INTRO TO REGRESSION ANALYSIS

UW
DATA SCIENCE
CLUB.

Presented by Jack Douglas

Workshop Overview

UW
DSC.

Intended Audience: Beginner Data Scientists

Goals:

- Learn about the importance of regression and fundamental machine learning concepts that regression uses
- Learn the theory behind three main types of regression
- Learn how to implement regression models in Python on your own

Workshop Outline

UW DSC.

1. Background Information
    a. Traditional Programming vs. Machine Learning
    b. Machine Learning Basics
2. What is Regression Analysis?
3. Cost Functions
4. Optimizer
5. Three Types of Regression
    a. Linear Regression
    b. Polynomial Regression
    c. Logistic Regression
6. Applications of Regression in Google Colab

UW DSC.

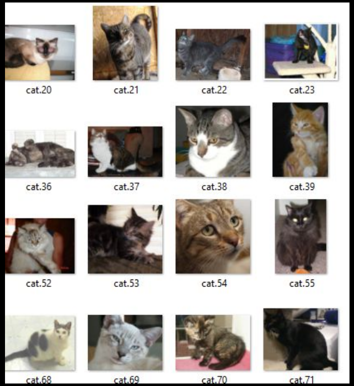
Classifying Dogs vs. Cats Example

- Suppose you wanted to create a program which classifies photos of cats and dogs

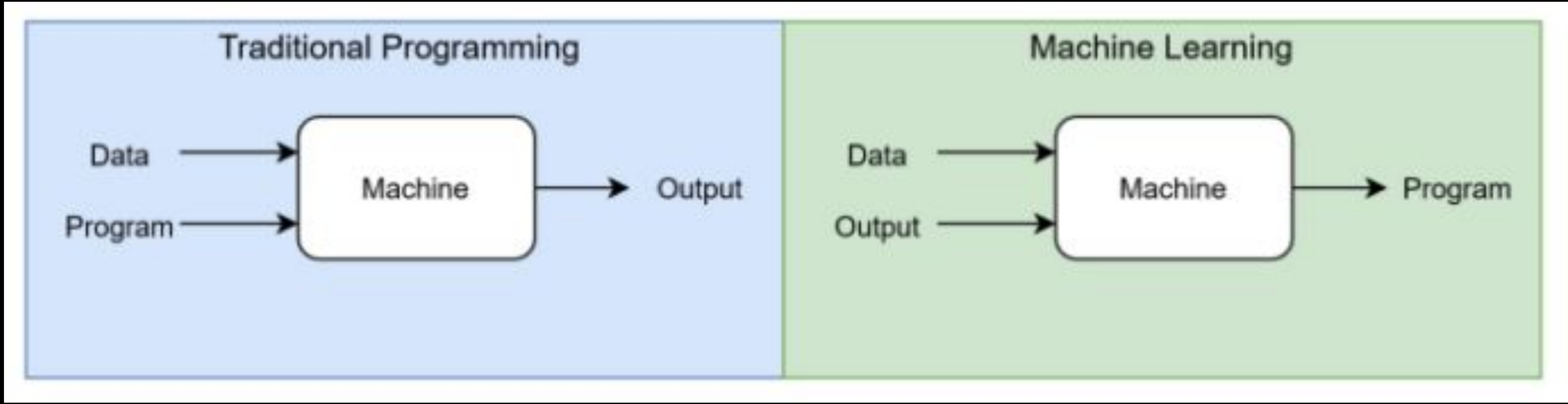
- The images are the input (*x*), the program is the function (*f*), and the label is the output (*y*)
    - Ie. f(x) = y

- Traditionally, we would have to create the function which is very challenging given the amount of variation in input

- Machine learning proposes using the many inputs and outputs to determine the function



