Q1.

a. $\sum_Y P(X,Y) =$

X=1	X=2	X=3
.15	.35	.5

b. $\sum_X P(X,Y) =$

Y=a	Y=b
.6	.4

c. $\frac{P(X,Y)}{P(Y)} =$

	X=1	X=2	X=3
Y=a	.167	.333	.5
Y=b	.125	.375	.5

d. $\frac{p(X,Y)}{P(X)} =$

	X=1	X=2	X=3
Y=a	.667	.571	.6
Y=b	.333	.429	.4

e. No. $P(Y) \neq P(Y|X)$

f. $-.15\,log_2(.15) - .35\,log_2(.35) - .5\,log_2(.5) = 1.441$

g. $-.6\,log_2(.6) - .4\,log_2(.4) = .971$

h. $-\sum_X \sum_Y P(X,Y) \log_2 P(X,Y)$
$= -.10\,log_2(.10) - .20\,log_2(.20) - .30\,log_2(.30) - .05\,log_2(.05) - .15\,log_2(.15) - .20\,log_2(.20)$
$= 2.409$

i. $H(X,Y) - H(Y) = 1.438$

j. $H(X,Y) - H(X) = .968$

k. $H(Y) - H(Y|X) = .003$

l. $KL\big(P(X,Y||Q(X,Y)\big) = -\sum_X P(X)\,log_2\,Q(X,Y) + \sum_X P(X)\,log_2\,P(X,Y) =$
$-.10\,log_2(.10) - .20\,log_2(.20) - .30\,log_2(.40) - .05\,log_2(.01) - .15\,log_2(.09) - .20\,log_2(.20) - (-.10\,log_2(.10) - .20\,log_2(.20) - .30\,log_2(.30) - .05\,log_2(.05) - .15\,log_2(.15) - .20\,log_2(.20)) = 2.51082 - 2.40869 = .102$

$KL\big(Q(X,Y||P(X,Y)\big) = -\sum_X Q(X)\,log_2\,P(X,Y) + \sum_X Q(X)\,log_2\,Q(X,Y) =$
$-.10\,log_2(.10) - .20\,log_2(.20) - .40\,log_2(.30) - .01\,log_2(.05) - .09\,log_2(.15) - .20\,log_2(.20) - (-.10\,log_2(.10) - .20\,log_2(.20) - .40\,log_2(.40) - .01\,log_2(.01) - .09\,log_2(.09) - .20\,log_2(.20)) = 2.24530 - 1.86985 = .375$

These are not equal.

Q2.

a. $-p\,log_2\,p - (1-p)\,log_2(1-p)$

b. .5

c. $\frac{d}{dp}\big((-p\,log_2\,p) - (1-p)\,log_2(1-p)\big) = 0$
$-(\ln(x)+1) - \ln(1-x) + 1 = 0$
$\ln\left(\frac{1-x}{x}\right) = 0$
$\frac{1-x}{x} = 1$

$x = .5$

Additionally, $\frac{d^2}{dp^2}\big((-p\ log_2\ p) - (1-p)\ log_2(1-p)\big) = -\frac{1}{p(1-p)\ln(2)} < 0$ for p = .5, so this is a local maximum.

Q3.

a. $\frac{\binom{n}{\frac{n}{2}}}{2}$

b. $\frac{10!}{5!3!2!} = 2520$

c1. $\frac{N!}{t_1!t_2!\ldots t_n!}$

c2. $\frac{N!}{t_1!t_2!\ldots t_n!}\prod_i (P(w_i))^{t_i}$

Q4.

a. $\prod_i P(w_i|t_i)P(t_i|t_{i-2}, t_{i-1})$

b. T states, corresponding to the number of POS tags
$a_{ij} = P(s_j|s_i)$
$b_{jk} = P(o_k|s_j)$

Q5.

a. $O(V^2T^2)$

b. x is a word; y is a tag

c.

<Mike NN w_{-1} <s> w_0 Mike w_1 likes $w_{-1}w_{+1}$ BOS_likes t_{-1} BOS $t_{--2}t_{-1}$ BOS_BOS>
<likes VBP w_{-1} Mike w_0 likes w_1 cats $w_{-1}w_{+1}$ Mike_cats t_{--1} NN $t_{-2}t_{-1}$ BOS_NN)>
<cats NNS w_{-1} likes w_0 cats w_1 EOS $w_{-1}w_{+1}$ likes_EOS t_{-1} VBP $t_{-2}t_{-1}$ NN_VBP>

Q6.

a. I would build a classifier, where x is a document and y is a language. I would choose the following features:
   1. Word unigrams
   2. Word bigrams
   3. Word trigrams
   4. Letters
   5. Word length
   6. Document size

b.
   1. Languages poorly represented in training data
   2. Documents in closely related languages
   3. Some languages are harder than others to identify word boundaries