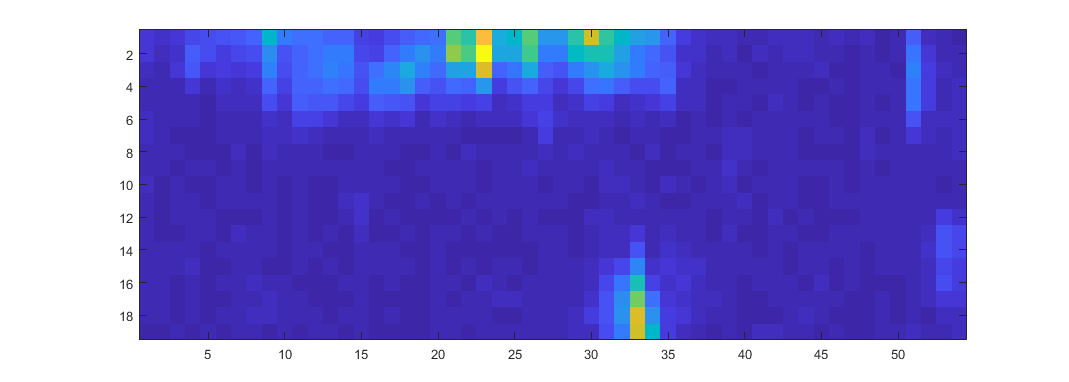
Cancer Cell Classification via Multiple Supervised Learning Methods

By: Jake Rhodes, Erick Stone, Andres Correa

Our dataset consists of two types of cancer cells: 231 (breast cancer) and A549 (lung cancer). Each of cell types treated varying dosages of two drugs, metformin (at five different dosage levels) and IFN gamma (at three different dosage levels.

Given the two cell types and varying dosages of the two drugs, we have twenty different groups of cells (including control groups) and from five to fifteen samples from each group.

Each cell underwent a series of non-invasive chemical analysis called Raman spectroscopy. The result was, for each cell, an m-by-n grid of Raman spectra, which can be reduced to a single image of intensity levels. The measure of intensity might be determined by the single highest peak of the Raman spectrum or by the average of the Raman measurements.

**Example:** Image using top peak intensity for pixel measurement

Our goal is to be able to use a single pixel or small array of pixels to classify between

1. The two types of cancer cells
2. The treatment groups given a type of cancer cell

As a part of the analysis, we would like to be able to determine the locations in the image where the cell is present or absent. Where the cell is present, we will try to classify into groups as described in (a) and (b) above.

Proposed Algorithms:

1. MCR-ALS (Multivariate curve resolution, for cell presence detection)
2. Random Forests (Non-linear classification)
3. Logistic Regression (Linear classification)
4. SVM (Linear classification)
5. Other methods as seen fit

Contributions:

Jake - Data organization and MCR-ALS analysis

Erick - Random Forests and Logistic Regression analysis

Andres - SVM and other methods