Traditional Programming vs. Machine Learning



UW
DSC.
–

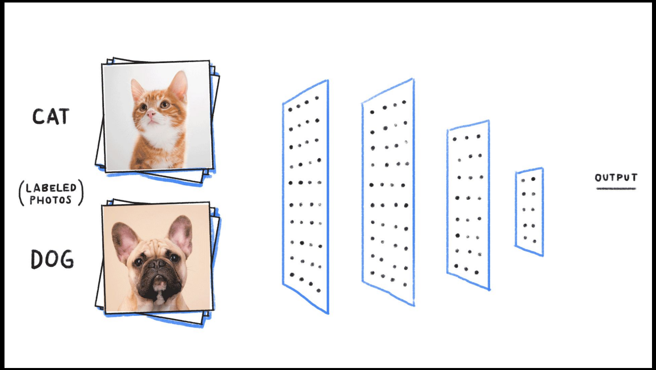
Machine Learning Basics

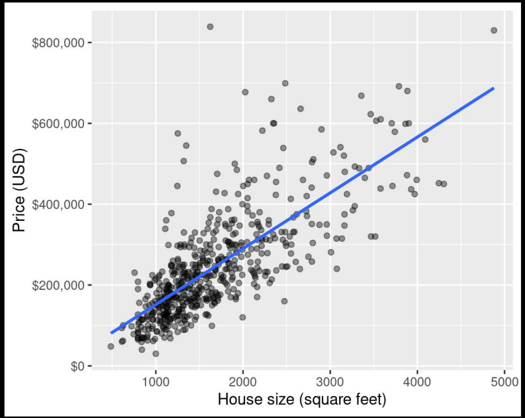
- Def'n: Machine learning refers to the field of study that gives computers the ability to learn without being explicitly programmed

- There are 3 main types of machine learning
  - Supervised Learning: Finds a correlation between given inputs and outputs (labels)
  - Unsupervised Learning: Finds how to structure unlabelled inputs
  - Reinforcement Learning: Performs a task and improves by maximizing a reward

Types of Supervised Learning Problems

- There are 2 main types of supervised learning problems
    - Classification: Predicting a label (discrete)
        - Ex. Distinguish between a cat and dog, given a labelled dataset with photos of both
    - Regression: Predicting a quantity (continuous)
        - Ex. Predict the price of house, given a labelled dataset of housing prices along with other factors (lot area, year built, etc.)





UW DSC.

What is Regression Analysis?

UW DSC.

- Def’n: Regression analysis refers to a set of statistical methods used for estimating the relationship between a dependent (target) variable and independent (predictor) variable(s)

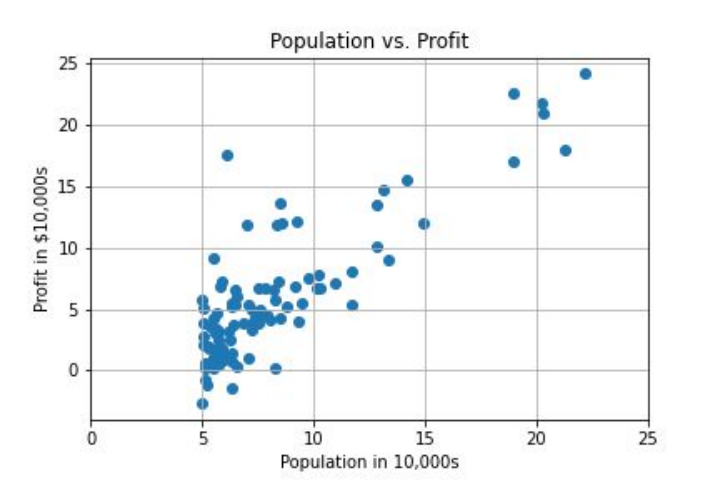
- Motivation
  - Provides a powerful statistical method to examine the relationship between two or more variables and make predictions about future data
  - A data analysis tool which uses several ubiquitous machine learning concepts, namely cost functions and optimizers

Restaurant Owner Example

- Suppose you own a restaurant chain and you have data about the profit and population of locations in various cities

- You want to use this data to help decide which city to open a new location

- You have to come up with a model for this data so that you can predict the profit for a given city population...



Notation

Population in 10,000s (x)	Price ($) in 10,000s (y)
6.1101	17.592
5.5277	9.1302
8.5186	13.6620
…	…

m = number of samples
x = input variable (a.k.a. feature)
y = output variable (a.k.a. target variable)

$$x^{(1)} = 6.1101$$

$$y^{(3)} = 13.6620$$

UW DSC.

