

Data Hackathon 2019

King's College London Health Science DTC

2019 04 08

Our goals for today

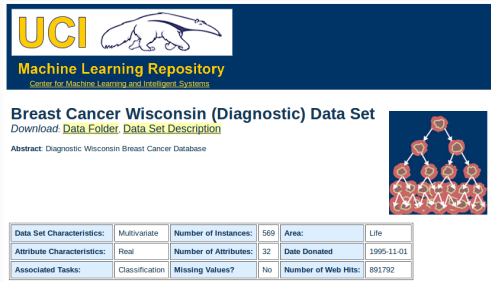
- Learn about **Data Science** approaches.
- Use the time today to **code together** and **exchange ideas and experiences**.
- Using the **group effort to solve computational biology orientated challenge**.
- A group of experienced programmers (Anna, Joseph, Mat and Paul) will be available to help during the session.

The goal is: **to learn about and use machine learning algorithms for classification:**

- Setup a set of common machine learning models.
- Obtaining unbiased estimates of a model's performance.
- Fine-tune machine learning models.
- Evaluate predictive models using different performance metrics.

Coding challenges for today

Dataset of interest: Breast Cancer Wisconsin, which contains 569 samples of malignant and benign tumour cells.



The screenshot shows the UCI Machine Learning Repository website. At the top is the UCI logo with a sloth and the text "Machine Learning Repository" and "Center for Machine Learning and Intelligent Systems". Below this is the title "Breast Cancer Wisconsin (Diagnostic) Data Set" with links for "Data Folder" and "Data Set Description". An abstract describes it as the "Diagnostic Wisconsin Breast Cancer Database". To the right is a diagram of a neural network. At the bottom is a table with dataset characteristics.

Data Set Characteristics:	Multivariate	Number of Instances:	569	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	32	Date Donated	1995-11-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	891792

Figure 1: UCI machine learning repository

Your task will be to build a model to predict whether a tumour is benign or malignant.

Happy Hacking!

- General information about the Hackathon:
https://khsdtc.github.io/Hackathon_Spring2019/
- Hacking challenge is placed at: https://github.com/KHSDTC/Hackathon_Spring2019_Challenge