Healthcare Capstone Project

The following are a set of available materials, datasets, and questions to get you started thinking about how to apply machine learning to better understanding healthcare.

Available Course

This free course, developed mutually by Udacity and Georgia Tech, could be a valuable resource if you are interested in exploring machine learning in healthcare more deeply:
Health Informatics in the Cloud

Available Datasets

The following datasets are available in a number of different formats, including CSV, and can be processed by the tools and techniques of your choice that you have learned thus far in this Nanodegree program.

- MIMIC, from the MIT Lab for Computational Physiology
  - **Note that MIMIC is not open source will require completion of a CITI "Data or Specimens Only Research" course to get access.**
- HealthData.gov, from the United States Department of Health and Human Services
- National Survey of Family Growth, from the Center for Disease Control and Prevention
- Medicare Datasets, from the United States Department of Health and Human Services

You are also welcome to find additional educational datasets and pursue your own research.

If you have elected to use the MIMIC dataset, please use the following information when requesting access to the dataset (after completing the certification course):

Organization: Udacity
Title or position: Nanodegree student
Reference Category: Supervisor
Reference: Farhan Ahmad, Learning Technologist (farhan@udacity.com)
Research area: Machine Learning Engineer Nanodegree Capstone project, or Data Scientist Nanodegree Capstone project

Sample Questions

The following are some questions you may answer with the datasets above. You are also free and encouraged to develop your own questions based on the datasets in which you are interested; these are just intended to get you started.

- Can patient healthcare records predict the likelihood of mortality in an emergency room visit? (Answerable with MIMIC)
- What is the trade-off between more-effective medications with more side effects to manage and less-effective ones with fewer side effects? (Answerable with MIMIC)
- During labor, when should natural labor be interrupted for a C-section to maximize the health of both the mother and the child? (Answerable with MIMIC)
- What medications should be administered when multiple conflicting symptoms are present? (Answerable with MIMIC)

Relevant Resources

- The Future of Health Care is in the Data, from the Carnegie-Mellon Machine Learning Department
- Big Data in Healthcare Made Simple: Where It Stands Today and Where It's Going, by Doug Adamson
- The big-data revolution in US health care: Accelerating value and innovation, by Basel Kayyali, David Knott, and Steve Van Kuiken