Linear Regression

- Def'n: Linear regression refers to a form of regression analysis where a linear model estimates the relationship between a dependent variable and 1 or more independent variables
  - "Simple" linear regression has 1 independent variable (a.k.a. line of best fit)
  - "Multiple" linear regression has more than 1 independent variables

- Regression can be broken down into 3 components
  1. Start off with a model, h(x), that has arbitrary initialized parameters
  2. Use a cost function to measure the error in our hypothesis
  3. Use an optimizer to minimize the error and adjust our hypothesis

UW DSC.

Model

- Our model, h(x), is made up of two components: features and parameters
  - X represents the features and θ represents the parameters

- The model maps inputs to outputs

- We are given the features and we are trying to determine the parameters of our model
  - X_0 is called the **bias** and it must be initialized to 1 for regression analysis

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

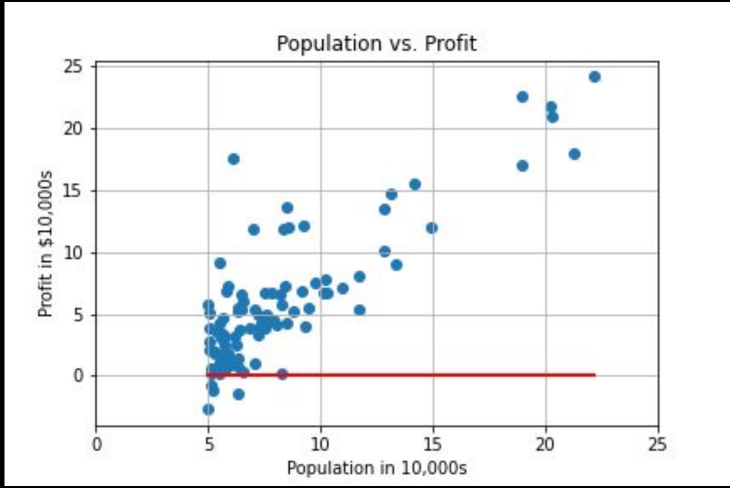
Parameter vector

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Feature Vector

Simple Linear Regression Model

UW DSC.



θ_0 and θ_1 are initialized to 0

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

Feature vector

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

Parameter vector

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 = \theta_0 + \theta_1 x_1$$

Simple Linear Regression Model

Multiple Linear Regression Model

UW
DSC.

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

Parameter vector

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

Multiple Linear
Regression Model

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Feature Vector

Implementation Detail (Linear Algebra)

UW
DSC.
–

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 \cdots + \theta_n x_n = \theta^T x$$

Implemented as the inner product of
the transpose of the parameter vector
with the feature vector

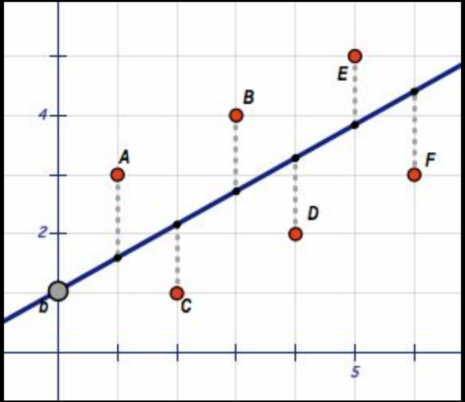
Cost Functions

- Def'n: A cost function refers to functions which measure the error in the predictions of a model compared to the actual results
  - A.k.a loss function, error function
  - Often represented with either the symbols J or L

- Examples of cost functions
  - Mean Square Error (MSE)
  - Log Loss / Binary Cross-Entropy
  - Focal Loss
  - Categorical Cross-Entropy

