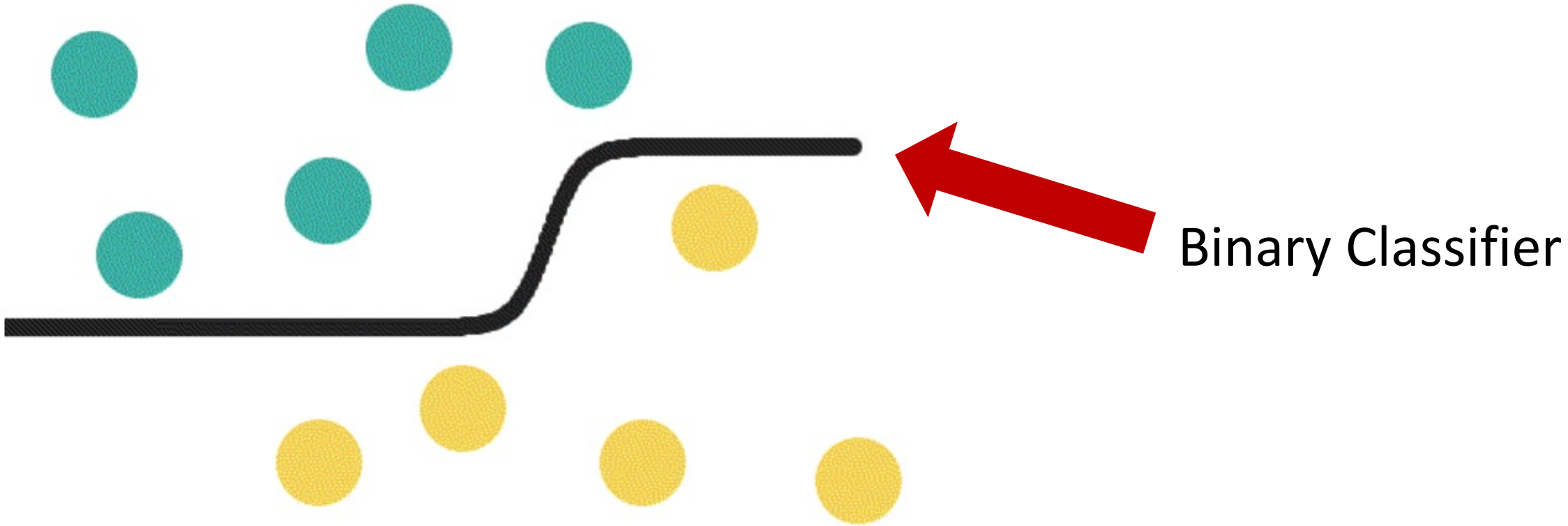
Applied Machine Learning Classification: Introduction

