Homework 6

1. KL-divergence: $f(q,d) = \sum_{w \in d, p(w|\hat{\theta}_Q) > 0} p(w|\hat{\theta}_Q) \log \frac{P_{seen}(w|d)}{\alpha_d P(w|C)} + \log \alpha_d$

assuming $p(w|\hat{\theta}_Q) = \frac{c(w,Q)}{|Q|}$

$f(q,d)$ becomes

$$\sum_{\substack{w \in d \\ w \in Q}} \frac{c(w,Q)}{|Q|} \cdot \log \frac{P_{seen}(w|d)}{\alpha_d P(w|C)} + \log \alpha_d$$

$|Q|$ can be ignored because it is constant

$\therefore$

$$f(q,d) = \sum_{\substack{w \in d \\ w \in Q}} c(w,Q) \log \frac{P_{seen}(w|d)}{\alpha_d P(w|C)} + n \log \alpha_d$$

which is the query likelihood function

$\blacksquare$

2.

a.) $p(\text{"the"}) = (.80)(.3) + (.2)(.3) = .3$

b.) $p(\text{"the"}) = (.8)(.3) + (.2)(.3) = .3$

c.) $$p(H \mid w = \text{"data"}) = \frac{p(w = \text{"data"} \mid H) \cdot p(H)}{p(H)p(\text{"data"} \mid H) + p(\text{"data"} \mid T)p(T)}$$

$$= \frac{(.8)(.1)}{(.8)(.1) + (.2)(.1)} = .8$$

d. $p(\text{"the"}) = .3$

$p(\text{"computer"}) = .12$

$p(\text{"data"}) = .1$

$p(\text{"baseball"}) = .18$

$p(\text{"game"}) = .18$

$\therefore$ data would occur least frequently b/c $p(\text{"data"})$ is the smallest

e. $p(\text{"computer"}) = \frac{3}{10} = .3$

$p(\text{"game"}) = \frac{2}{10} = .2$

2. ~~the~~

i a. when $\boxed{\lambda = 0}$

$$\log p(D \mid \Theta, \Pi) = \sum_{i=1}^{N} \sum_{j=1}^{|d_i|} \log \left[0 \cdot p(d_{i,j} = w \mid D) + (1-0) \sum_{k=1}^{K} p(z_{i,j} = k \mid \Pi_i) \cdot p(d_{i,j} = w \mid \theta_k) \right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{|d_i|} \log \left[\sum_{k=1}^{K} p(z_{i,j} = k \mid \Pi_i) \, p(d_{i,j} = w \mid \theta_k) \right]$$

b.) ~~$\tau(w_i$~~

$$\frac{\sum_{i=1}^{N} c(w, D_i)}{\sum_{i=1}^{N} |D_i|}$$

c.) If λ is very large the log liklehood will depend less on the topics and more on the background language model. If λ is smaller the likelihood will depend more on the topics found by the model

To test this hypothesis, one could run the PLSA multiple times with different values of lambda. We could then use the data to see if the likelihood reflects the different topics more or the background model ii

a.) $$p(z_{d,w}=k) = \frac{p(z_{i,j}=k \mid \pi_i^{(n)})\, p(w_{i,j} \mid \theta_k^{(n)})}{\sum_{k=1}^{K} p(z_{i,j}=k \mid \pi_i^{(n)})\, p(w_{i,j} \mid \theta_k^{(n)})}$$

$$p(z_{d,w}) = B = \frac{\lambda\, p(w_{i,j} \mid \theta_B^{(n)})}{\lambda\, p(w_{i,j} \mid \theta_B^{(n)}) + (1-\lambda) \sum_{k=1}^{K} p(z_{i,j}=k \mid \pi_i^{(n)})\, p(w_{i,j} \mid \theta_k)}$$

$$h_{w,k} = \sum_{d \in D} c(w,d)\, p(z_{d,w}=k) \cdot (1 - p(z_{d,w}=B))$$

$$h_{d,k} = \sum_{w \in d} c(w,d)\, p(z_{d,w}=k) \cdot (1 - p(z_{d,w}=B))$$

b.) $$p(w \mid \theta_k^{(n+1)}) = \frac{h_{w,k}}{\sum_{w' \in V} h_{w',k}}$$

$$p(z_{d,w}=k \mid \pi_k^{(n+1)}) = \frac{h_{d,k}}{\sum_{k'=1}^{K} h_{d,k'}}$$