Exercise 1 :

Consider that we have the following data-points:

x_1	x_2	y
1	1	2
0	0	0
1	0	3
2	1	2

What is the Mean Squared Error (MSE) error for the following sets of parameters:

i) $w_0 = 0, w_1 = 1, w_2 = -1$

ii) $w_0 = 1, w_1 = 1, w_2 = 0$

Which model is preferable according to the MSE error function?

Exercise 2 :

Consider a database of cars represented by the five training examples below. The target attribute *Acceptable*, which can have values `yes` and `no`, is to be predicted based on the other attributes of the car in question. These attributes indicate a) the age of the car (*Age* having values $<$ `5 years` and $\geq$ `5 years`), b) the make of the car (*Make* having states `Toyota` and `Mazda`), c) the number of previous owners (*#Owners* having values `1`, `2` and `3`), d) the number of kilometers (*#Kilometers* having values $> 150k$ and $\leq 150k$) and e) the number of doors (*#Doors* having values 3 and 5).

	Attributes					Target
	Age	*Make*	*#Owners*	*#Kilometers*	*#Doors*	*Acceptable*
1	< 5	Mazda	1	$> 150k$	3	yes
2	≥ 5	Mazda	3	$> 150k$	3	no
3	≥ 5	Toyota	1	$\leq 150k$	3	no
4	≥ 5	Mazda	3	$> 150k$	5	yes
5	≥ 5	Toyota	2	$\leq 150k$	5	yes

a) Calculate the entropy for the attribute *#Owners*.[1]

b) Show the decision/classification tree that would be learned by the learning algorithm assuming that it is given the training examples in the database.

c) Show the value of the information gain for each candidate attribute at each step in the construction of the tree.

Exercise 3 :

Solve Exercise 7.3 (except sub-question f) in PM.

---

[1] Note that $\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$.

Exercise 4 :

Given below is a trainings data set about esoteric programming languages. We want to predict whether people are willing to learn the languages depending on Form, Usefulness, and Torment.

ID	Usefulness	Form	Torment	Appealing
1	useful	Other	pleasant	No
2	useless	Text	torture	No
3	useless	Text	torture	No
4	useless	Images	pleasant	Yes
5	useful	Images	pleasant	Yes
6	useful	Other	torture	No
7	useless	Images	torture	Yes
8	useless	Text	pleasant	Yes
9	useless	Text	pleasant	Yes
10	useful	Other	torture	No

(a) Compute the information gain associated with choosing the attribute Form as root.

(b) Draw a decision tree with Form as root.

(c) Use your decision tree to predict the following records:

Usefulness	Form	Torment
useful	Other	pleasant
useless	Text	torture

Exercise 5 :

Harry wants to invite his friends to a Pizza-Party. He does not want to ask them which pizza preferences they have in order to keep the party secret. He asks other people to get an impression what pizza preferences most people have.

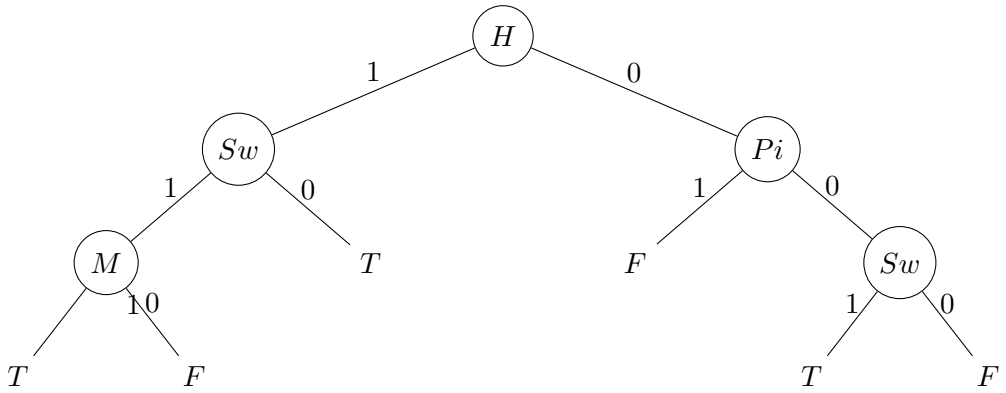
1. Use the Decision-Tree-Learning algorithm from the lecture to build up a decision tree for his observations depicted in the table below. Make sure to show your calculations. For each step, begin with the calculation of the entropy, and then after that, calculate the gain of each ingredient, calculating them in the same order as they are presented in the table ($Pineapple \rightarrow Mushrooms \rightarrow Ham \rightarrow Sweetcorn$). Explain which variable you take in each step and why. Draw the resulting Decision-Tree
2. Harry gets bored of all the calculations and decides to just pick variables at random. Why is this a bad idea. Specify what the resulting problem could be, and what the Decision-Tree-Learning algorithm tries to do.

Pineapple	Mushrooms	Ham	Sweetcorn	Rating
1	1	1	1	true
1	0	1	0	false
1	1	0	1	false
1	1	1	0	false
0	1	0	1	true
0	0	1	1	true
0	1	1	1	true
0	0	0	1	false
0	1	1	0	true
0	1	0	1	true

Exercise 6 :

In the table below the favorite pizzas of Harry's friends are listed. For each of the pizzas, use the following tree to say if Harry orders this pizza or not. (Note that this is not necessarily the solution to Exercise 5). Do not forget to declare your steps. We use the following abbreviations for the ingredients: Pi is Pineapple, M is Mushrooms, H is Ham, Sw is Sweetcorn.

Hermione	Mushroom, Ham, Sweetcorn and Onion
Neville	Pineapple, Sweetcorn and Mushrooms
Ron	Salami and Sweetcorn
Ginny	Ham
Luna	Salami, Bolognese sauce and Pepper



Exercise 7 :

Consider a fair six-sided die and another loaded die such that the probability of obtaining an outcome of 6 is 50%, the probabiliy of obtaining 5 is 25% and the remaining outcomes are equally likely. Calculate the entropy for both dice. How do the results differ? Explain!

Exercise 8 :

Download and install the WEKA data-mining toolbox:

http://www.cs.waikato.ac.nz/ml/weka/

WEKA provides several user-interfaces. Select the 'Explorer' interface from the 'Applications' menu, and try the following:

- Load the 'Iris' dataset. This dataset contains measurements from 150 indi- vidual plants of the genus Iris, belonging to 3 different species 'Iris setosa', 'Iris versicolor', and 'Iris virginica'. The machine learning task associated with this dataset is: predict the species from the four measurement values.
- Use the 'Visualize' tab to get an overview of the attribute values and their relation to the class label. Sketch by hand a small decision tree for predicting the class label.
- Use WEKA's decision tree construction methods to build a decison tree (under the 'Classify' tab select e.g. J48 or the SimpleCart classifier). Compare with your own proposed decision tree.

Exercise 9 (Optional):

Download the Pregnancy dataset. Note that the format of this file does not follow the standard file-format used by Weka. When trying to load the file you will therefore have to use the 'converter' suggested by Weka.

- Construct a decision tree for classification. Try to reason about the structure of the tree. Hint: have a look at the underlying Bayesian network model (which can be found here) that we have previously looked at in the course.
