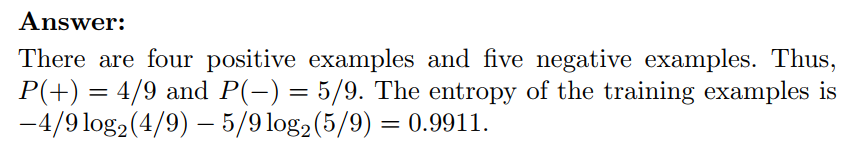
**Lab sheet N°9: Classification (1) (solutions)**

**Exercise 1:**

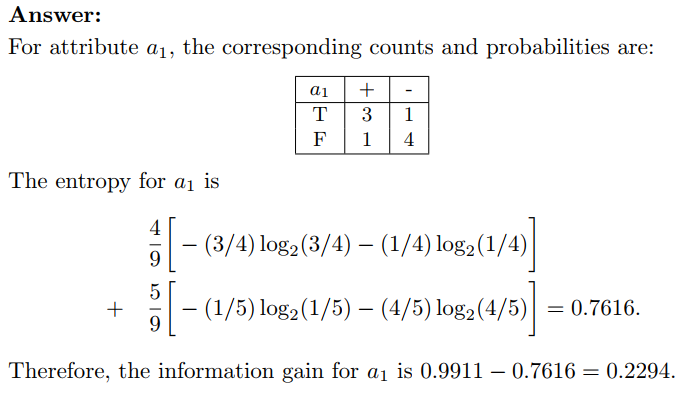
| **Instance** | **a1** | **a2** | **a3** | **Target Class** |
| --- | --- | --- | --- | --- |
| 1 | T | T | 1.0 | **+** |
| 2 | T | T | 6.0 | **+** |
| 3 | T | F | 5.0 | **-** |
| 4 | F | F | 4.0 | **+** |
| 5 | F | T | 7.0 | **-** |
| 6 | F | T | 3.0 | **-** |
| 7 | F | F | 8.0 | **-** |
| 8 | T | F | 7.0 | **+** |
| 9 | F | T | 5.0 | **-** |

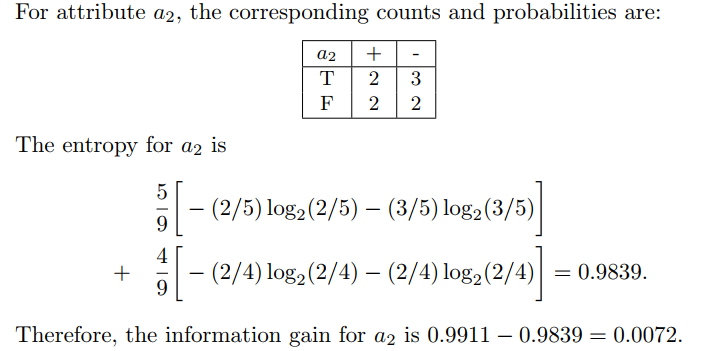
Consider the training examples shown in the table for a binary classification problem.

1. **What is the entropy of this collection of training examples with respect to the class attribute?**

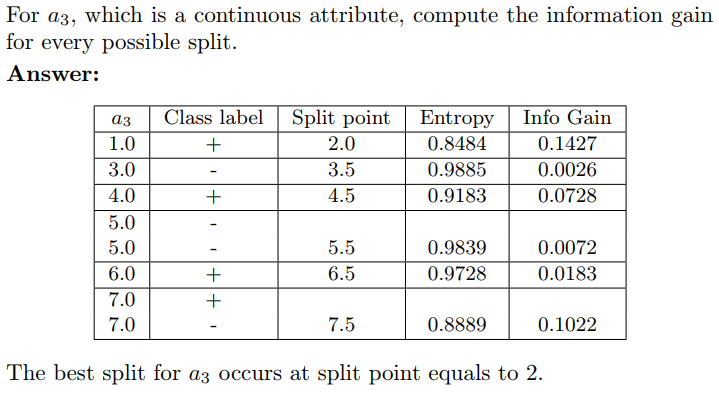


1. **What are the information gains of a1 and a2 relative to these training examples?**

****

****

1. **For a3, which is a continuous attribute, compute the information gain for every possible split.**