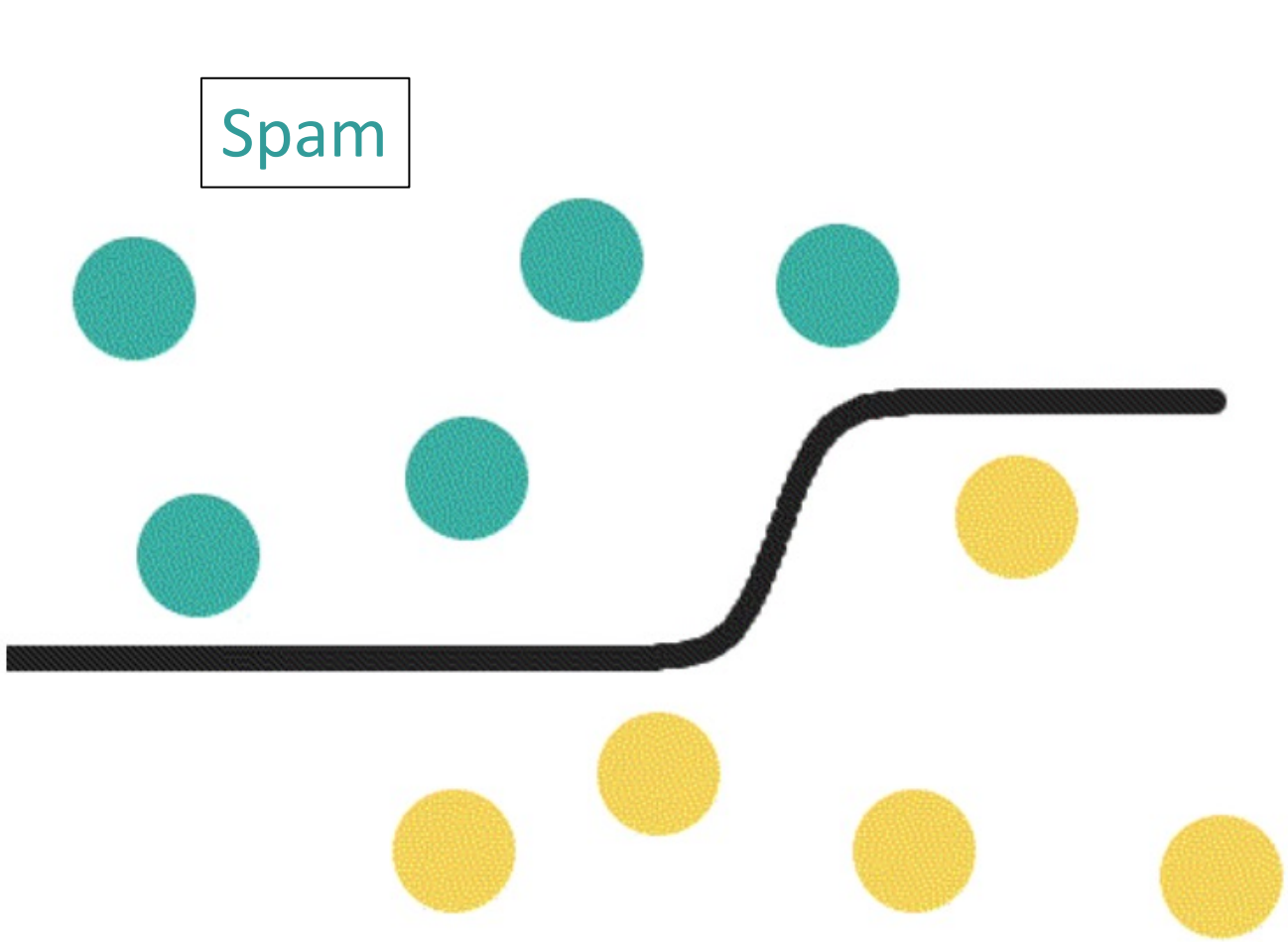
Ngan Le

thile@uark.edu

Classification



