

k-Nearest Neighbors

What's the Difference?

k-means

- Unsupervised
- For clustering
- No x/y variables
- Based on the center points of the data

k-nearest neighbors

- Supervised
- For classification
- Need to specify an x and y
- Based on distance between data



General Steps in KNN

- Data Wrangling
- Fit the model
- Choose the best model
- Find the accuracy of that model

Data Wrangling Requirements

- Scale the data
- Train/Test/Split

Interpreting the Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP 	FP
	Negative (0)	FN	TN 

TP: True Positive
TN: True Negative

FP: False Positive
FN: False Negative

Eeek, what matters here?

- Look at the accuracy on the bottom

	precision	recall	f1-score	support
Female	0.64	0.91	0.75	35
Male	0.70	0.28	0.40	25
accuracy			0.65	60
macro avg	0.67	0.60	0.58	60
weighted avg	0.67	0.65	0.61	60