Programming Assignment 1: Predicting Newborn's Weight

Spring 2019 Machine Learning

I am quoting from the WebMD article [1]:
"A new baby's gender, name, time of birth, and birth weight are nice information for a birth announcement, but birth weight is especially important for an obstetrician. A large size at delivery has long been associated with an increased risk of injuries to a newborn and its mom. So the better a doctor can predict birth weight, the easier the delivery may be."

Ultrasound is a popular way of doing it. But, aha! You are a Data Scientist (or going to be one). You can amaze people by predicting the birth weight way earlier than ultrasound, right? In this assignment, let's do this.

Dataset

- Login to Canvas > Assignments > Programming Assignment 1
- You will get the following 3 files
  - **baby-weights-dataset2.csv**
    - It has **101400** rows (samples) with **37** columns (variables). Each sample represent a case for a new-born. It contains 37 variables (just mentioned! Haha) about it. Very last column of it is "BWEIGHT", that true weight of the new-born (in lbs unit). Actually, this needs to be considered as the target variable here.
  - **data-description.txt**
    - You will see that the name of the 37 variables are actually contracted form of some sort. And, the source of the dataset did not offer me description of every single of them. But, after studying about them, I could elaborate only few of them. Please pardon my laziness. Okay, this file contains few descriptions for the variables. All the rest are mostly talking about the Mother's medical history and all. No big deal, I guess, for you to work with these variables without knowing their meaning.
  - **judge-without-label.csv**
    - This is an interesting file. It contains new samples: additional 2001 rows with 36 columns (without the BWEIGHT target column). Once again, this should be part of the training, as there are no ground truth target labels, right? Once the training is complete with the dataset provided above, you must apply your prediction algorithm to predict BWEIGHT of these 2001 samples, and submit the result as part of your assignment submission.

Tasks

- Please read the PA1-skeleton-Ashis-Biswas.ipynb file using Jupyter Notebook to learn about 15 mandatory tasks, and 2 additional tasks for graduate students (CSCI-5930).

Good luck!