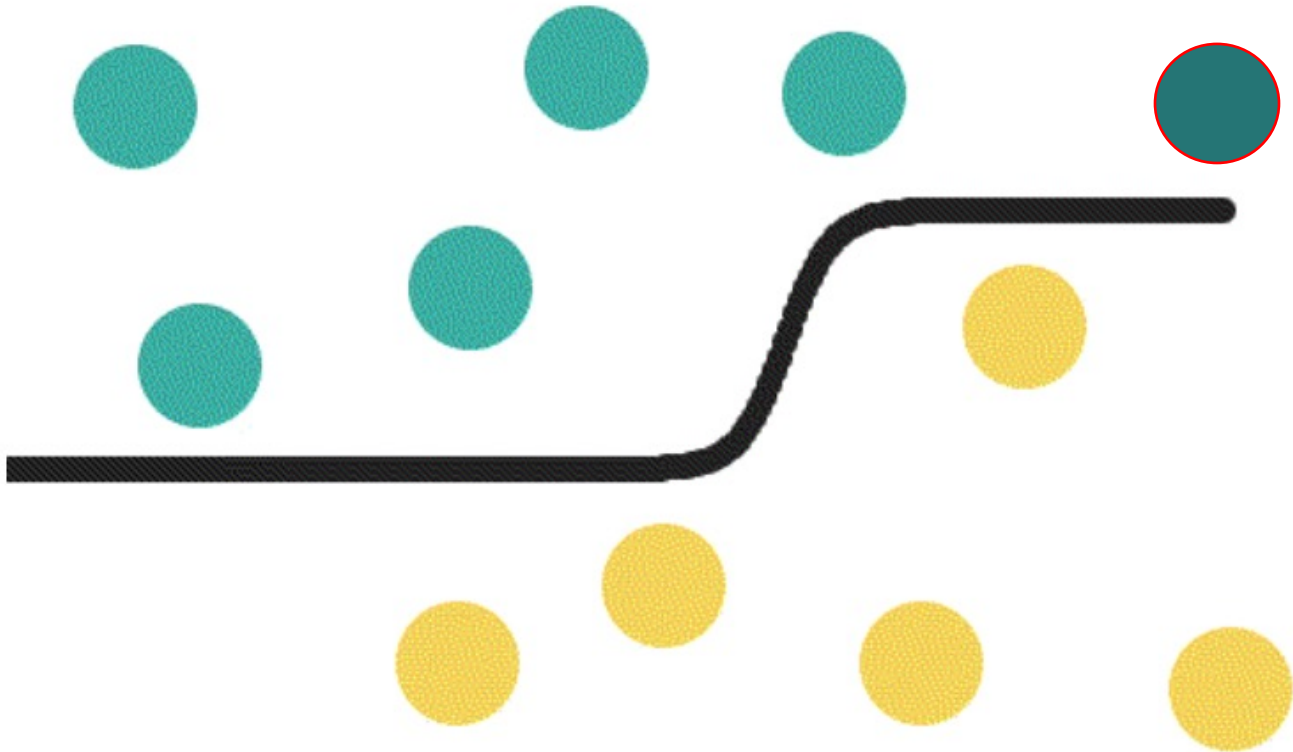
Classification



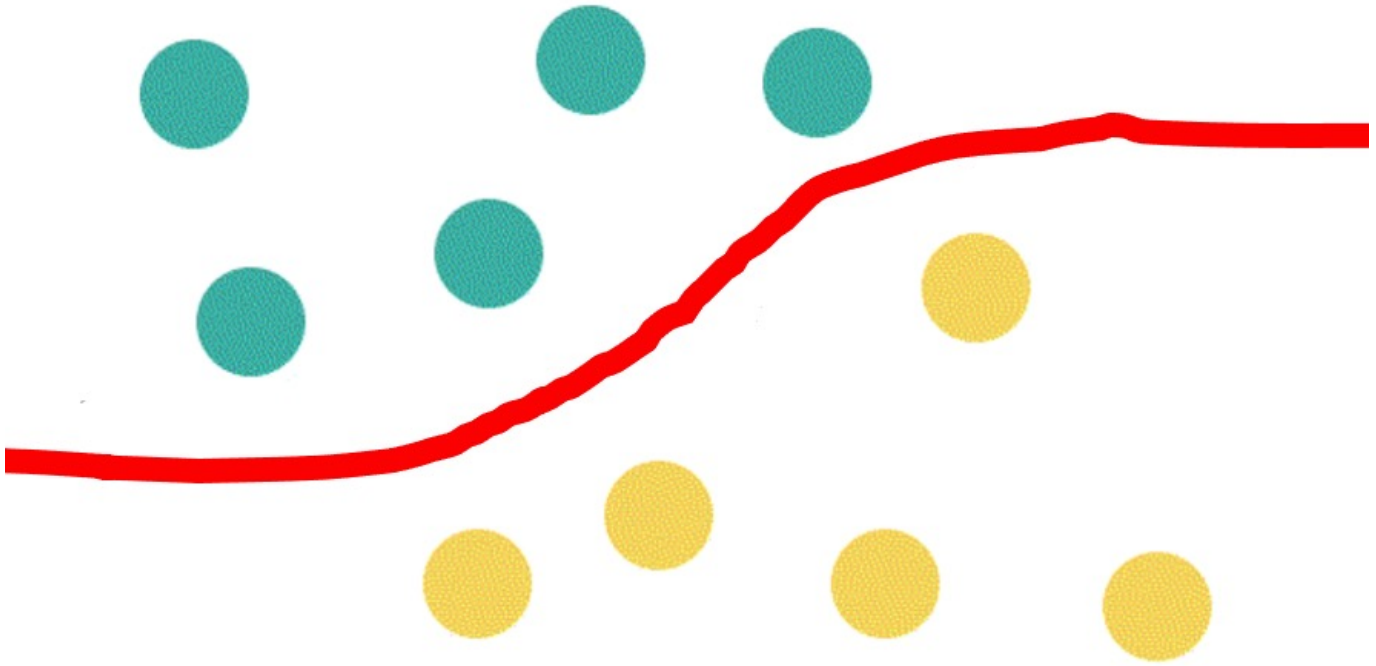
Is this spam?