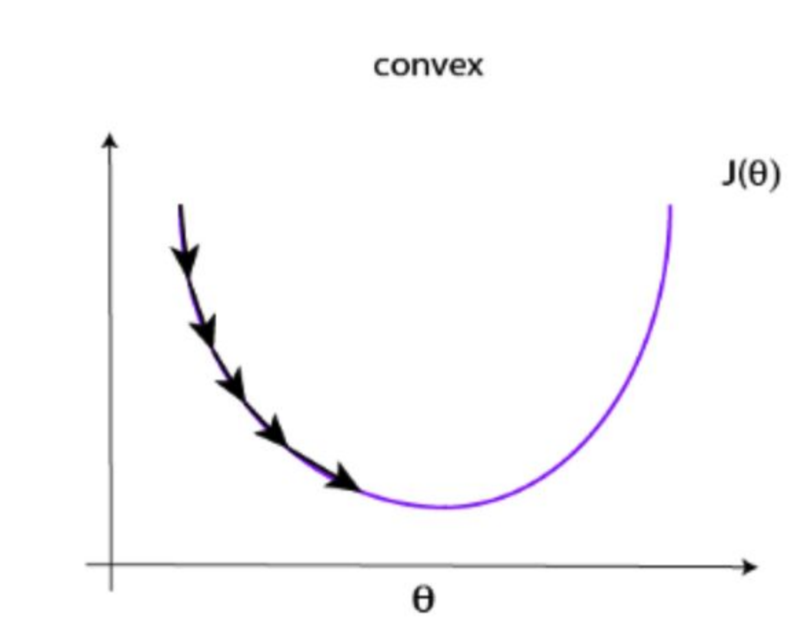
Mean Squared Error (MSE)

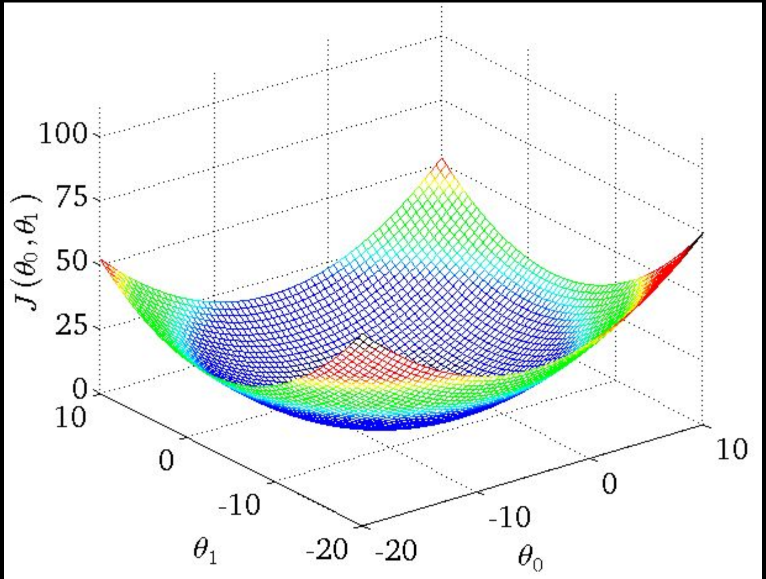
$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$



Cost Function Visualization

UW DSC.





Optimizers

UW
DSC.

- Def’n: Optimizers refers to algorithms which minimize the cost function of a model

- Examples
  - Gradient Descent
  - Stochastic Gradient Descent (SGD)
  - Batch Gradient Descent
  - Adam
  - Adaptive Gradient Descent (AdaGrad)

General Gradient Descent Algorithm

UW DSC.

Repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \qquad \text{for all } j \text{ from } 0 \text{ to } n$$

