

# Statistical Machine Learning

Day 36

Review, part 1

# Steps in a Machine Learning Project

1. Reframe a prediction question in terms of math and statistics
2. Find, clean and transform an appropriate dataset
3. Understand the dataset (visualize, summarize)
4. Train, select and assess a prediction model
5. Report results

# This week: review of methods and examples of all 5 steps

Four prediction questions:

1. How old is someone?
2. How many pairs of shoes does someone own?
3. Does a person have heart disease?
4. Is there a stop light in front of my car? What does the signal say?

# Steps in a Machine Learning Project:

1. Translate the prediction question to math/statistics

*What is being predicted?*

*What information is available to predict from?*

*Who (or what) will the model be used to predict for?*

*Does the model need to be interpretable?*

*How should the quality of the predictions be measured?*

# Steps in a Machine Learning Project:

2. Find, clean and transform an appropriate dataset

*Find – government, your client/employer, researchers, internet...*

*Clean – are the values of the variables valid? Standard format?*

*Transform – are the values of the variable in a format suitable for putting into a model?*

# Datasets for Four Questions

1. How old is someone? **Math 407 surveys**
2. How many pairs of shoes does someone own? **Math 361 surveys**
3. Does a person have heart disease?  
<https://www.kaggle.com/ronitf/heart-disease-uci>
4. Is there a stop light in front of my car? What does the signal say?  
<http://cvrr.ucsd.edu/vivachallenge/index.php/traffic-light/traffic-light-detection/>

# Cleaning Math 361 surveys

Variable Name	Question	Acceptable values	Type of variable
height	What is your height in inches?		
favoriteNumber	What is your favorite number?		
CraterLake	How much do you enjoy visiting Crater Lake, on a scale of 1 to 9?		
commute	How do you typically commute to your classes at OIT? (Walk, car or public transit?)		
shoes	How many pairs of shoes do you own?		
drink	Do you prefer the taste of coffee or coca cola? (1 = coffee, 2 = coca cola)		
chocolate	Do you like milk chocolate better than dark chocolate? (Yes or no)		
Color	What is your favorite color?		