WEEK 2

CLASS 3: MACHINE LEARNING, K-NEAREST NAIGHBORS (KNN)

- Iris dataset [What does an iris look like?]
- [Data] hosted by the UCI Machine Learning Repository

Machine learning and KNN

- Reddit AMA with Yann LeCun
- Characteristics of your zip code
- Introduction to scikit-learn (Classroom Notebook)
- Documentation: [user guide] [module reference] [class documentation]

Homework:

- Complete your visualization homework assigned in class 2
- Reading assignment on the Pre-work Bias-Variance tradeoff

Optional:

- For a more in-depth introduction to machine learning, read section 2.1 (14 pages) of Hastie and Tibshirani's excellent book, [An Introduction to Statistical Learning]. (It's a free PDF download!)
- For a really nice comparison of supervised versus unsupervised learning, plus an introduction to reinforcement learning, watch this [video] (13 minutes) from Caltech's [Learning From Data] course.

Resources:

- Quora has a [data science topic FAQ] with lots of interesting Q&A.
- Five most popular similarity measures – Euclidean, Manhattan, Minkowski, Cosine Similariy, and Jaccard Similarity
- Intuitive Explanation of Bias Variance – discussion on Quora

CLASS 4: BIAS-VARIANCE TRADEOFF AND MODEL EVALUATION

- Exploring the bias-variance tradeoff (Classroom Notebook)
- Discussion of the [assigned reading] on the bias-variance tradeoff
- Model evaluation procedures (Classroom Notebook)

Homework:

- Complete the glass_id problem.

Resources:

- If you would like to learn the IPython Notebook, the official [Notebook tutorials] are useful.
- To get started with Seaborn for visualization, the official website has a series of [tutorials] and an [example gallery].
- Hastie and Tibshirani have an excellent [video] (12 minutes, starting at 2:34) that covers training error versus testing error, the bias-variance tradeoff, and train/test split (which they call the "validation set approach").
- Caltech's Learning from Data course includes a fantastic [video] (15 minutes) that may help you to visualize bias and variance.