}

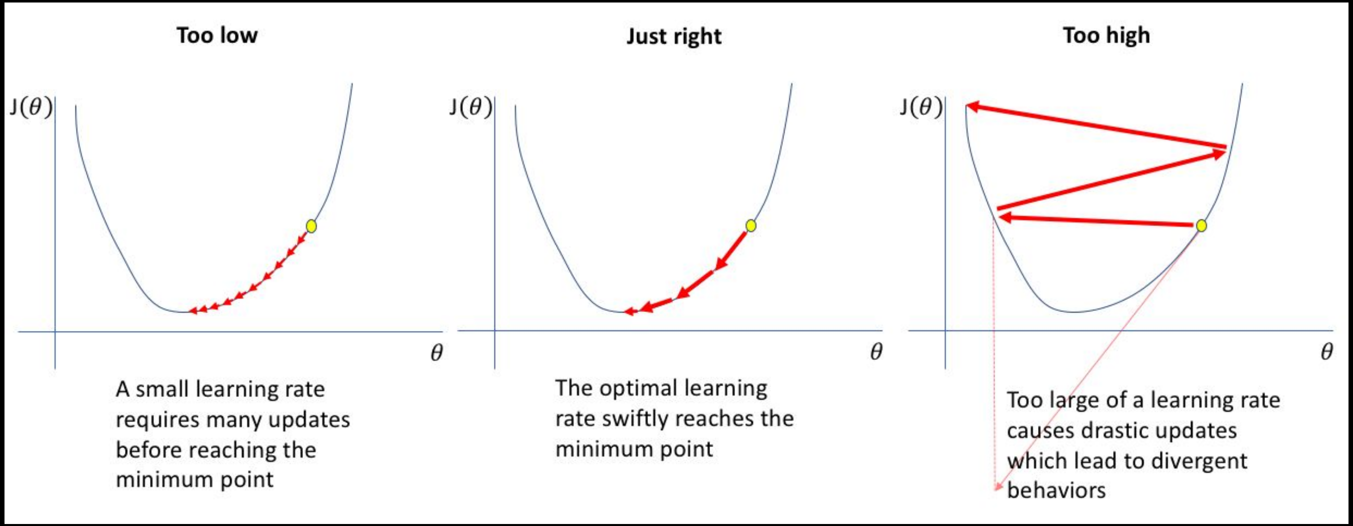
$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$\boldsymbol{\alpha}$ = learning rate

$\frac{\partial}{\partial \theta_j}$ = partial derivative w/ respect to θ_j

UW DSC.

Learning Rate



Repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \qquad \text{for all } j \text{ from } 0 \text{ to } n$$

}

Gradient Descent for Linear Regression

UW
DSC.
–

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left((\theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n) - y^{(i)}\right)^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j$$

Partial Derivative Derivation

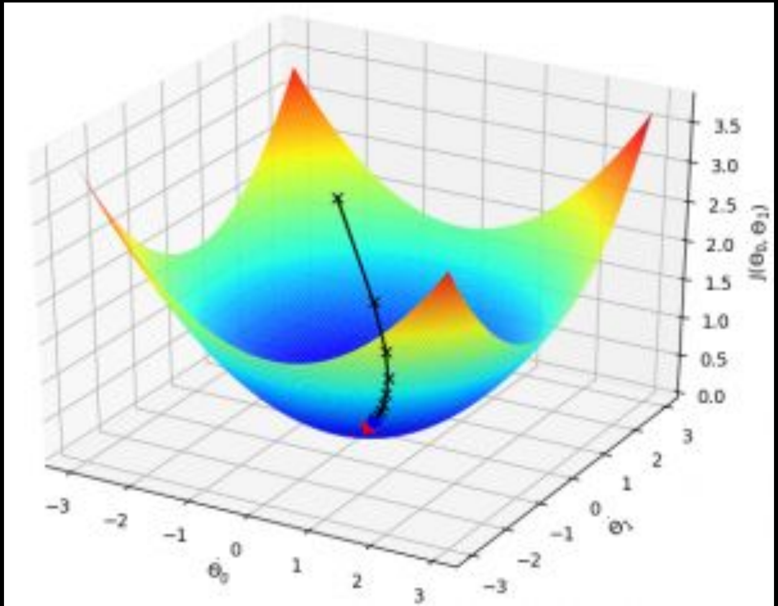
Repeat until convergence (for all *j* from *0* to *n)* {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j$$

