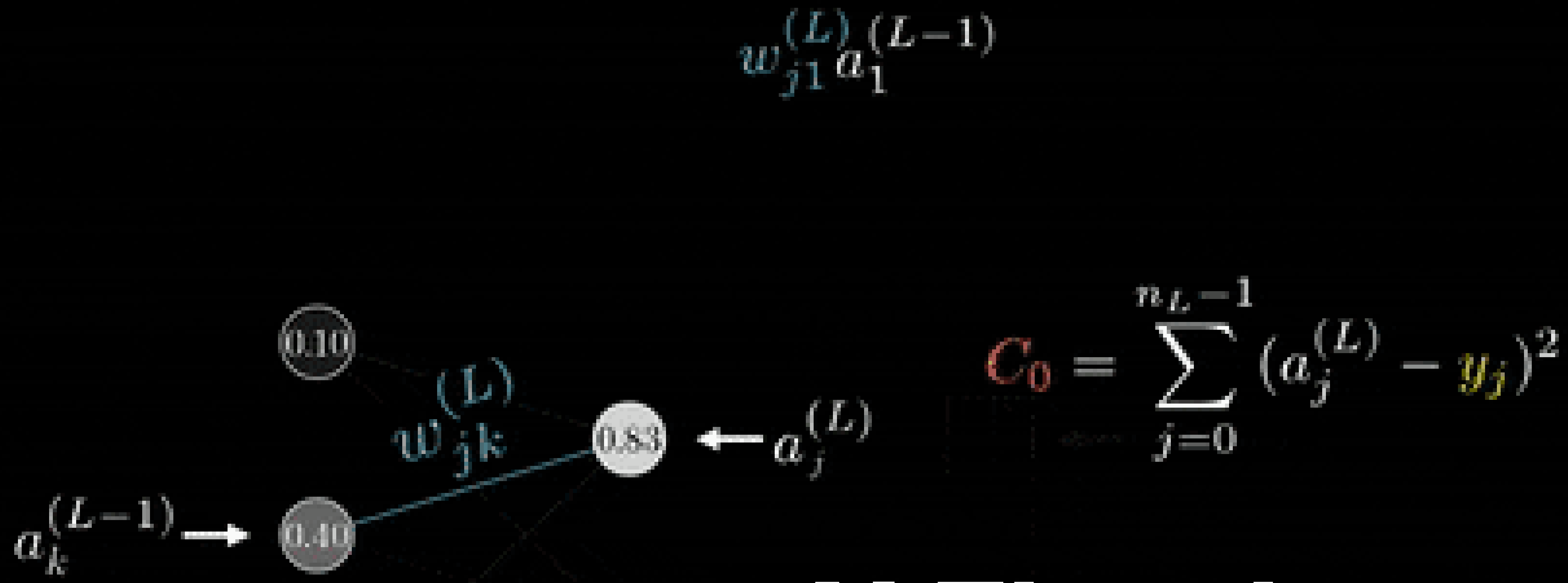


AI Theories

Machine Learning Concepts

Module 09 Algorithm Description

IDARE®

Types of Algorithms utilized in AI

Statistical Algorithm

- Designed understands relationship between variables
- Learn From Past and Predict Future
- Serves as the basis of explainability of the solution
- Useful when the problem is well understood

Linear Regression, logistic Reg. , SVR etc.

Machine Learning Algorithm

- Designed to provide accurate prediction
- Learn From Past's Mistake and Predict Future
- Useful when the problem is less understood

