Machine Learning PS 6 — Meet Barot

1. A: Testing positive
B: Having disease

$P(A|B) = 0.97 \quad P($

$P(B) = \frac{1}{20000}$

Need to find $P(B|A)$.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A) = P(A|B) * P(B) + P(A|\neg B) * P(\neg B)$$

$$= .97 \cdot \frac{1}{20000} + .03 \cdot \frac{19999}{20000}$$

$$= .030047$$

$$P(A \cap B) = P(A|B)\,P(B) = .97 \cdot \frac{1}{20000} = 0.0000485$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.0000485}{.030047} = 0.001614$$

You have a 0.1614% chance of having the disease.

2. $\arg\max_y \left(Pr(Y) \prod_i P(X_i = x_i | Y) \right)$

$$= \arg\max_y \left(\theta_c \prod_i (\theta_{yi}^{x_i})(1-\theta_{yi})^{(1-x_i)} \right)$$

$$= \arg\max_y \ln\left(\theta_c \prod_i (\theta_{y:}^{x_i})(1-\theta_{y:})^{(1-x_i)} \right)$$

$$= \arg\max_y \ln\theta_c + \sum_i \left(x_i \ln\theta_{yi} + (1-x_i)\ln(1-\theta_{yi}) \right)$$

$$= \arg\max_y \underbrace{\ln\theta_c + \textstyle\sum_i \ln(1-\theta_{yi})}_{w_0} + \sum_i \underbrace{\left(\ln\theta_{yi} - \ln(1-\theta_{yi})\right)}_{w_i} x_i$$

Thus,

$$\underset{y}{\operatorname{argmax}} \Pr(y|x;\vec{\theta}) = \underset{x}{\operatorname{argmax}}\, y(w \cdot x + b).$$

3. a. Using War and Peace text file as corpus
b. Using sklearn's LDA implementation,

5 topics:

- "War" topic: troops, shouted, regiment, rostov, reached
- "People" topic: princess, mary, men, man, said
- "officer" topic: Napoleon, officer, kutuzov, knows, voice
- Some other ill-defined topics

10 topics:

- "Loss" topic: mother, taken, understood, took, saw, crowd, battle, away, soon
- "Action" topic: Napoleon, eyes, look, action, moved, forward, regiment, activity.
- "Russian" topic: Russian, moscow, time, people

These are more specific and well-defined than the 5 topic run's topics.

4. a. 0.9

b. $P(Y_2 = \text{frown}) = P(Y_2 = \text{frown} \mid X_2 = \text{Happy}) * P(X_2 = \text{Happy}) + P(Y_2 = \text{frown} \mid X_2 = \text{Angry}) * P(X_2 = \text{Angry})$

$= 0.9 * 0.1 + 0.1 * 0.6 = 0.15$

c. $P(X_2 = \text{Happy} \mid Y_2 = \text{frown}) = \frac{P(Y_2 = \text{frown} \mid X_2 = \text{Happy}) P(X_2 = \text{happy})}{P(Y_2 = \text{frown})}$

$= \frac{0.1 * 0.9}{0.15} = \frac{0.09}{0.15} = \frac{3}{5} = 0.6$

d. $P(Y_{80} = \text{yell}) = P(X_{80} = \text{Happy}) P(Y_{80} = \text{yell} \mid X_{80} = \text{Happy}) + P(X_{80} = \text{Angry}) P(Y_{80} = \text{yell} \mid X_{80} = \text{Angry})$

$= 0.9 \cdot 0.1 + 0.1 \cdot 0.2 = 0.11$

e. $P(X_1 = \text{Happy} \mid Y_1 = \text{frown}) = 1 \leftarrow \text{given}$

$P(X_2 = \text{Happy} \mid Y_2 = \text{frown}) = 0.6$

$P(X_3 = \text{Happy} \mid Y_3 = \text{frown}) = 0.6$

$P(X_4 = \text{Happy} \mid Y_4 = \text{frown}) = 0.6$

$P(X_5 = \text{Happy} \mid Y_5 = \text{frown}) = 0.6$

The most likely sequence is Happy, Happy, Happy, Happy, Happy.