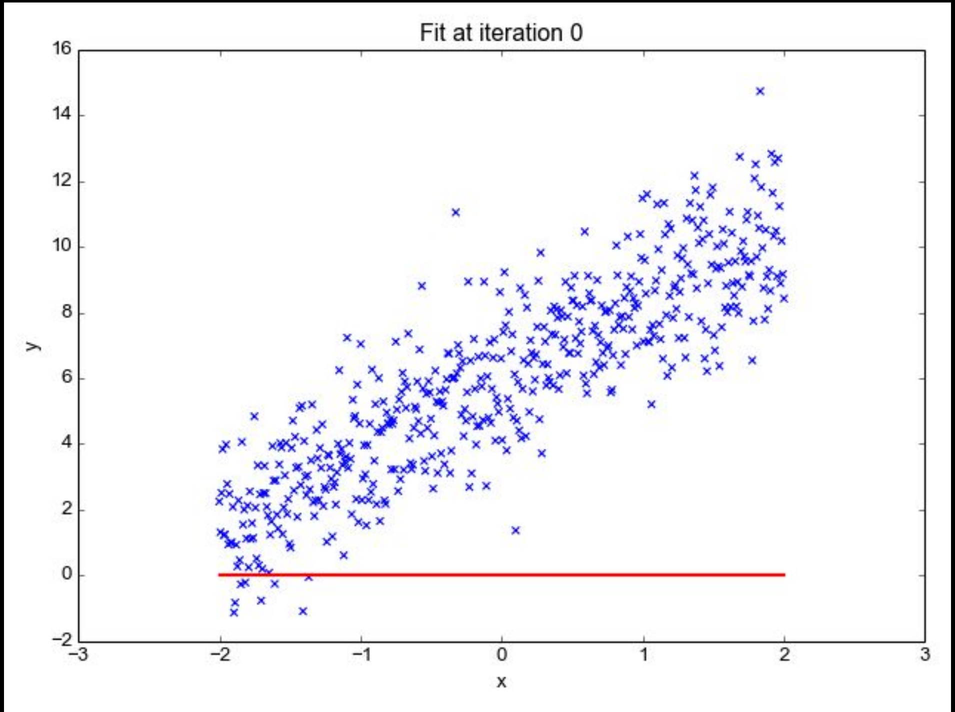
}

Gradient Descent for Linear Regression

Gradient Descent Visualization



Linear Regression Visualization

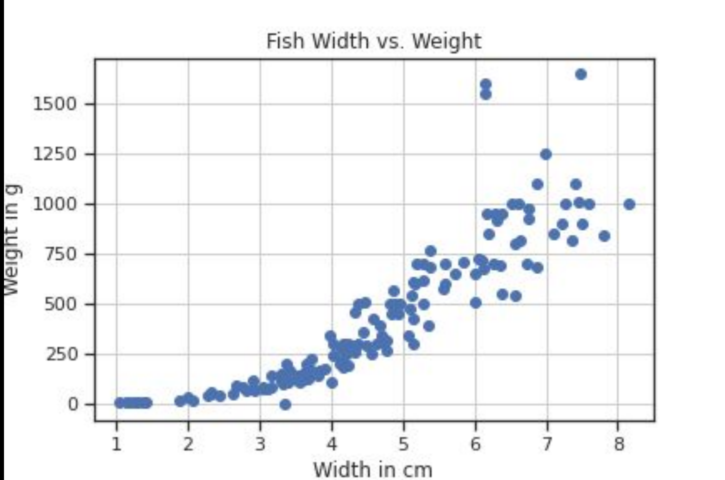


Fish Example

- Suppose you work for a fishing company and you're given a dataset of fish width and weight

- Your boss wants you to model this data to help quickly estimate the weight of a fish

- You have to come up with a model for this data but linear regression doesn't look suitable...



Polynomial Regression

UW DSC.

- Def'n: Polynomial regression refers to a form of regression analysis where the data is modelled by a nth degree polynomial

- Polynomial regression is an extension of linear regression
  - Recall we are given the values of **x** and **y**, and we are trying to determine the parameter vector which has degree 1

- Unlike with linear regression, the number of features and the number of parameters don't necessarily need to match

Polynomial Regression Model

UW DSC.

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 {x_1}^2 \cdots + \theta_n {x_1}^n$$

Polynomial Regression Model for Single Variable

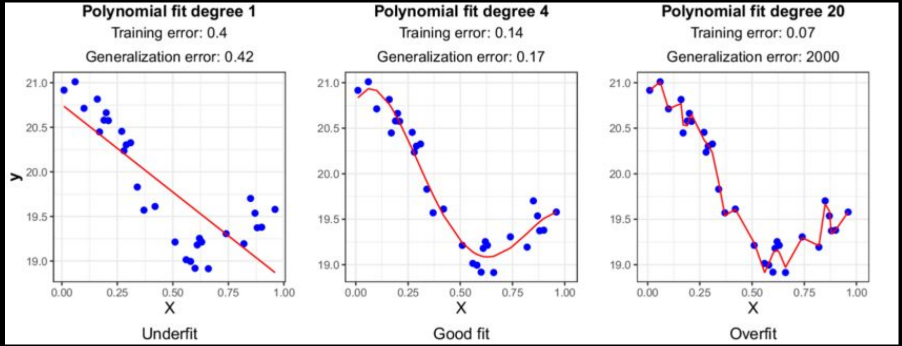
$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 {x_2}^2 \cdots + \theta_n {x_n}^n$$