Decision Tree, Neural Network, Deep Learning, Gradient Boosting etc

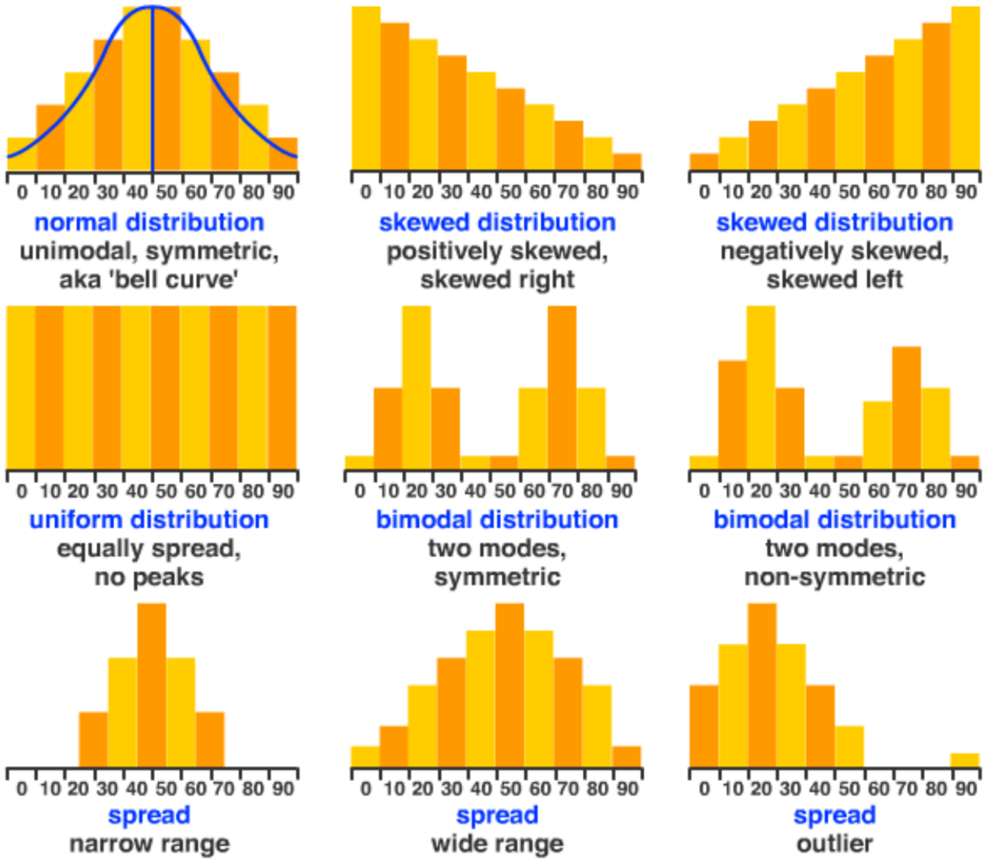
IDARE®

Basic concepts of Statistics used in AI

Sample, Description and Distribution

1. Sample: Statistics allows us to take a sample, or portion of the entire population or possibilities. Need to keep in mind
   - Is that sample represents entire population?

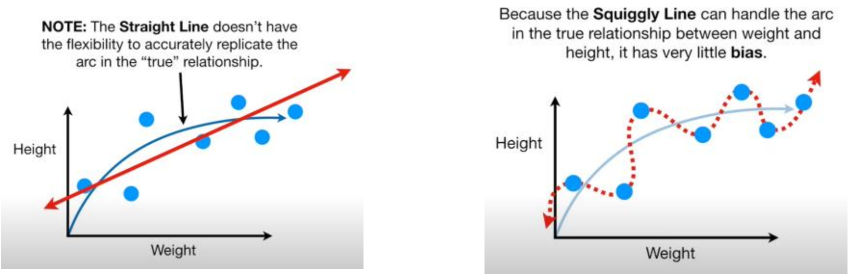
2. Descriptive Statistics
   - Mean = Average
   - Median: the middle Value
   - Mode : the value occurs most often
3. Data Distribution:
   - The distribution of data often represent with histogram plot to understand how the values of the data are distributed.



