Models for Regression vs. Models for Classification

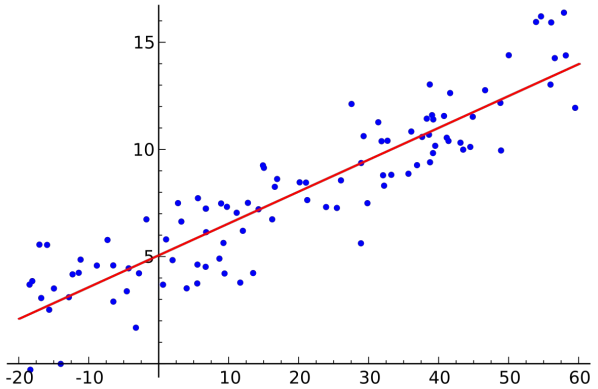
Presented by Ted Haley

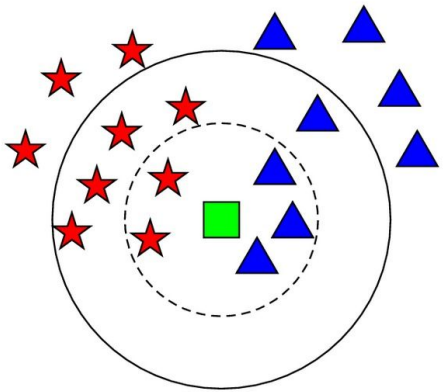
Brief Introduction

- Ted Haley
- Born in Toronto, Ontario
- Bachelor's of Civil Engineering (UBC)
- Master's of Data Science (UBC)

Agenda

- Brief overview of Regression
- Models for Regression
  - Example: Regression Model
- Brief overview of Classification
- Models for Classification
  - Example: Classification Model
- Summary
- Activity

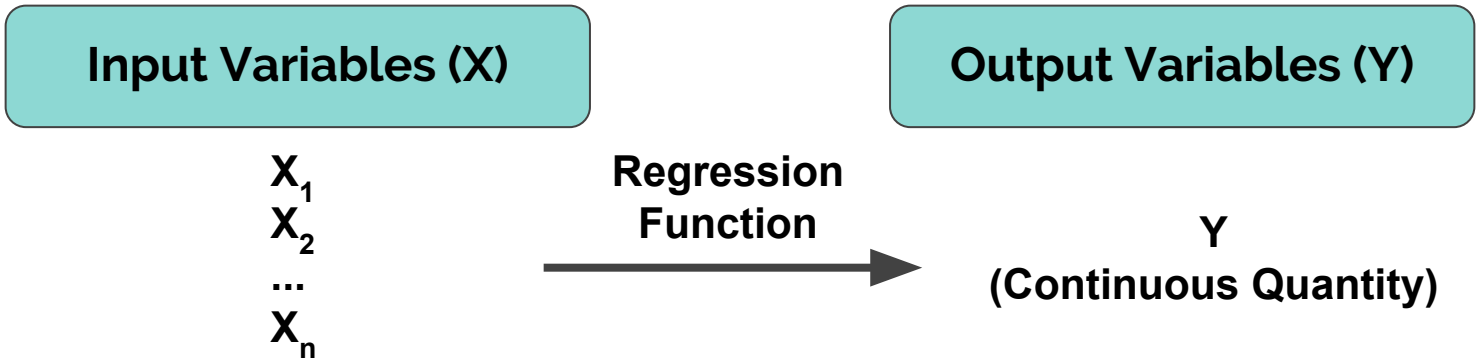




Regression

Regression is the statistical process for analyzing the relationship between a dependent (output) variable and one or more independent (input) variables.

Regression is about ***predicting a quantity.***



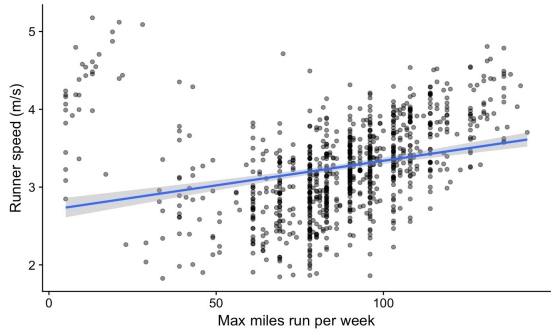
Models for Regression

Many different flavours for regression:

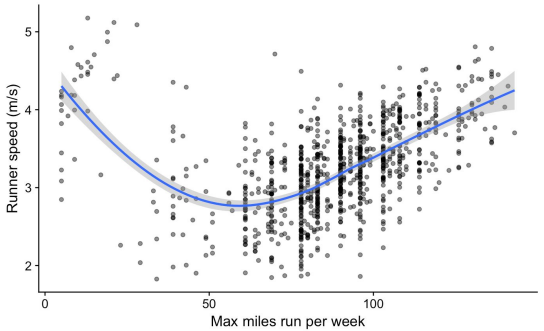
- **Simple Linear Regression:** 1 independent var
- Multiple Linear Regression: 2+ independent var
- Polynomial Regression: Polynomial expression
- Etc.

