(-y_i + 1) + \sum b_i(-y_i + 1) + \sum c_i(-y_i + 1)},$$

$$\beta_0 = \frac{\sum b_i(-y_i + 1)}{\sum a_i(-y_i + 1) + \sum b_i(-y_i + 1) + \sum c_i(-y_i + 1)},$$

$$\gamma_0 = \frac{\sum c_i(-y_i + 1)}{\sum a_i(-y_i + 1) + \sum b_i(-y_i + 1) + \sum c_i(-y_i + 1)}.$$

2 Hidden Markov Models

(a) The missing transition probabilities are $P(q_{t+1} = 1|q_t = 2) = 1 - q_{11} = 0$ and $P(q_{t+1} = 2|q_t = 2) = 1 - q_{12} = 0$.

The missing probabilities are $e_1(B) = P(O_t = B|q_t = 1) = 1 - e_1(A) = 0.01$ and $e_2(A) = P(O_t = A|q_t = 2) = 1 - e_2(B) = 0.49$

(b)

$$P(\text{output } A) = \pi_1 \times e_1(A) + \pi_2 \times e_2(A) = 0.49 \times 0.99 + 0.51 \times 0.49 = 0.735$$

$$P(\text{output } B) = \pi_1 \times e_1(B) + \pi_2 \times e_2(B) = 0.49 \times 0.01 + 0.51 \times 0.51 = 0.265$$

Hence, A is the most frequenct output symbol to appear in the 1st position.

(c) Since the output symbol with the highest probability is A and also due to $q_1 1 = q_1 2 = 1$, the 2nd and 3rd states will be 1, regardless of 1st state being 1 or 2. Also, $e(\text{A}) = 0.99$, so the 2nd and 3rd symbol with highest probability is A. Hence, the first 3 symbols with highest probability are A.

3 Facial Recognition

(a) The reason of that it is a bad idea is that k is not fixed. Under this circumstance, $J(c, \mu, k)$ is 0 and n centoids are at positions of n datapoints. Hence, $k = n$, $\mu_i = x^{(i)}$, and $c^{(i)} = i$.