

# Model Interpretation Methods

Welcome to the final assignment of course 3! In this assignment we will focus on the interpretation of machine learning and deep learning models. Using the techniques we've learned this week we'll revisit some of the models we've built throughout the course and try to understand a little more about what they're doing.

In this assignment you'll use various methods to interpret different types of machine learning models. In particular, you'll learn about the following topics:

- Interpreting Deep Learning Models
  - Understanding output using GradCAMs
- Feature Importance in Machine Learning
  - Permutation Method
  - SHAP Values

Let's get started.

## This assignment covers the following topics:

- 1. Interpreting Deep Learning Models
  - 1.1 GradCAM
    - 1.1.1 Getting Intermediate Layers
    - 1.1.2 Getting Gradients
    - 1.1.3 Implementing GradCAM
      - Exercise 1
    - 1.1.4 Using GradCAM to Visualize Multiple Labels
      - Exercise 2
- 2. Feature Importance in Machine Learning
  - 2.1 Permutation Method for Feature Importance
    - 2.1.1 Implementing Permutation
      - Exercise 3
    - 2.1.2 Implementing Importance
      - Exercise 4
    - 2.1.3 Computing our Feature Importance
  - 2.2 Shapley Values for Random Forests
    - 2.2.1 Visualizing Feature Importance on Specific Individuals
    - 2.2.2 Visualizing Feature Importance on Aggregate