Optimization Methods:

- **Least Squares Estimation**
- Ridge Regression Estimation
- Bayesian Estimation
- Etc.

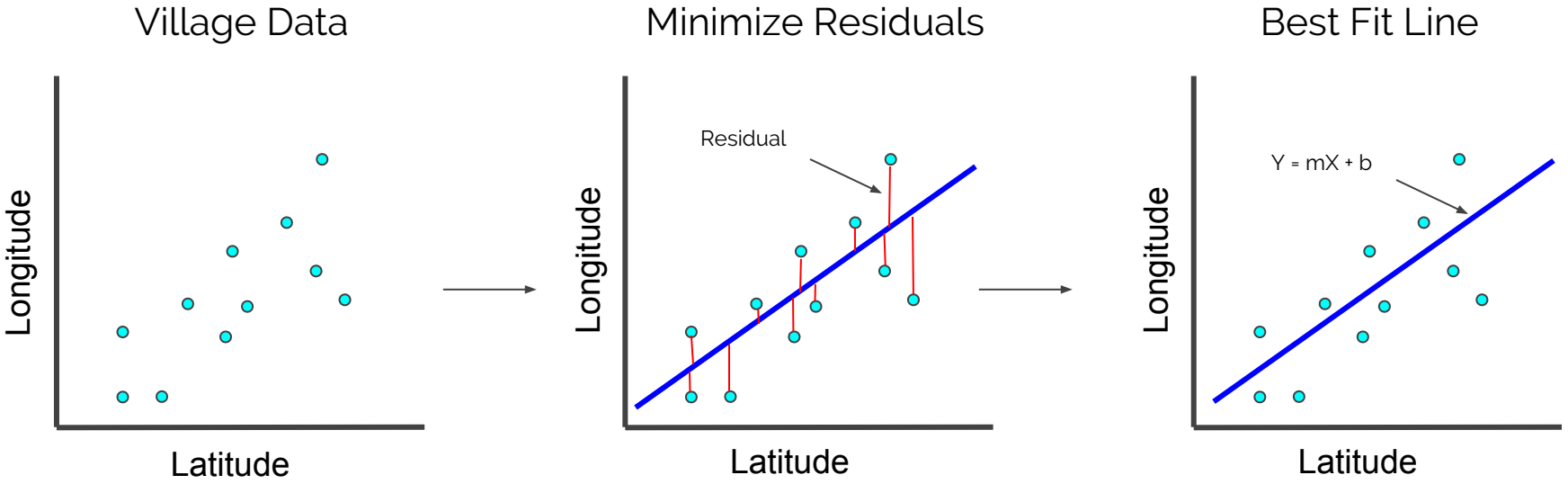


Linear Regression



Polynomial Regression

Example: Simple Linear Regression

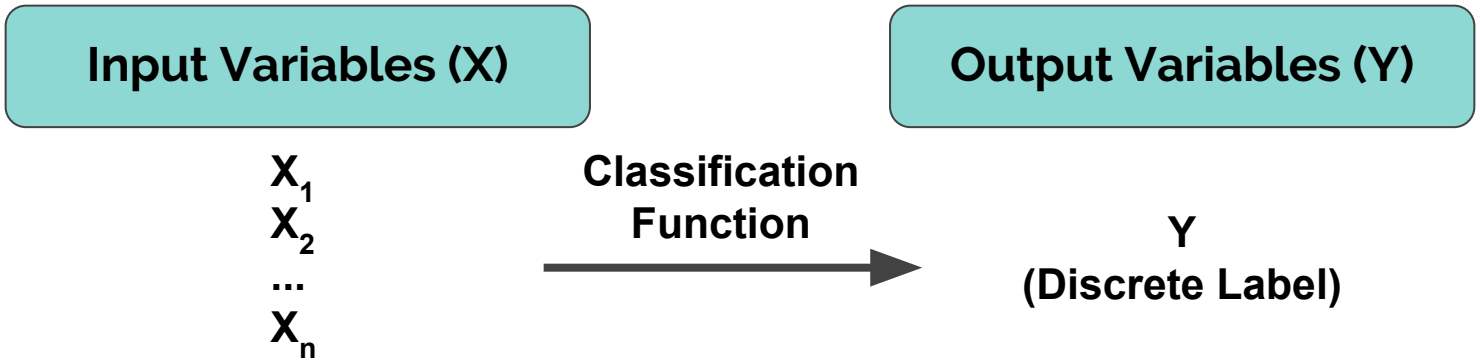


Optimize by reducing the sum of squares of the residuals.

