CS760 Spring 2016 Homework 3

Assigned Feb. 16, Due Feb. 24, 2016 before class

Instructions:

- Homeworks are to be done individually.
- Hand in a *electronic copy i.e., pdf* of your homework *before* class.
- We do not accept hand-written homeworks.
- Please show the derivations/calculations in detail.
- *Please note that a maximum of 1 late day can be used for this homework.*
- Use log to the base 2 for computing mutual information.
- Round off the final answers to 4 digits after the decimal point.
- *Include the following details on the first page of your submission.*

Name:

Email:

Student Id:

1. Consider the following training set with two boolean features and one continuous feature.

	A	**B**	**C**	**Class**
Instance 1	F	T	120	Benign
Instance 2	T	F	1090	Benign
Instance 3	T	T	245	Malignant
Instance 4	F	F	589	Malignant
Instance 5	T	T	877	Malignant

   (a) How much information about the class is gained by knowing whether or not the value of feature C is less than 475?

   (b) How much information about the class is gained by knowing whether or not the value of features A and B are different?

2. Suppose we want to learn a k-nearest neighbor model with the following data set and we are using Leave One Out Cross Validation (LOOCV) to select k. What would LOOCV pick - $k = 1$ or $k = 2$ or $k = 3$. Use Manhattan distance for calculations.

	Feature1	**Feature2**	**Class**
Instance 1	2	3	Positive
Instance 2	4	4	Positive
Instance 3	4	5	Negative
Instance 4	6	3	Positive
Instance 5	8	3	Negative
Instance 6	8	4	Negative

3. Suppose we wish to construct a Bayes Network for 3 features X, Y and Z using Sparse Candidate algorithm. We are given data from 100 independent experiments where each feature is binary and takes value **T** or **F**. Below is a table summarizing the observations of the experiment:

X	Y	Z	Count
T	T	T	36
T	T	F	4
T	F	T	2
T	F	F	8
F	T	T	9
F	T	F	1
F	F	T	8
F	F	F	32

(a) Suppose we wish to compute a single candidate parent for Z. In the first round of the sparse candidate algorithm, we compute the mutual information between Z and the other random variables.

i. Compute the mutual information between Z and X i.e. $I(X, Z)$ based on the frequencies observed in the data.

ii. Compute the mutual information between Z and Y i.e. $I(Y, Z)$ based on the frequencies observed in the data.

(b) Based on your observations in **part (a)**, which feature should be selected as candidate parent for Z? Why?

(c) In the first round of the algorithm, suppose that we choose Y to be the parent of Z in our network, X to be the parent of Y, and that X remains parent less. Estimate the parameters of the current Bayes net, given the data.

(d) In the next round of the algorithm, we wish to see if X can also be considered as a candidate parent for Z. To do this, we will use the *Kullback-Leibler* divergence between the marginal distributions of X and Z as estimated from the data $\hat{P}(X, Z)$ and that implied by the current network $P_{net}(X, Z)$. Compute the KL-divergence $D_{K,L}(\hat{P}(X, Z)||P_{net}(X, Z))$ between the two distributions.

(e) Based on your answer to **part (d)**, should we consider X as a candidate parent as Z? Explain.