Polynomial Regression Model for Multiple Variables

Underfitting vs. Overfitting

- **Underfitting** refers to models which are too simple to accurately capture the relationship between the features and the target variable

- **Overfitting** refers to models which correspond to closely to the training data such that it generalizes poorly to unseen data

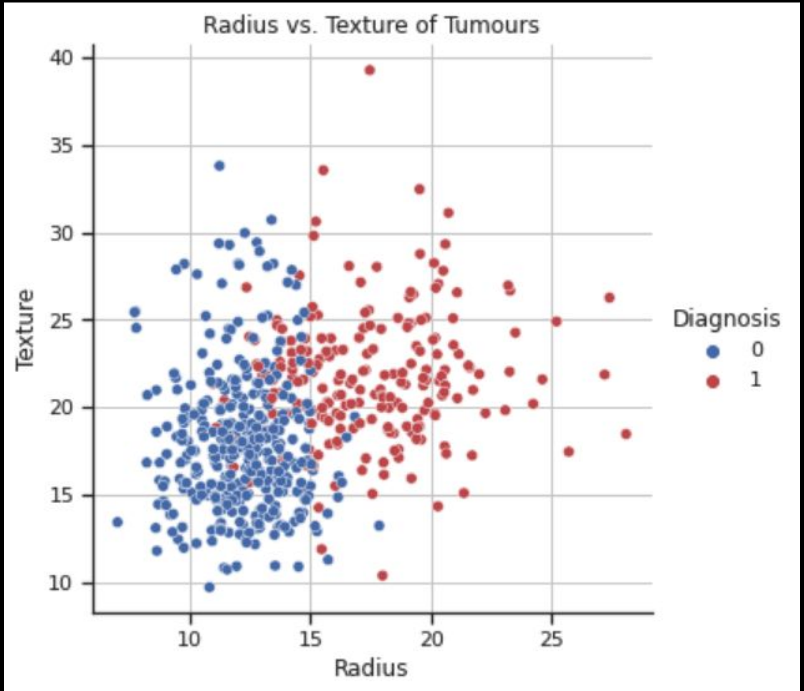


Tumour Classification Example

- Suppose you are a researcher at a hospital and you are trying to find alternative ways of diagnosing tumours

- You have access to a dataset that classifies tumours as malignant or benign and it has information about the tumour radius and texture

- You have to come up with a way for classifying tumours given their radius and texture…



Logistic Regression

UW DSC.

- Def’n: Logistic regression refers to a form of regression analysis where a model is used to predict the probability of a class
    - Typically used as a classification algorithm!

- Why is it called regression?
    - It uses the same underlying techniques as other forms of regression analysis
    - It is another generalized linear model but it is regressing to a probability (continuous value) of a categorical outcome

Logistic Regression Model

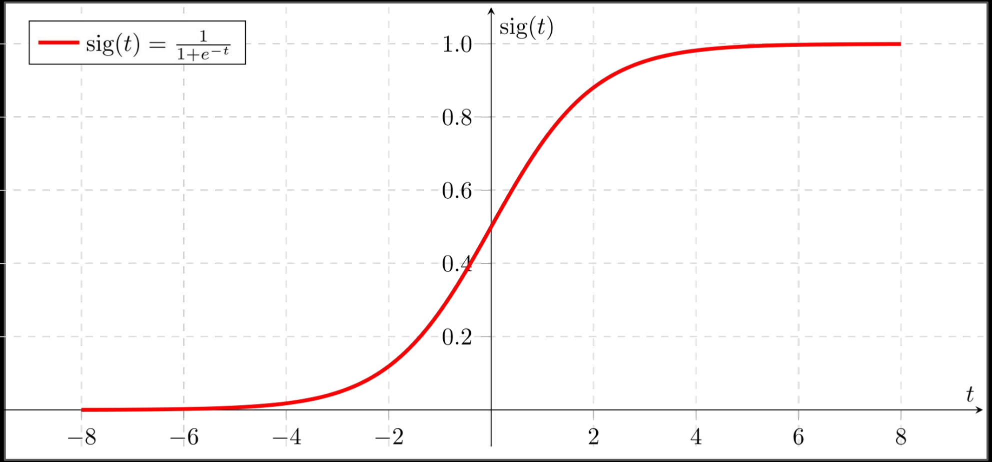
- For binary classification problems, we say $y = 0$ is the negative class and $y = 1$ is the positive class

