Summer 2017 Research: Machine Learning, Face Recognition, and Disease Prediction

1. Learning about machine learning and face recognition
   1. Data Mining textbook
   2. Machine Learning is Fun! series
      1. Face detection—gradients
      2. Face landmarks—projection
      3. Face encodings—neural network
      4. Face comparisons—match to attribute
2. ColorFERET database
   1. Race/gender metadata
   2. Use Python face\_recognition library to generate 128 number encoding
   3. Run through WEKA classifiers to develop machine learning scheme
      1. Gender classification—very good
      2. Race classification—not very good
3. Other databases
   1. LFW
      1. Metadata—names, but no race/gender
      2. Genderize.io API to find most likely gender (and confidence level) for each name
      3. Some manual gender partitioning
      4. Gender classification using WEKA
   2. FDDB
      1. No metadata
      2. Make gender predictions with WEKA by using other databases as training sets, then FDDB as test set
   3. IMDb
      1. Age/gender metadata, but no race
      2. Much larger database
4. Disease prediction
   1. 22q11.2 deletion syndrome
   2. Pictures from Kruszka paper, NHGRI website, and a few other places
      1. Age, race, gender metadata
   3. Age partitions—includes all races
      1. 0 to 3 y/o
      2. 4 to 12 y/o
      3. 13 y/o and up
   4. Race partitions—includes all ages
      1. Asian
      2. Black/African-American
      3. Latino
   5. Control pictures
      1. IMDb database (age, no race) for 12 and under age partitions
      2. ColorFERET database (race, no age) for 13 and up and for race partitions
   6. WEKA results
      1. Several classification methods including SMO (SVM), Naïve Bayes, Multilayer Perceptron (neural network), Voted Perceptron
      2. Best results from SMO and Multilayer Perceptron methods
      3. Accuracy above 90% for all age and race partitions
      4. Highest accuracy for Latino partition (98%)
5. Future directions
   1. Build classifier for other diseases, including Down syndrome
   2. Compare between diseases
   3. Supplement to doctors’ manual predictions, not replacement