**Name: Lianhan Huang 3700459; Qu Wang 3700666**

## 1. Entropy, Cross Entropy and KL divergence

**1.1**

H(P) = $-\sum \quad P(X=x) * logP(X=x)$

H(P, Q) = $-\sum \quad P(X=x) * logQ(X=x)$

H(P, P) = $-\sum \quad P(X=x) * logQ(X=x)$ = H(P)

**1.2**

H(P, Q) = $-\sum \quad P(x) * log(Q(x))$

H(P) = $-\sum \quad P(x) * log(P(x))$

D(P||Q) = $-\sum \quad P(x) * log(Q(x)/P(x))$

= $-\sum \quad P(x) * (log(Q(x)) - (log(P(x)))$

= $-\sum \quad P(x) * log(Q(x)) + \sum \quad P(x) * log(P(x))$

= H(P, Q) - H(P)

-> H(P, Q) = H(P) + D(P||Q)

**1.3**

P = [p1; p2], Q = [q1; q2]

D(P||Q) = $-(P(p1) * log(Q(q1)/P(p1)) + P(p2) * log(Q(q2)/P(p2))$ )

D(Q||P) = $-(Q(q1) * log(P(p1)/Q(q1)) + Q(q2) * log(P(p2)/Q(q2))$ )

If D(P||Q) = D(Q||P)

So:

$$P(p1) * log(Q(q1)/P(p1)) + P(p2) * log(Q(q2)/P(p2))$$
$$=$$
$$Q(q1) * log(P(p1)/Q(q1)) + Q(q2) * log(P(p2)/Q(q2))$$

So:

$$log((Q(q1)/P(p1)) ** P(p1) * (Q(q2)/P(p2)) ** P(p2))$$
$$=$$
$$log((P(p1)/Q(q1)) ** Q(q1) * (P(p2)/Q(q2)) ** Q(q2))$$

Unless p1=q1 and p2=q2, D(P||Q) = D(Q||P)

Or D(P||Q) $\neq$ D(Q||P)

**1.4**

Likelihood = $\prod$ P(yi|xi)

Log(L) = log($\prod$ P(yi|xi)) = $\sum \quad log$(P(yi|xi))

H(P, Q) = $-\sum \quad P(x) * log(Q(x))$

Let Q = P(yi|xi), H(P, Q) = $-\sum \quad P$(xi) $* log$(P(yi|xi))

If each observed data equally distributed, P(xi) = 1/n

So, H(P, Q) = $-\sum \quad \frac{1}{n} * log$(P(yi|xi)) =-$\frac{1}{n}\sum \quad log$(P(yi|xi))

So, H(P,Q) = - Log(L)

So max likelihodd is to min cross entropy.

**1.5**

As mentioned before:

If P=Q, D(P||Q) = D(Q||P) = $log((Q(q1)/P(p1)) ** P(p1) * (Q(q2)/P(p2)) ** P(p2))$

= $log((P(p1)/Q(q1)) ** Q(q1) * (P(p2)/Q(q2)) ** Q(q2))$

= $log(1)$

= 0