LING 572 Hw1
Due: 11:45pm on Jan 14, 2010

Q1 (25 points): Let X and Y be two random variables. The values for P(X,Y) are shown in Table 1, H(X) is the entropy of X, and MI(X,Y) is the mutual information of X and Y. Please write down the formulas and the results for the following.

(a) 1 pt: $P(X)$

(b) 1 pt: $P(Y)$

(c) 1 pt: $P(X \mid Y)$

(d) 1 pt: $P(Y \mid X)$

(e) 2 pts: Are X and Y independent? Why or why not?

(f) 2 pts: $H(X)$

(g) 2 pts: $H(Y)$

(h) 2 pts: $H(X,Y)$

(i) 2 pts: $H(X \mid Y)$

(j) 2 pts: $H(Y \mid X)$

(k) 2 pts: $MI(X,Y)$

(l) 7 pts: The value for $Q(X,Y)$ are shown in Table 2. What is the value for $KL(P(X,Y) \mid\mid Q(X,Y))$? What is the value for $KL(Q(X,Y) \mid\mid P(X,Y))$? Are they the same?

Table 1: The joint probability P(X,Y)

	X=1	X=2	X=3
Y=a	0.10	0.20	0.30
Y=b	0.05	0.15	0.20

Table 2: The joint probability Q(X,Y)

	X=1	X=2	X=3
Y=a	0.10	0.20	0.40
Y=b	0.01	0.09	0.20

Q2 (10 points): Let X be a random variable for the result of tossing a coin. $P(X = h) = p$; that is, p is the possibility of getting a head, and $1 - p$ is the possibility of getting a tail.

(a) 1 pt: $H(X)$ is the entropy of X. Write down the formula for $H(X)$.

(b) 2 pts: Let $p^* = arg\ max_p\ H(X)$; that is, p^* is the p that results in the maximal value of $H(X)$. What is p^*?

(c) 7 pts: Prove that the answer you give in (b) is correct. Hint: recall how you calculate the optimal solution for a function $f(x)$ in your calculus class? In this case, $H(X)$ is a function of p.

Q3 (15 points): Permutations and combinations:

(a) 3 pts: There are 10 balls: 5 are red, 3 are blue, and 2 are white. Suppose you put the balls in a line, how many different color sequences are there?

(b) 6 pts: Suppose you want to create a document of length N by using only the words in a vocabulary $\Sigma = \{w_1, w_2, ..., w_n\}$. How many different documents are there which satisfy the condition that for each w_i in Σ the occurrence of the word w_i in the document is exactly t_i? That is, how many different word sequences are there which contain exactly t_i w_i's for each w_i in Σ?

Hint: The answer to (b) is very similar to the answer to (a).

(c) 6 pts: Suppose you want to create a document of length N by using only the words in a vocabulary $\Sigma = \{w_1, w_2, ..., w_n\}$. The document is created with the following procedure: for each position in a document, you pick a word w_i from the vocabulary with probability p_i, where $\sum_i p_i = 1$. What is the probability that you will end up with a document where the occurrence of the word w_i (for each $w_i \in \Sigma$) in the document is exactly t_i, assuming $\sum_i t_i = N$?

Hint: As (b) shows, there will be many documents that contain exactly t_i w_i's. The answer to (c) should be the sum of the probabilities of all these documents.

Q4 (10 points): Suppose you want to build a trigram POS tagger. Let T be the size of the tagset and V be the size of the vocabulary.

(a) 2 pts: Write down the formula for calculating $P(w_1, ..., w_n, t_1, ..., t_n)$, where w_i is the i-th word in a sentence, and t_i is the tag for w_i.

(b) 8 pts: Suppose you will use an HMM package to implement the tagger.

- What does each state correspond to? How many states are there?
- What probabilities in the formula for (a) do transition probability a_{ij} and b_{jk} correspond to?

Q5 (10 points): In a POS tagging task, let V be the size of the vocabulary (i.e., the number of words), and T be the size of the tagset. Suppose we want to build a POS tagger that predicts the tag of the current word by using the following features:

1. Previous word w_{-1}
2. Current word w_0

3. Next word w_{+1}
4. Surrounding words w_{-1} w_{+1}
5. Previous tag t_{-1}
6. Previous two tags t_{-2} t_{-1}

(a) 3 pts: How many unique features are there **in total**?

(b) 2 pts: A classifier predicts class label y given the input x. In this task, what is x? what is y?

(c) 5 pts: For the sentence **Mike/NN likes/VBP cats/NNS**, write down the feature vector for each word in the sentence. The feature vector has the format "InstanceName classLabel featName1 val1 featName2 val2 ....". For the instanceName, just use the current word.

Q6 (30 free points): Same as Hw8 from Ling570, which is available at http://courses.washington.edu/ling570/fei_fall09/hw8.pdf.

Submission: In your submission, include the following:

- Your note file hw1.* that include your answers to Q1-Q5.
- If you did NOT take ling570 last quarter, you **must** turn in your code and solution to Q6 (The table/question numbers below refer to the ones in Hw8):
  - Shell scripts for proc_file.sh and create_vectors.sh, and the code called by the shell scripts.
  - The directories q2c, q2d and q2e created in Q2(c)-(e).
  - Completed Tables 1 and 2.
  - The answers to (e)-(h) in Q1.
  - No need to submit anything for (a)-(d) in Q1.