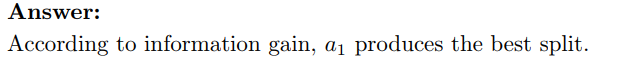
****

**For detailed calculations:** [Exercise1\_continuous\_split (solution).xlsx](https://docs.google.com/spreadsheets/d/1rn_GF2iT_lyHsC98IwcZ9mDCkLxsJssP/edit?usp=sharing&ouid=107299365240764970508&rtpof=true&sd=true)

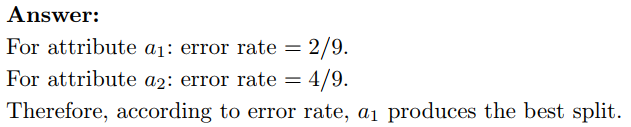
Information gain of the best split (a3): **0.142**

1. **What is the best split (among a1, a2, and a3) according to the information gain?**

**Information gain(a1) =0.229, Information gain(a2) =0.007, Information gain(a3) =0.142**



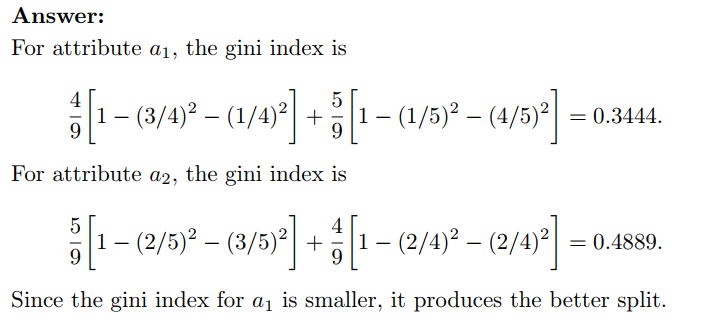
1. **What is the best split (between a1 and a2) according to the misclassification error rate?**



Collective misclassification error rate(a1) = 4/9 (1-3/4)+5/9(1-4/5) = 4/9\*1/4 +5/9\*1/5 = 2/9

Collective misclassification error rate(a2) = 5/9 (1-3/5)+4/9(1-1/2) = 5/9\*2/5 +4/9\*1/2 = 4/9

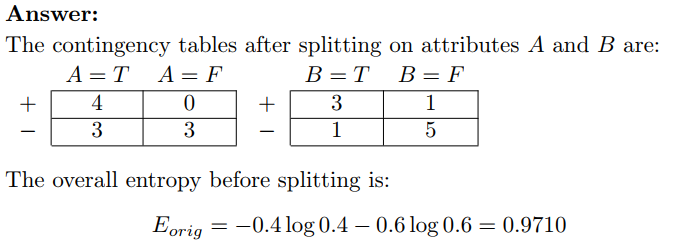
1. **What is the best split (between a1 and a2) according to the Gini index?**

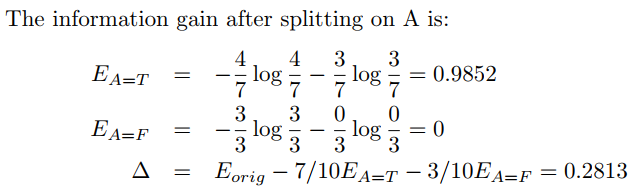


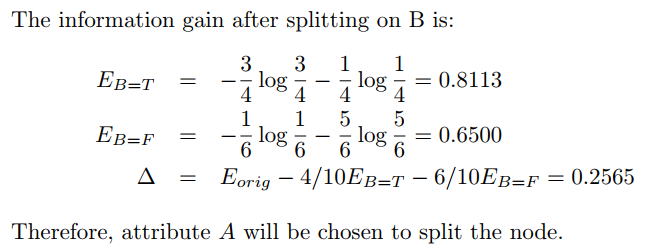
**Exercise 2 (For students):**

Consider the following data set for a binary class problem.

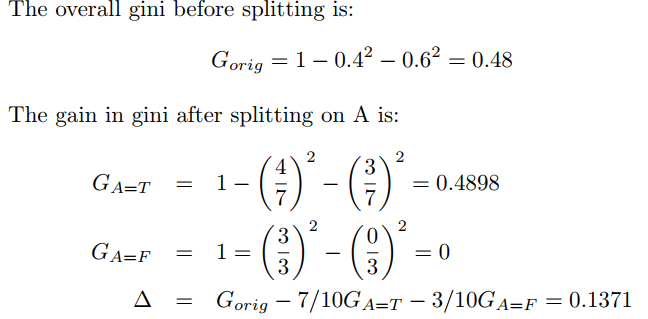
| **A** | **B** | **Class Label** |
| --- | --- | --- |
| T | F | **+** |
| T | T | **+** |
| T | T | **+** |
| T | F | **-** |
| T | T | **+** |
| F | F | **-** |
| F | F | **-** |
| F | F | **-** |
| T | T | **-** |
| T | F | **-** |

