**Intro to ML using Sklearn (1) Outline**

Goal 1.5 hour to 2 hours

1. Background knowledge **45 mins**
   1. What is ML?
      1. Succinct definition
         1. Supervised vs Unsupervised (Prediction vs discovery?)
         2. Classification vs Regression
            1. I think focusing on Regression might be good
      2. Why is it useful/Brief example of exciting work
   2. What is the model we will use?
      1. Linear regression (brief)
      2. Decision Trees/Random Forests
      3. Knn
   3. How do we measure success?
      1. MSE
      2. Overfitting
      3. Cross Validation
2. Intro to Package, coding how to **45 mins**
   1. Look at dataset, import package etc
   2. Declare object, train, make prediction
   3. Test on held out data
3. Practice section **30 mins**
   1. Two levels of challenges? One with cleaned data, one without?
   2. Should be harder than just copying, but not sure how
4. Recap/summary **2 mins**

Audience: People who have limited background coding (but have some basic knowledge of Python from the bootcamp), and little to no exposure in the subject.

Goal: Students hopefully will be equipped with the bare basic knowledge of what ML is, will have a decent working knowledge of one model (and maybe linear model), will know the basics for how to perform the model in Scikit-learn, and will know how to test for accuracy.

Notes: Probably should do some coding on the spot or live coding. I think that way people won’t be too distracted. Likewise, maybe some of the focus should be shifted towards the important lessons of ML, such as the checking, overfitting, small data, curse of dimensionality etc. Also some real value will come from how interesting the assignment is. If it’s more creative, the lessons will stick better. Maybe could do something like predict the next pixel in a video (this could have a worksheet to help some of the more technical parts) or the next chord in a song etc.

It's worth going through the basic commands to show what exactly each part is doing.

Perhaps I’ll show one example of classification with Knn if extra time

When our first model is up, and people have their’s coded probably should have them play with the hyperparameters to get a sense of how important it is.

Video Resources:

<https://www.youtube.com/watch?v=HC0J_SPm9co>

(pedagogy) Resources:

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006023>

<https://scottberkun.com/2013/run-a-good-workshop/>

<https://academia.stackexchange.com/questions/80777/how-to-make-programming-lectures-more-effective-and-stimulating>