Springboard--DSC Program
Capstone Project 1 Proposal

Histopathologic Cancer Detection

By Morgan Fry

January 2020

The Problem

Cancer detection is commonly done by human visual inspection of biopsy samples treated with contrast. Healthcare providers and patients both have an interest in faster and more accurate analysis of these, as more accurate and earlier detection can ultimately improve treatment outcomes. A sufficiently accurate model for detection may save analysts time by reducing their workload to chiefly edge cases or may improve on human results altogether.

The Data

Data from Kaggle Competition 'Histopathologic Cancer Detection' will be used to train and test models. This dataset consists of images of slides of lymph node samples treated with contrast, along with their classification of positive or negative for cancer.

Proposed solution

Proposed process to find a solution will be:

- Convert the dataset to a useable format
- Explore the dataset through storytelling and applications of inferential statistics
- Try machine learning algorithms of increasing complexity and evaluate their effectiveness for image classification of this type.

Deliverables

Report with slide deck and Jupyter notebooks with code.