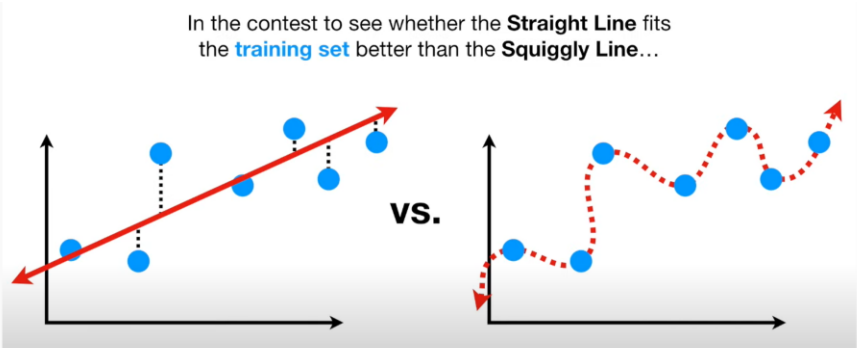
Bias and Variance

IDARE®

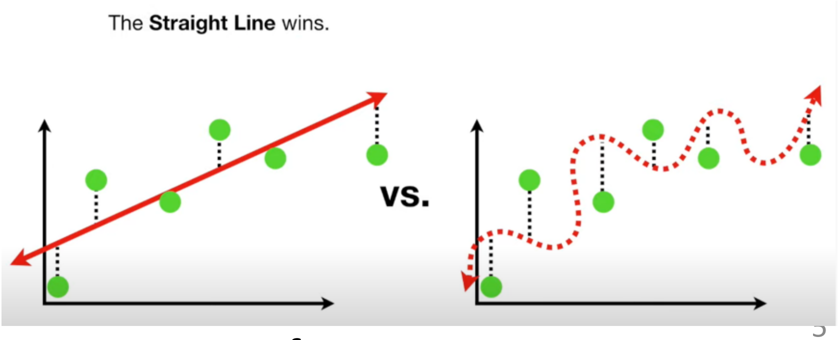
4. **Bias (Training Error): **How well the algorithm learns the true behavior or how complicated your model is
    - *Low bias: over complicated model*
    - *High bias: over simplified model*
    - Sum of Squared Error for each predicted points with training data set
5. Variance **(Testing Error)**: Measures the differences between actual and predictions.
    - Sum of Squared Error for each predicted points with test data set



