Objective

Learn how to use an off-the-shelf tool to explain a regression model for medical expense prediction. This lab will focus on LIME.

As before, the notebook will automatically grab the data with wget.

Your Tasks

- **(TURN THIS IN, 5 points)** First, read the assignment specification and ***estimate how long you think it will take you*** and write it down.

1. *Task 1: Expectations before beginning*
    A. **[Each, TURN THIS IN, 10 points]** Before beginning this assignment, briefly characterize in writing your expectation of what you will find to be predictive of an individual having high medical expenses. Also characterize any expectations of the explanations we might see during the lab.
2. *Task 2: Wrangle Data*
    A. Check out the allFeaturesList variable in the first cell of this section. Divide those feature labels up amongst the team members and learn what they mean. The codebookFIXME will be helpful here (I recommend searching from the bottom)
    B. **[Team, TURN THIS IN, 10 points]** Before proceeding, discuss among the team members which features everyone thinks might be predictive of high medical expenses. Come to some consensus on which would be good hypotheses and write them down now.
    C. **[Team, TURN THIS IN, 10 points]** Pick 8 features you think will be important and write down what they mean and what values they can take on (be sure to include those you identified in Task 2B).
    D. **[Each, TURN THIS IN, 10 points]** Look at the processing function in the first cell. It does two things: (1) drop rows from the data table which contain cells with missing data and (2) drop a bunch of features to make the rest of this notebook more computationally tractable. Pick 3 features out of the codebook that we cut from the dataframe and you think might be good to keep and describe why you think those features might be valuable.

3. *Task 3: Investigate LIME explanations on the MEPS dataset*

    A. **[Team, TURN THIS IN, 20 points]** Inspect the given explanations from LIME, examining indices 35, 273, and 300, at least. Describe what you observe within the explanations from LIME.
4. *Task 4: Explaining multiclass text prediction*
    A. **[Team, TURN THIS IN, 20 points]** Understand the given explanations from LIME, provide your interpretation to explanations given by LIME. Be sure to highlight ways the given explanations from LIME matches or does not match your judgement. Feel free to modify the helper function to assist in gaining an understanding of this model
5. Task 5: Big Picture
    A. **[Team, TURN THIS IN, 10 points]** Dig around in the links provided in the first cell and learn a bit about some of the other features LIME offers. Describe a few that look interesting to you and why.
    B. **[Each, TURN THIS IN, 10 points]** Having finished this lab, re-read what you wrote for task A1. How has your position changed?
    C. **[Each, TURN THIS IN, 5 points]** Upon completing the lab, determine how long you actually spent on the lab, and report that timeframe in addition to your estimate beforehand.