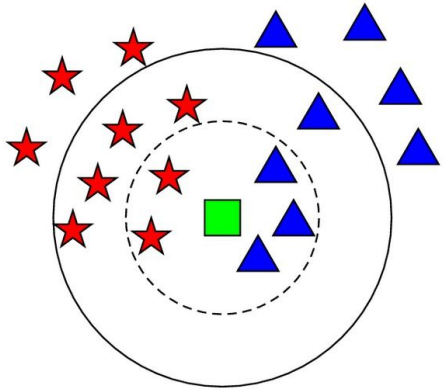
Classification

Classification is the process of identifying which category a new uncategorized observation belongs to, given the observation characteristics.

Classification is about ***predicting a label*** from a finite set of options.



Models for Classification

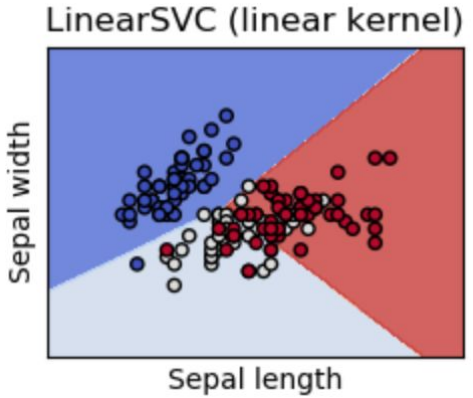


Many different flavours for classification:

- **k-Nearest Neighbors**
- Support Vector Classification (SVC)
- Stochastic Gradient Descent (SGD)
- Etc.

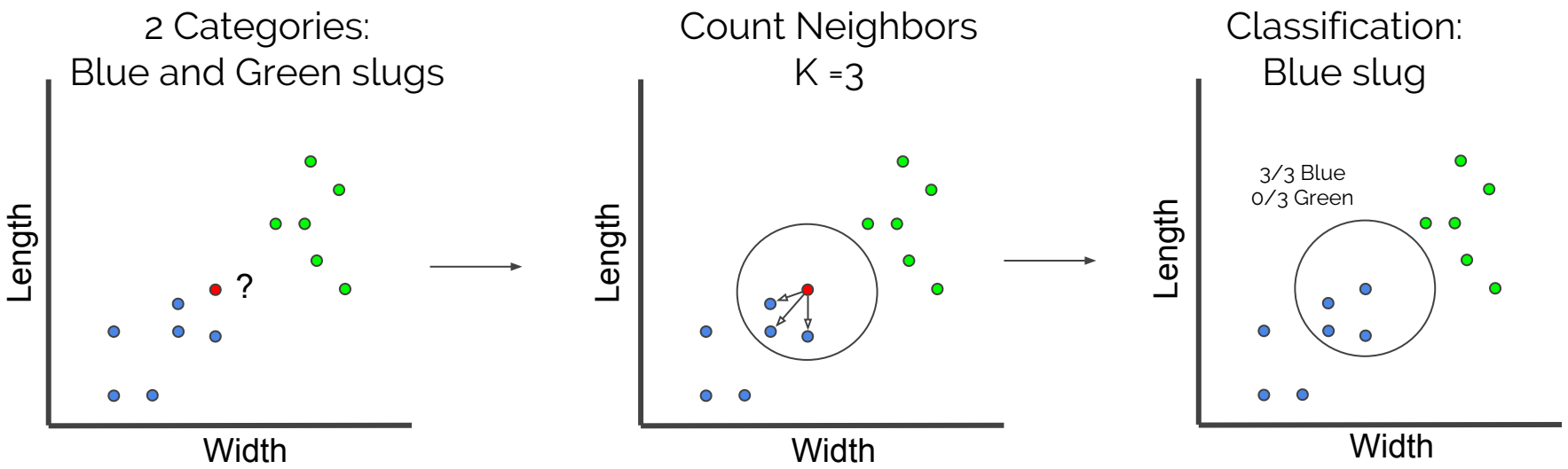
Model Optimization:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$



Example: k-Nearest Neighbors

K = Number of closest neighbors to consider



Summary: Regression vs. Classification

	Regression	**Classification**
Prediction	Continuous quantity	Discrete label
Model Optimization	Root Mean Square Error (RMSE)	Percentage classified correctly (Accuracy)
Type of Learning	Supervised	

Activity

Refer to the provided Jupyter Notebook for coding examples and questions for Simple Linear Regression and K-Nearest Neighbors.

Additional Readings:

SKLearn Nearest Neighbors
http://scikit-learn.org/stable/modules/neighbors.html#classification

SKLearn Ordinary Least Squares
http://scikit-learn.org/stable/modules/linear_model.html#ordinary-least-squares

Regression vs. Classification
https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/