**Homework #7: Machine learning on a cardiac health dataset**

Questions:

1. **What question are you trying to answer?**
   1. Do any variables have a higher correlation to death than others? Mostly, do any have any of significance?
2. **Which machine learning strategy/strategies will you use to answer your question?**
   1. To answer the previous question about correlation, I will use Lasso regularization. However, first I want to see general correlation values with a heat map. To do this I will use the seaborn package, as discussed in class. Then I will use the Lasso regularization step by step process to look at trained values v test values and see which are non-zero values. Thus, I can see variables that do have a significant correlation to death, whether that be negative or positive. I will do this by creating different functions, one for going through regularization, and then one for finding which are significant.
3. **[Submit the code you wrote to answer your question] (DONE)**
4. **[Submit unit tests for any functions you wrote] (DONE)**
5. **What were your results?**
   1. [('creatinine\_phosphokinase', 6.366701882671068e-06), ('ejection\_fraction', -0.004898767608426774), ('platelets', -6.50086892650224e-08)]
   2. This shows that these three variables have significant enough correlations for causing death based on data. Although the specific values depending on what run you do, because of the training value, all three are consistently found to have significant correlation.
   3. This is very helpful when you compare it to the heatmap. The heatmap is clearly crowded and not always able to see the exact significance of the correlation. When we implement Lasso regularization, we can find a line that fits each for the variables, thus slightly generalizing and then finding which are significant. The scatter of points makes it difficult to discern and this method shows the significant of these variables and levels.
6. **Provide a visualization of your machine learning algorithm results**.
   1. Below is the heat map created from the seaborn package for correlations of values

Chart

Description automatically generated

Seaborn heat map from the variable correlations.