Classification

YES/NO



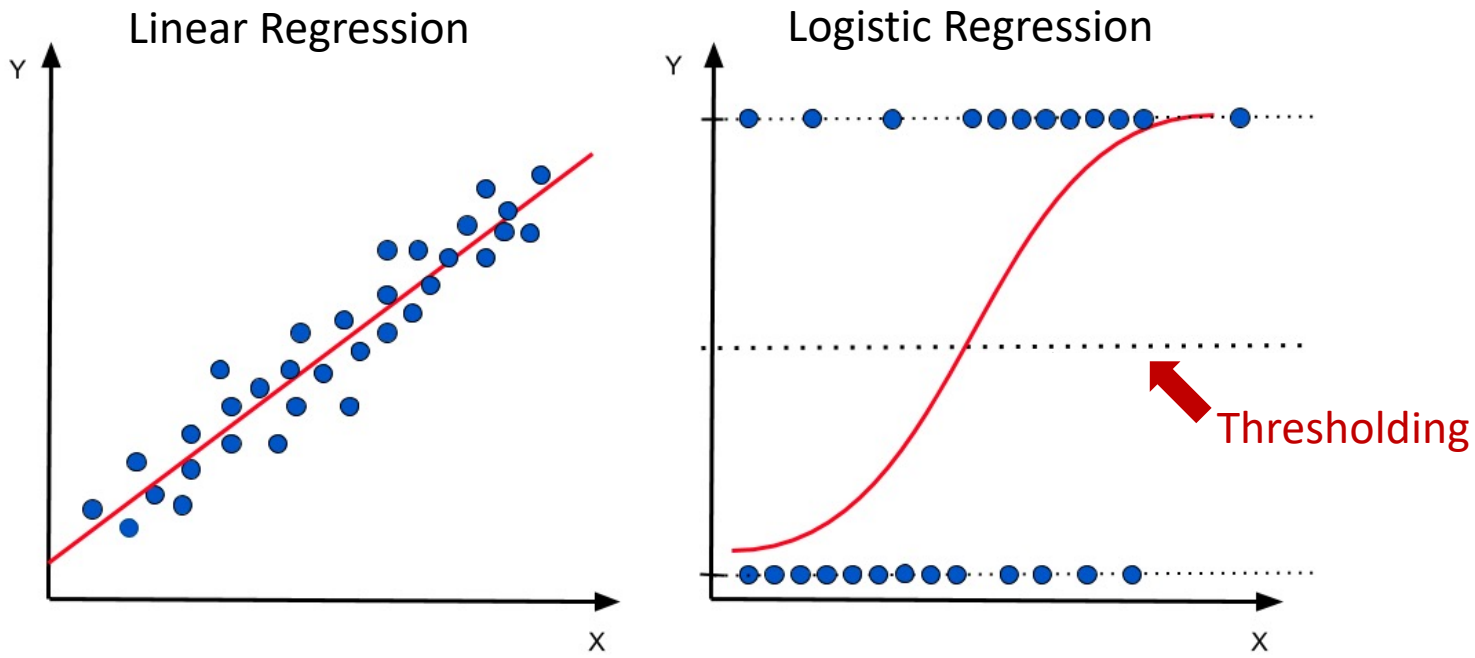
Classification



- Logistic regression
- Nearest Neighbors
- Decision trees
- Random forests
- SVM
- Naive Bayes
- Deep Learning (Neural Network)