Training data set trains with 2 algorithm



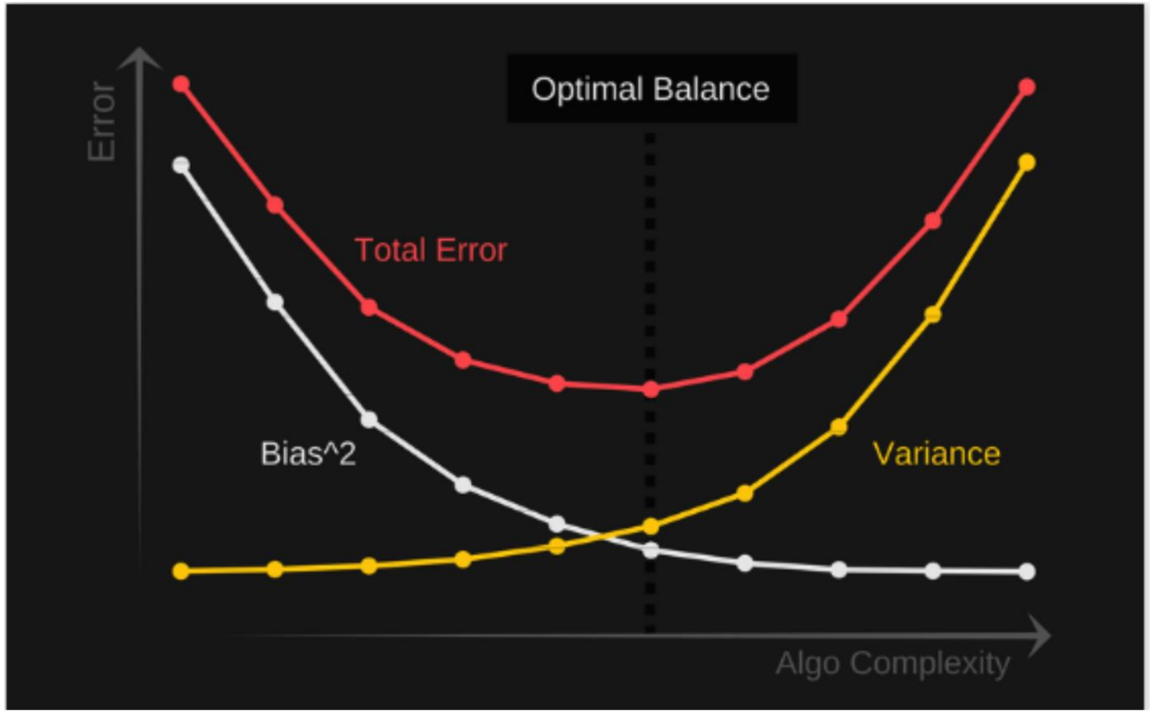
Error for Training Data sets



Error for Test Data sets

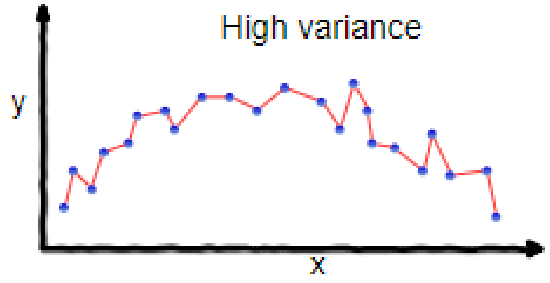
Finding What configuration of A Model provide best solution

Total Error = Bias^2 + Variance + Irreducible Error

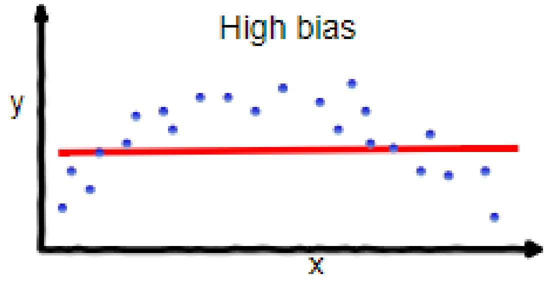


In machine learning bias and variance make up the overall expected error for our predictions. In an ideal world, we would have both low bias and low variance. However, in practice minimizing bias will usually result in an increase in variance and vice versa. The **bias/variance trade-off** describes the process of balancing these two errors to minimize the overall error for a model.

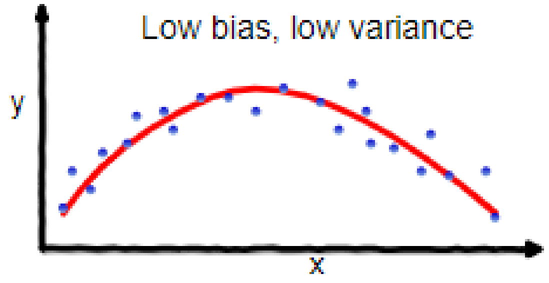
Overfitting-Underfitting



overfitting



underfitting

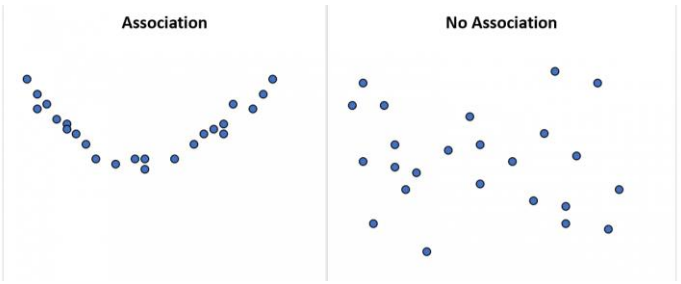


Good balance

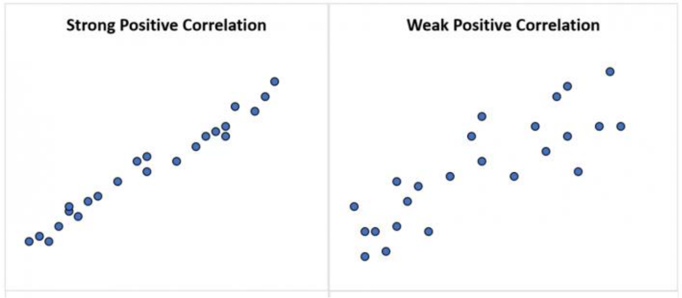
Overfitting: Training error low ,testing error high→ Model Low Bias high variance
Underfitting: Training error high, testing error high→ Model high Bias, high variance

