Problem 1

a) **[8 points]** The entropies of each attribute for each side are as follows:

Expert	Entropy for "Good"	Entropy for "Bad"	Maximum
1	$(2/6)\log_2(1/(2/6))$ + $(4/6)\log_2(1/(4/6))$ = **0.9183**	$(4/6)\log_2(1/(4/6))$ + $(2/6)\log_2(1/(2/6))$ = **0.9183**	**0.9183**
2	$(5/7)\log_2(1/(5/7))$ + $(2/7)\log_2(1/(2/7))$ = **0.8631**	$(1/5)\log_2(1/(1/5))$ + $(4/5)\log_2(1/(4/5))$ = **0.7219**	**0.8631**
3	$(3/6)\log_2(1/(3/6))$ + $(3/6)\log_2(1/(3/6))$ = **1**	$(3/6)\log_2(1/(3/6))$ + $(3/6)\log_2(1/(3/6))$ = **1**	**1**
4	$(5/7)\log_2(1/(5/7))$ + $(2/7)\log_2(1/(2/7))$ = **0.8631**	$(1/5)\log_2(1/(1/5))$ + $(4/5)\log_2(1/(4/5))$ = **0.7219**	**0.8631**

Either **Expert 2** or **Expert 4** should be put at the root.

b) Suppose Expert 2 is put at the root.

[6 points] On the "Good" side of Expert 2:

Expert	Entropy for "Good"	Entropy for "Bad"	Maximum
1	$(2/4)\log_2(1/(2/4))$ + $(2/4)\log_2(1/(2/4))$ = **1**	$(3/3)\log_2(1/(3/3))$ + $(0/3)\log_2(1/(0/3))$ = **0**	**1**
3	$(2/2)\log_2(1/(2/2))$ + $(0/2)\log_2(1/(0/2))$ = **0**	$(3/5)\log_2(1/(3/5))$ + $(2/5)\log_2(1/(2/5))$ = **0.9710**	**0.9710**
4	$(4/5)\log_2(1/(4/5))$ + $(1/5)\log_2(1/(1/5))$ = **0.7219**	$(1/2)\log_2(1/(1/2))$ + $(1/2)\log_2(1/(1/2))$ = **1**	**1**

One this side, **Expert 3** should be used.

[6 points] On the "Bad" side of Expert 2:

Expert	Entropy for "Good"	Entropy for "Bad"	Maximum
1	$(0/2)\log_2(1/(0/2))$ + $(2/2)\log_2(1/(2/2))$ = **0**	$(1/3)\log_2(1/(1/3))$ + $(2/3)\log_2(1/(2/3))$ = **0.9183**	**0.9183**
3	$(1/4)\log_2(1/(1/4))$ + $(3/4)\log_2(1/(3/4))$ = **0.8113**	$(0/1)\log_2(1/(0/1))$ + $(1/1)\log_2(1/(1/1))$ = **0**	**0.8113**
4	$(1/2)\log_2(1/(1/2))$ + $(1/2)\log_2(1/(1/2))$ = **1**	$(0/3)\log_2(1/(0/3))$ + $(3/3)\log_2(1/(3/3))$ = **0**	**1**

One this side, **Expert 3** should be used.

[5 points] The outcomes at leaves are as follows:

Expert 2 / Expert 3	Good	Bad
Good	**Good (0)**	**Good (2)**
Bad	**Bad (0)**	**Bad (1)**

(x) is the number of CD's misclassified by this tree.

(ALTERNATIVE) Suppose Expert 4 is put at the root.

[6 points] On the "Good" side of Expert 4:

Expert	Entropy for "Good"	Entropy for "Bad"	Maximum
1	$(2/4)\log_2(1/(2/4))$ + $(2/4)\log_2(1/(2/4))$ = **1**	$(3/3)\log_2(1/(3/3))$ + $(0/3)\log_2(1/(0/3))$ = **0**	**1**
2	$(4/5)\log_2(1/(4/5))$ + $(1/5)\log_2(1/(1/5))$ = **0.7219**	$(1/2)\log_2(1/(1/2))$ + $(1/2)\log_2(1/(1/2))$ = **1**	**1**
3	$(3/4)\log_2(1/(3/4))$ + $(1/4)\log_2(1/(1/4))$ = **0.8113**	$(2/3)\log_2(1/(2/3))$ + $(1/3)\log_2(1/(1/3))$ = **0.9183**	**0.9183**

One this side, **Expert 3** should be used.

[6 points] On the "Bad" side of Expert 4:

Expert	Entropy for "Good"	Entropy for "Bad"	Maximum
1	$(0/2)\log_2(1/(0/2))$ + $(2/2)\log_2(1/(2/2))$ = **0**	$(1/3)\log_2(1/(1/3))$ + $(2/3)\log_2(1/(2/3))$ = **0.9183**	**0.9183**
2	$(1/2)\log_2(1/(1/2))$ + $(1/2)\log_2(1/(1/2))$ = **1**	$(0/3)\log_2(1/(0/3))$ + $(3/3)\log_2(1/(3/3))$ = **0**	**1**
3	$(0/2)\log_2(1/(0/2))$ + $(2/2)\log_2(1/(2/2))$ = **0**	$(1/3)\log_2(1/(1/3))$ + $(2/3)\log_2(1/(2/3))$ = **0.9183**	**0.9183**

One this side, either **Expert 1** or **Expert 3** should be used.

[5 points] The outcomes at leaves are as follows:

Expert 2 / Expert 3	Good	Bad
Good	**Good (1)**	**Good (1)**
Bad	**Bad (0)**	**Bad (1)**

(x) is the number of CD's misclassified by this tree.

Note:

[-3] Use log base 10, instead of log base 2.

[-9] Not do the entropy calculation in Part (b).

[-15] Misunderstand the algorithm.

Problem 2

There are no right or wrong answers for this problem. Full credits are given as long as the answer is supported by some correct reasoning. 10 points are deducted if no explanation is given. The suggested answers are as follows:

a) Building decision trees.
b) Finding highly correlated pairs.
c) Finding highly correlated pairs.
d) Clustering.
e) Finding highly correlated pairs.