- With our previous models $h(x)$ can be greater than 1 and less than 0
  - In contrast, logistic regression is bounded by 0 and 1

- To bound our model predictions, the logistic regression model uses the logistic function which is a type of sigmoid

Logistic Function (Sigmoid)

UW DSC. –

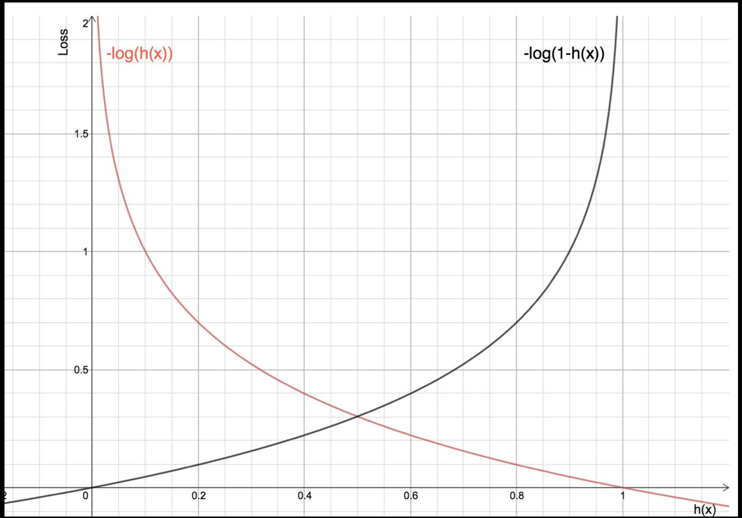


$$h_\theta(x) = sig(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Logistic Function Formula

Log Loss / Binary Cross-Entropy

$$J(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & if\ y = 1 \\ -\log(1 - h_\theta(x)) & if\ y = 0 \end{cases}$$



Gradient Descent for Logistic Regression

UW DSC.

$$h_\theta(x) = sig(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Logistic Function

$$J(h_\theta(x), y) = \begin{cases} -\log\big(h_\theta(x)\big) & if\ y = 1 \\ -\log\big(1 - h_\theta(x)\big) & if\ y = 0 \end{cases}$$

Log Loss / Binary Cross-Entropy

Repeat until convergence (for all *j* from *0* to *n)* {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \big(h_\theta\big(x^{(i)}\big) - y^{(i)}\big)\, x_j$$

}

Same gradient descent formula!

Evaluating Models

UW
DSC.

- Regression Problem Metrics
    - Mean Squared Error (MSE)
    - Root Mean Squared Error (RMSE)
    - Mean Absolute Error (MAE)
    - R-Squared (R^2)

- Classification Problem Metrics
    - Accuracy
    - Precision
    - Recall

Regression Analysis Notebook

Notebook: https://bit.ly/39kUSRm

Next Steps

UW DSC.

- Perform an exploratory data analysis (EDA) on your own!
- Check out other forms of regression analysis
  - Ex. Binary Regression, Poisson Regression, etc.
- Learn about feature normalization
  - Useful scaling technique to improve regression results
- Submit raffle and feedback form: https://bit.ly/2XUvrDJ!

Resources/Bibliography

- Machine Learning by Andrew Ng: https://www.coursera.org/learn/machine-learning/home/welcome

- Logistic Regression for Malignancy Prediction in Cancer by Luca Zammataro : https://towardsdatascience.com/logistic-regression-for-malignancy-prediction-in-cancer-27b1a1960184

- 3 Best Metrics to Evaluate Regression Models by Songhao Wu: https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b