Association, Correlation & Causation

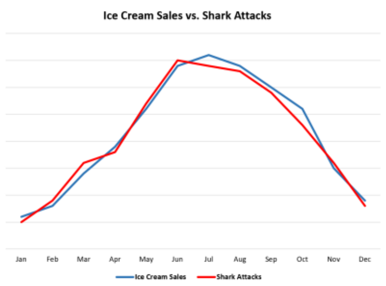
- Association that measure relationship between two variables either linear or non-linear



- Correlation: if the association assumes linear relationship between two variables i.e. Pearson Correlation coefficient
    - varies between -1 to +1



- Causation explicitly applies to cases where action A causes outcome B. p-value could be one of the measure of causation. Correlation does not mean causation.



p-value or probability value

- A measure of the affect of one variable to another variable or causation

- The value varies between 0 to 1
  - Lower value means one variable affects the other variable significantly

  - Higher value means one variable affects the other variable not too much.

  - For example, p-value is calculated between variable A and Variable B. A *p*-value less than 0.05 (typically ≤ 0.05) is statistically significant. It indicates strong evidence that there is more than a 95% probability that one variable affects the other variable

- However, low p-value does not always garounty causation

Types of Algorithms utilized in AI

IDARE®

Statistical Algorithm

- Designed understands relationship between variables
- Learn From Past and Predict Future
- Serves as the basis of explainability of the solution
- Useful when the problem is well understood

Linear Regression, logistic Reg. , SVR etc.

Machine Learning Algorithm

- Designed to provide accurate prediction
- Learn From Past's Mistake and Predict Future
- Useful when the problem is less understood

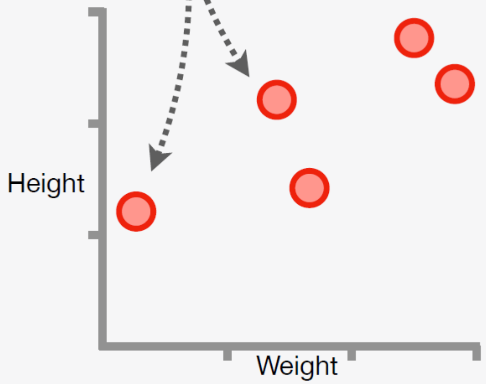
Decision Tree, Neural Network, Deep Learning, Gradient Boosting etc

IDARE®

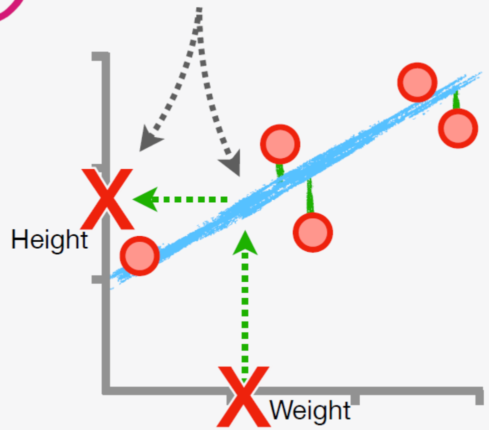
Statistical Algorithms

Linear Regression: Main Ideas

1 **The Problem:** We've collected Weight and Height measurements from **5** people, and we want to use Weight to predict Height, which is ***continuous***…



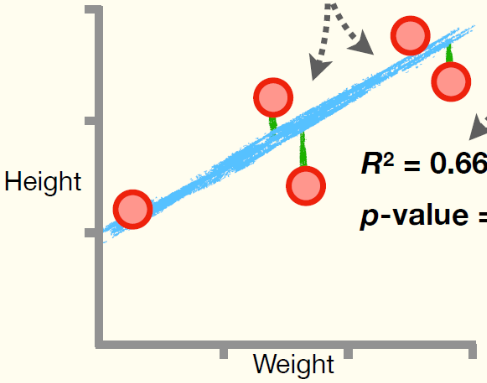
…and in **Chapter 3**, we learned that we could fit a line to the data and use it to make predictions.



However, **1)** we didn't talk about how we fit a line to the data…

…and **2)** we didn't calculate a ***p*-value** for the fitted line, which would quantify how much confidence we should have in its predictions compared to just using the mean y-axis value.

