COSC 2753 | Machine Learning

Week 7+1 Lab Exercises: **Rule Learning**

⌄ Introduction

In this lab you will be:

1. Implement the entropy calculation
2. Implement a simplified proposition rule learning algorithm, outputting rules

sklearn does not have an implementation of a rule learner. Instead you will implement a simplified CN2 algorithm. This algorithm will construct pre-conditions that contain a single term, that is, the rule precondition will not contain conjunctions. This will require you to implement functions in python, and use simple loops and if-statements. If you are unfamiliar with these, first revise the Python tutorials from Lab01.

This lab only requires Pandas/Numpy to load with work with the data set, and the math library.

```
1 import pandas as pd
2 import numpy as np
3 import math
```

⌄ Datasets

You will be looking at two data sets for this lab which you have seen before:

1. Sailing days
2. Zoo (animal) classification

You can download these from Canvas or BitBucket code repo.

```
1 sailData = pd.read_csv('./Lab/sailing-custom-python.txt',delim_whitespace=True)
2 zooData = pd.read_csv('./Lab/zoo-python.txt',delim_whitespace=True)
```

```
1 sailData.head()
```

	Outlook	Company	Sailboat	Sail
0	rainy	big	big	yes
1	rainy	big	small	yes
2	rainy	med	big	no
3	rainy	med	small	no
4	sunny	big	big	yes

```
1 zooData.head()
```

	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	type	name
0	Yes	No	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	4.0	No	No	Yes	mammal	aardvark
1	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	No	No	4.0	Yes	No	Yes	mammal	antelope
2	No	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes	0.0	Yes	No	No	fish	bass
3	Yes	No	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	4.0	No	No	Yes	mammal	bear
4	Yes	No	No	Yes	No	No	Yes	Yes	Yes	Yes	No	No	4.0	Yes	No	Yes	mammal	boar

remove unnecessary columns:

```
1 zooData = zooData.drop(columns='name')
```

› Simple Rule Learner

You will develop the simple rule learner over three parts:

1. Entropy calculation function
2. Majority class calculation function
3. Rule learner

Entropy function

First you will need a function that calculates the entropy of a data set.

```
[ ]  ↳ 24 cells hidden
```

Sample Solutions

If you are struggling with the first two functions, a sample solution has been provided for these. Only use this if you have **made your absolute best attempts** at implementing these functions yourself. The purpose of this lab is to understand common aspects of symbolic machine learning algorithms, though the CN2 algorithm. You will gain significantly less out of this lab if you don't try to solve the problems yourself.