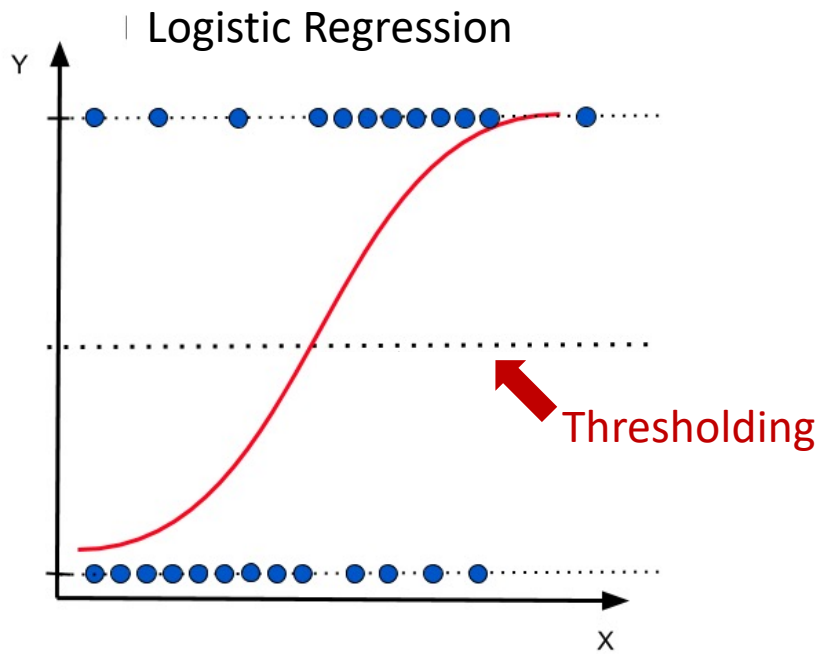
Classification: Logistic Regression

A variation of linear regression that performs a regression, then uses some threshold to make a classification decision



Classification: Logistic Regression

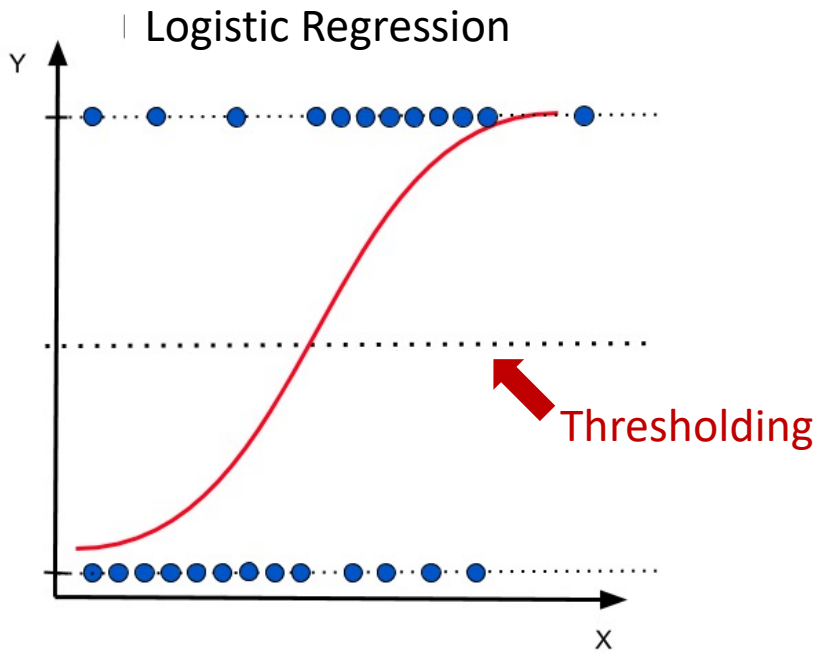
A logistic regression finds a **logistic function** and uses it to divide two classes of data.
A logistic function is a function in the shape depicted in this slide. It can range in values from zero to one.



