CS532 Project/Lab Proposal
Due: 10/24/15

Tentative Title: Predicting bone tumor response using imaging features

Team Members: Christie Lin, Alison Roth, Matthew Scarpelli

Brief Overview of Topic and Motivation:

The three of us are in the Image-Guided Therapy Lab group within the School of Medicine and Public Health. We are interested in investigating how to improve patient-specific disease management. We want to use molecular imaging to predict patient outcome.

We will try to use imaging features derived from PET imaging to predict tumor response to therapy. Successfully predicting response to treatment would allow for adaptive therapy early during treatment and better management of patient. We will build a classifier using least squares to separate those patients who will respond better or worse to treatment. We also might utilize clustering to determine if there are subpopulations of lesions.

The goal is to determine if tumor lesions respond to the therapy, and can be categorized as responding or non-responding lesions. Overall, we have the imaging feature values of the lesion at three time points. We will see if it is possible to predict the response at the third time point based on the values at the first two time points. To test if our algorithm is accurate, we will use a set of lesions as the training set where we know the imaging feature values at all three time points, and then evaluate the accuracy on the testing set.

Core Concepts:
The core concepts that may be involved in our project are:

- linear dimensionality reduction
    - We will reduce the number of features we are using to predict since some of them are highly correlated or not useful for prediction
- support vector machines
    - Support vector machines can be used for regression analysis to categorize if each tumor is a responder or non-responder.
- independent component analysis
- spectral clustering
    - We can cluster the tumors into similarly behaving groups, or subpopulations.
- Regularization
    - Regularization is needed since we have a large number of imaging features relative to the number of tumors thus, the setup will likely be underdetermined

Related Papers, Datasets, or Resources:

Dataset

We will study a real dataset of tumor imaging features where there are m lesions and n imaging features at three time points. This data is from our lab which can be anonymized and shared with our class.

Papers and Resources

- **Machine Learning methods for Quantitative Radiomic Biomarkers.** Chintan Parmar, Patrick Grossmann, Johan Bussink, Philippe Lambin & Hugo J. W. L. Aerts. *Scientific Reports* 5, Article number: 13087 (2015) doi:10.1038/srep13087. http://www.nature.com/articles/srep13087
- **A Tutorial on Spectral Clustering.** Ulrike von Luxburg. *Statistics and Computing*, 17 (4), 2007.http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/attachments/Luxburg07_tutorial_4488%5b0%5d.pdf
- **Support vector machine classification and validation of cancer tissue samples using microarray expression data.** Terrence Furey et al. *Bioinformatics* (2000) 16 (10):906-914.doi:10.1093/bioinformatics/16.10.906. http://bioinformatics.oxfordjournals.org/content/16/10/906.abstract