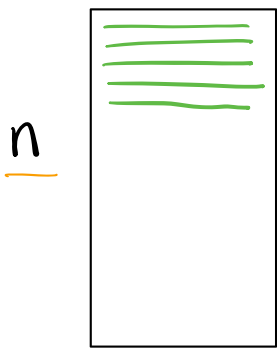


Data 311 - Lecture 17: Supervised Learning

- Classification and regression
- K nearest neighbors
- Feature engineering basics

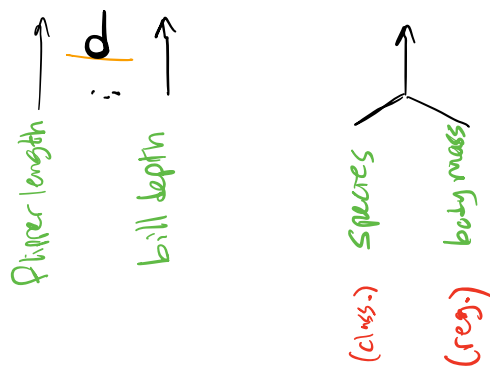
Recall: Supervised learning setup:

X , y



IF y is Continuous,
this is a regression problem.

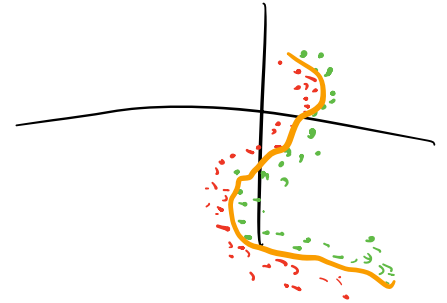
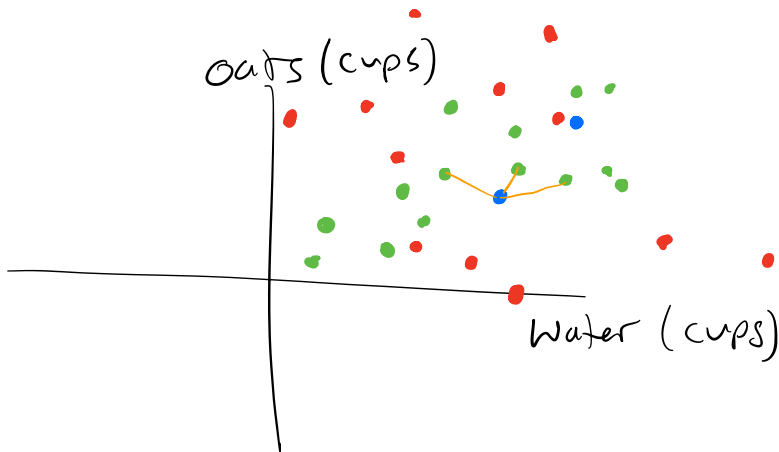
IF y is Categorical,
this is a classification problem.



Penguin
example:

K-Nearest Neighbors

Let's make oatmeal.



What k should we use?

Ex: Which is which? { lower k : less sensitive to detail
higher k : less sensitive to noise/outliers

Can KNN do regression?

Sure - take the average of the k neighbors' labels.