Classification: Logistic Regression

This model is relatively easy to interpret and train

But, it doesn't perform well unless an actual logistic boundary can be found.
To create a logistically divisible set of classes, you sometimes have to perform some very heavy data manipulation.



The Lab

differentiate between oranges and grapefruit

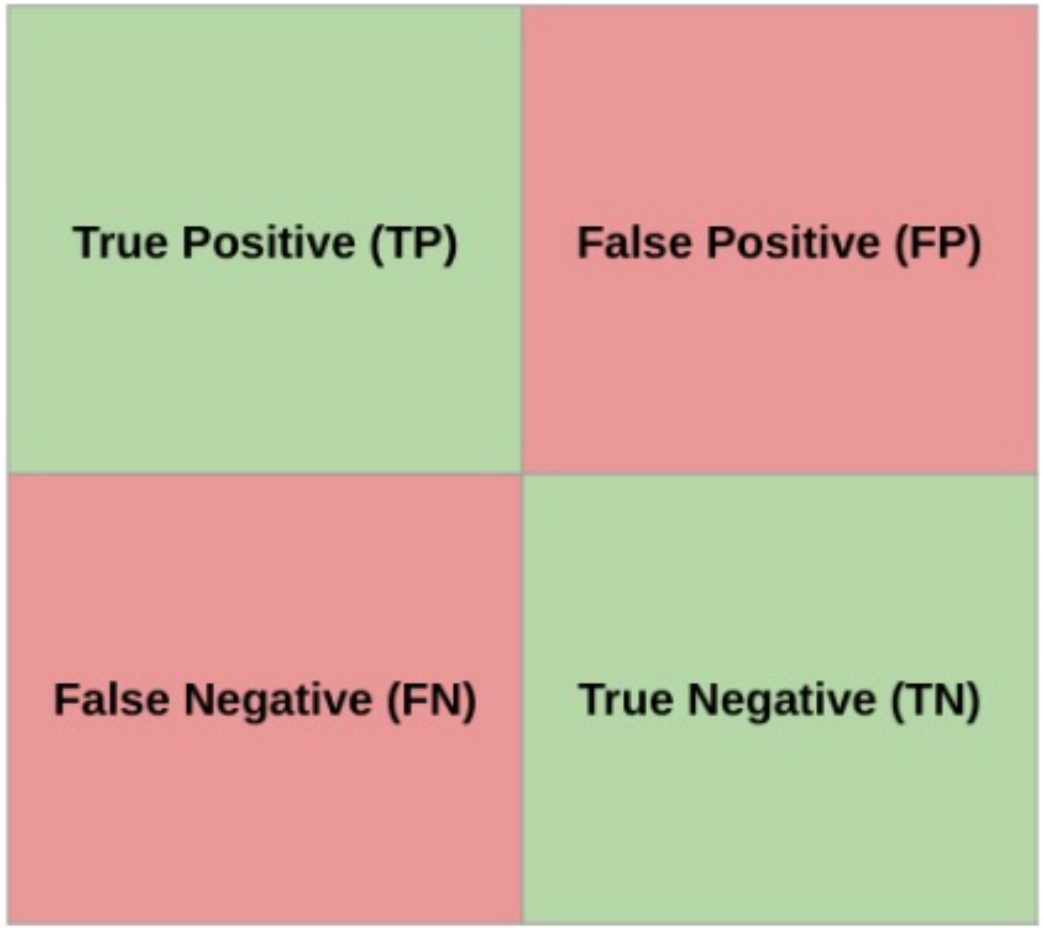
Input: a dataset with weight, size, and color information
Your task: build a logistic regression model

The Lab: Grid Search

- Test different hyperparameters for a model: tries every combination of parameters and finds which performs the best
- The search accepts a dictionary
  - keys that map to model parameters.
  - values are lists for hyperparameters that you want to experiment with or single values for parameters that you want to keep constant.

```
search = GridSearchCV(model, {
  'learning_rate': [1e-3, 1e-4],
  'max_iter': [10000, 15000],
  'C': 1,
})
```

The Lab: Confusion Matrix



Your Turn!