Tasks: Machine Learning and Statistics

ian.mcloughlin@atu.ie

Last updated: 13 September 2023

1. Square roots are difficult to calculate. In Python, you typically use the power operator (a double asterisk) or a package such as `math`. In this task,[1] you should write a function `sqrt(x)` to approximate the square root of a floating point number x without using the power operator or a package.

   Rather, you should use the Newton's method.[2] Start with an initial guess for the square root called z_0. You then repeatedly improve it using the following formula, until the difference between some previous guess z_i and the next z_{i+1} is less than some threshold, say 0.01.

$$z_{i+1} = z_i - \frac{z_i \times z_i - x}{2z_i}$$

2. Consider the below contingency table based on a survey asking respondents whether they prefer coffee or tea and whether they prefer plain or chocolate biscuits. Use `scipy.stats` to perform a chi-squared test to see whether there is any evidence of an association between drink preference and biscuit preference in this instance.

		Biscuit	
		Chocolate	**Plain**
Drink	**Coffee**	43	57
	Tea	56	45

3. Perform a t-test on the famous penguins data set[3] to investigate whether there is evidence of a significant difference in the body mass of male and female gentoo penguins.

4. Using the famous iris data set,[4] suggest whether the setosa class is easily separable from the other two classes. Provide evidence for your answer.

5. Perform Principal Component Analysis on the iris data set,[5] reducing the number of dimensions to two. Explain the purpose of the analysis and your results.

[1] *A Tour of Go*. Aug. 18, 2023. URL: `https://go.dev/tour/flowcontrol/8` (visited on 08/18/2023).

[2] *Square Roots via Newton's Method*. Feb. 4, 2015. URL: `https://math.mit.edu/~stevenj/18.335/newton-sqrt.pdf` (visited on 08/18/2023).

[3] *mwaskom/seaborn-data: Data repository for seaborn examples*. Aug. 30, 2023. URL: `https://github.com/mwaskom/seaborn-data/blob/master/penguins.csv` (visited on 08/30/2023).

[4] *Iris – UCI Machine Learning Repository*. Aug. 17, 2023. URL: `https://archive.ics.uci.edu/dataset/53/iris` (visited on 08/17/2023).

[5] *Iris – UCI Machine Learning Repository*.