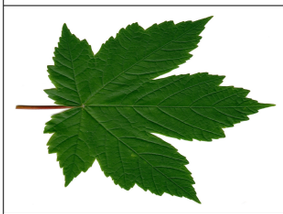
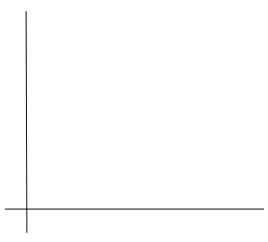
ARTIFICIAL INTELLIGENCE: Understanding Data

STUDENT REGISTRATION ID (NRP): _____
NAME: _____
CLASS: _____

Picture	Attribute	Species
	Width : 4.0 cm Height: 2.5 cm Width/Height: ____	A
	Width : 4.0 cm Height: 3.5 cm Width/Height: ____	B
	Width: 2.8 cm Height: 1.5 cm Width/Height: ____	A
	Width: 3.5 cm Height: 1.8 cm Width/Height: ____	A
	Width: 5.1 cm Height: 4.9 cm Width/Height: ____	B
	Width: 2.0 cm Height: 2.1 cm Width/Height: ____	B

ACTIVITY: Plot the data above based on these features

Width	Height	Width vs Height	Width/Height
			

ARTIFICIAL INTELLIGENCE: Linear Regression

STUDENT REGISTRATION ID (NRP): _____
NAME: _____
CLASS: _____

ACTIVITY

1. Try to guess the value of x
   * [1,2], [2,3], [3,4], [4,x]
   * [1,1], [2,4], [3,9], [4,x]

2. Install this app on your phone: https://play.google.com/store/apps/details?id=com.successcrazzy.datascience101 (You can use `Data Science 101` as keyword)

3. Read the `Linear Regression` section

4. Try to guess the value of x
   * [2,10], [4,9], [3,6], [6,6], [8,6], [8,3], [10,2], [12, x]

5. Try to make a program by using scikit-learn (http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html) to do the 4^{th} task.

ARTIFICIAL INTELLIGENCE: Decision Tree

STUDENT REGISTRATION ID (NRP): _____
NAME: _____
CLASS: _____

ACTIVITY

1. Install this app on your phone: https://play.google.com/store/apps/details?id=com.successcrazzy.datascience101 (You can use `Data Science 101` as keyword)

2. Read the `Decision Tree` section

3. How to calculate Information, Entropy, and Gain?

4. Consider the following dataset, determine the choice for the last data

Size	Shape	Color	Choice
M	Brick	Blue	Yes
S	Wedge	Red	No
L	Wedge	Red	No
S	Sphere	Red	Yes
L	Pillar	Green	Yes
L	Pillar	Red	No
L	Sphere	Green	Yes
M	Pillar	Green	?

5. Build a decision tree program for the case above by using scikit-learn (http://scikit-learn.org/stable/modules/tree.html)

6. Consider this ruleset:
   * finalScore = 0.2 * assignment + 0.3 * midTest + 0.5 * finalTest
   * if finalScore >= 80, then finalMark = A
   * if finalScore < 80 and finalScore >= 70, then finalMark = B
   * if finalScore < 70 and finalScore >= 60, then finalMark = C
   * if finalScore < 60 and finalScore >= 40, then finalMark = D
   * if finalScore < 40, then finalMark = E

   You can collect the student data and build a decision tree to determine the finalMark.
   Do you think that using the decision tree to determine the finalMark is a good decision? Why? Or why not?