2 **A Solution: Linear Regression** fits a line to the data that *minimizes* the **Sum of the Squared Residuals** (**SSR**)…



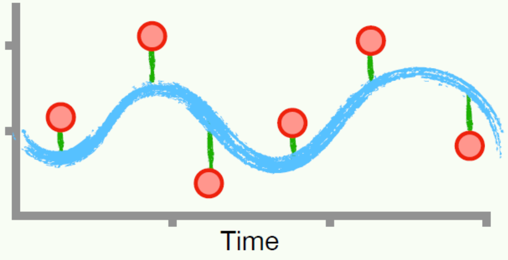
R^2 = 0.66

***p*-value = 0.1**

…and once we fit the line to the data, we can easily calculate R^2, which gives us a sense of how accurate our predictions will be…

…and **Linear Regression** provides us with a ***p*-value** for the R^2 value, so we can get a sense of how confident we should be that the predictions made with the fitted line are better than predictions made with the mean of the y-axis coordinates for the data.

NOTE: Linear Regression is the gateway to a general technique called **Linear Models**, which can be used to create and evaluate models that go way beyond fitting simple straight lines to data!!!



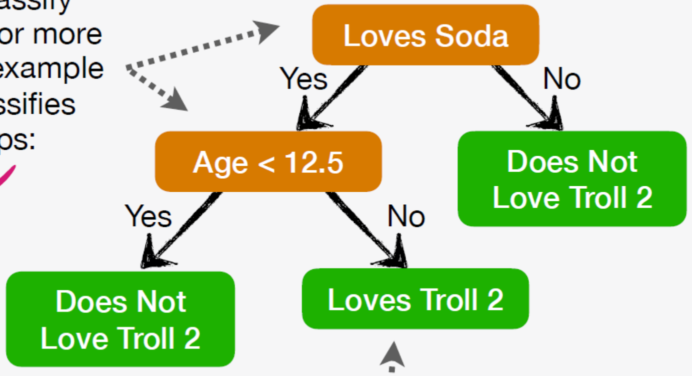
BAM!!!

Machine Learning Algorithms

Classification and Regression Trees: Main Ideas

1 There are two types of **Trees** in machine learning: trees for **Classification** and trees for **Regression**.

2 **Classification Trees** classify people or things into two or more *discrete* categories. This example **Classification Tree** classifies people into two groups:



…those who *do not* love Troll 2…

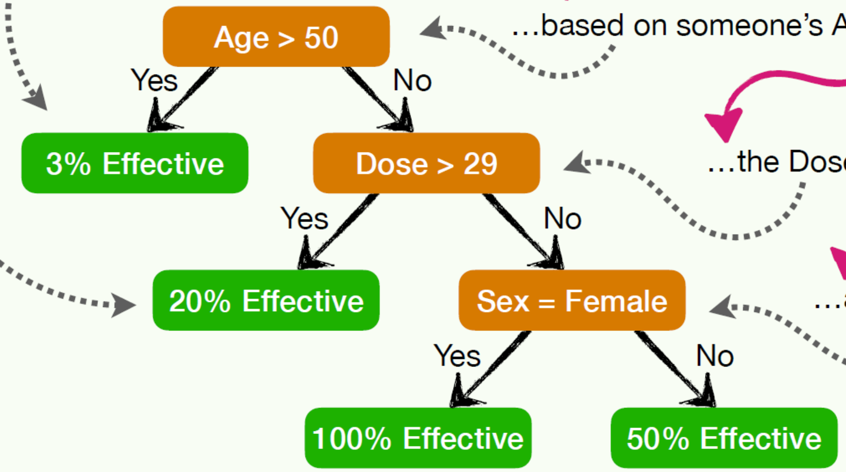
…and those who love Troll 2.