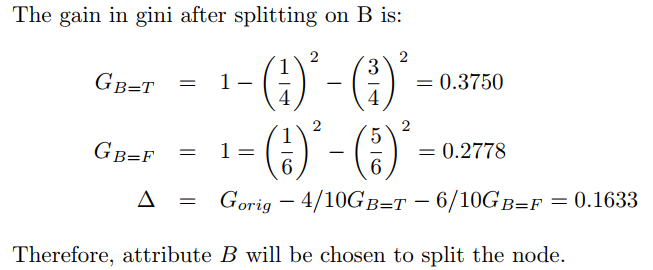
1. **Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?**

****

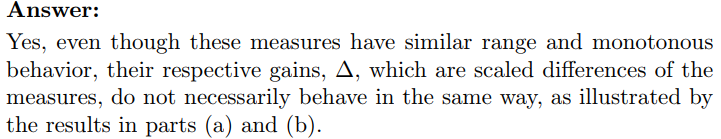
****

1. **Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?**

****

****

1. **In the lecture, we have shown that entropy and the Gini index are both monotonically increasing in the range [0, 0.5] and they are both monotonically decreasing in the range [0.5, 1]. Is it possible that information gain and the gain in the Gini index favor different attributes? Explain.**

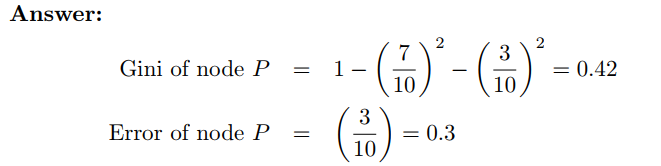
****

**Exercise 3:**

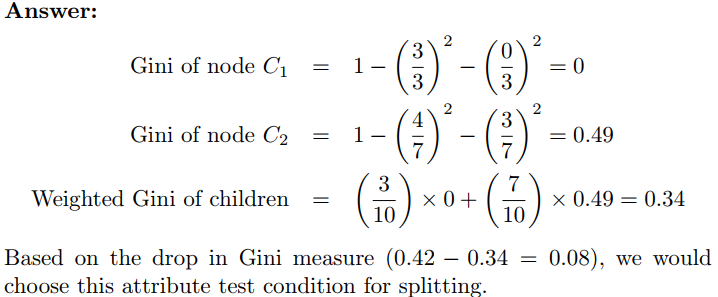
Consider splitting a parent node P into two child nodes, C1 and C2, using some attribute test condition. The composition of labeled training instances at every node is summarized in the table below.

|  | **P** | **C1** | **C2** |
| --- | --- | --- | --- |
| **Class 0** | 7 | 3 | 4 |
| **Class 1** | 3 | 0 | 3 |

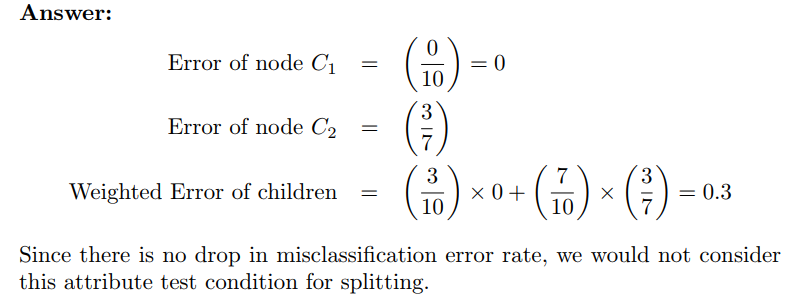
1. **Calculate the Gini index and misclassification error rate of the parent node P.**

****

1. **Calculate the weighted Gini index of the child nodes. Would you consider this attribute test condition if Gini is used as the impurity measure?**

****

1. **Calculate the weighted misclassification rate of the child nodes. Would you consider this attribute test condition if the misclassification rate is used as the impurity measure?**

****

**Exercise 4 (for students):**

Show that the entropy of a node never increases after splitting it into smaller successor nodes.

***Indication:*** *Jensen's inequality can be used without proof.*

