1. We consider the training examples shown in the following table for a binary classification problem.

Instance	a_1	a_2	a_3	Target Class
1	T	T	1	+
2	T	T	6	+
3	T	F	5	-
4	F	F	4	+
5	F	T	7	-
6	F	T	3	-
7	F	F	8	-
8	T	F	7	+
9	F	T	5	-

a) What is the original entropy of this set of training instances?

The original entropy is $-\frac{4}{9}\log_2\frac{4}{9}-\frac{5}{9}\log_2\frac{5}{9}=0.991$ bit.

b) What are the information gains when a_1 and a_2 are used for partitioning the training set respectively?

After splitting on a_1, the entropy becomes

$$\frac{4}{9}(-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4})+\frac{5}{9}(-\frac{1}{5}\log_2\frac{1}{5}-\frac{4}{5}\log_2\frac{4}{5})=0.762 \text{ bit.}$$

As a result,

gain (a_1) = 0.991 – 0.762 = 0.229 bit.

After splitting on a_2, the entropy becomes

$$\frac{5}{9}(-\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5})+\frac{4}{9}(-\frac{2}{4}\log_2\frac{2}{4}-\frac{2}{4}\log_2\frac{2}{4})=0.984 \text{ bit.}$$

As a result,

gain (a_2) = 0.991 – 0.984 = 0.007 bit.