**Modeling**

Walkthrough Video:

<https://berkeley.zoom.us/rec/share/22513E-n3ZOwlHfetLwVK6VTIExo8SQVmVYv1j1-IVUjwDTEHso_8MPInn-d0Aae.MCTHoeBIjuRaBoJd?startTime=1603596789000>

In this week’s example, we explored ways that we could go about to determine a good approximation of what a constant value might be for our charges. That is, given no other variables at play with no information besides the charges column, the best value we can give as an approximation is the ones we have found using various functions.

Here are some potential things for you to explore further.

* Try to find the values that would minimize the error under both loss functions.
* Explore using modelling for other pieces of quantitative data (e.g. BMI).
* When should you use the root mean square error as your loss function? When should mean absolute error be better as your loss function? Look into how these two might be different.
* Research different loss functions! This can potentially be very difficult as you go deep into certain topics, but also rewarding in learning advanced ways we can determine this and find good values to fit models. (here are some relatively simple alternative loss functions you might want to explore and try to implement: MSE, MSLE)
* So far, we have only explored for quantitative data. How should we go about for predicting and constructing classification models for qualitative data? You may explore a bit on Cross Entropy loss, and consider implementing this for something such as the smoker data.

This is all pretty cool, but we really want to be able to get better at predicting these values, and making use of other features such as ‘age’ to also be incorporated into insurance cost predictions. Next week we will be exploring a bit on linear regression.