3 In contrast, **Regression Trees** try to predict a *continuous* value. This example **Regression Tree** tries to predict how Effective a drug will be…

…based on someone's Age…

…the Dose…

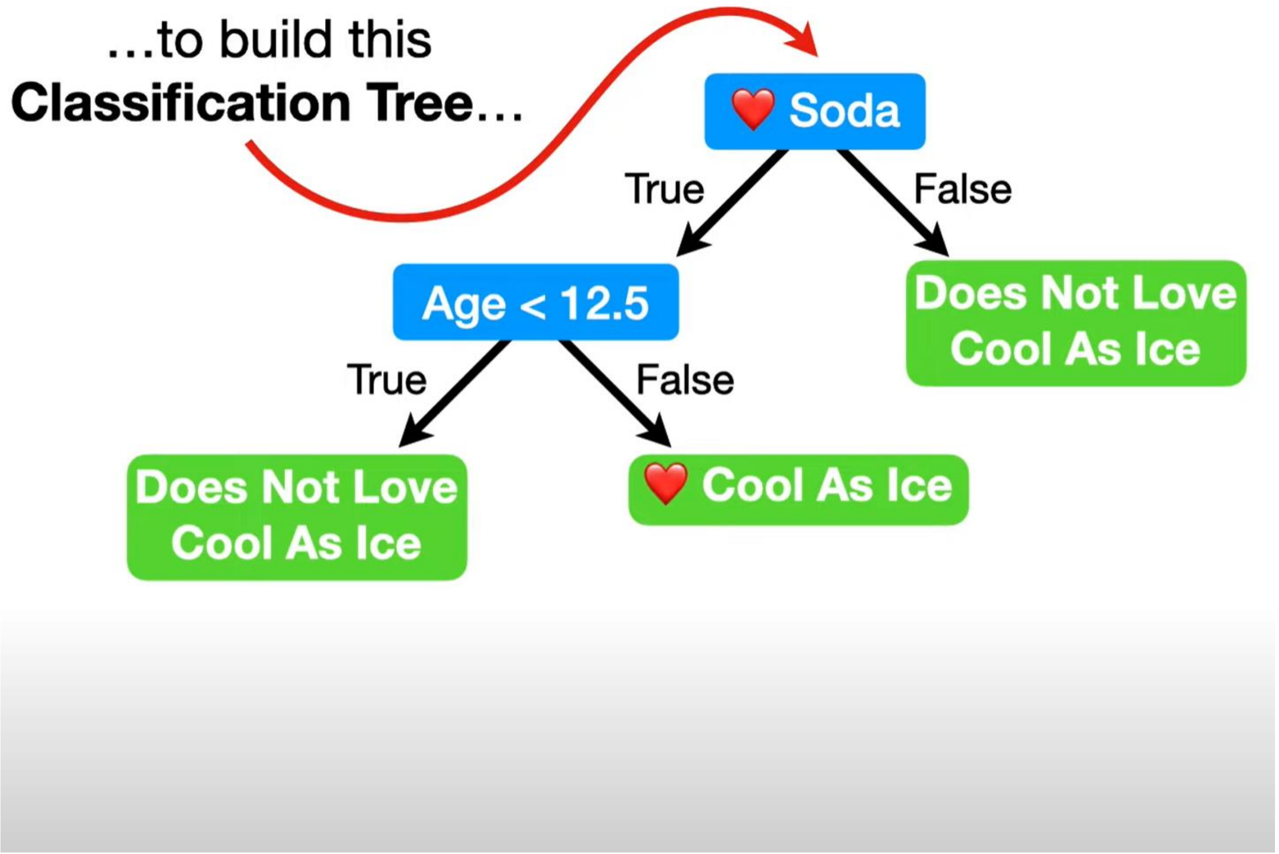
…and their Sex.



4 In this chapter, we'll cover the **Main Ideas** behind **Classification Trees** and **Regression Trees** and describe the most commonly used methods to build them. But first, it's time for the dreaded…

ML Algorithm Decision Tree

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No



ML Algorithm: Random Forest

Original Dataset

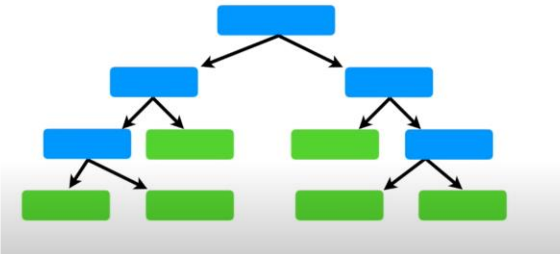
Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
Yes	No	Yes	167	Yes

Bootstrapped Dataset

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	Yes	Yes	180	Yes
No	No	No	125	No
Yes	No	Yes	167	Yes
Yes	No	Yes	167	Yes

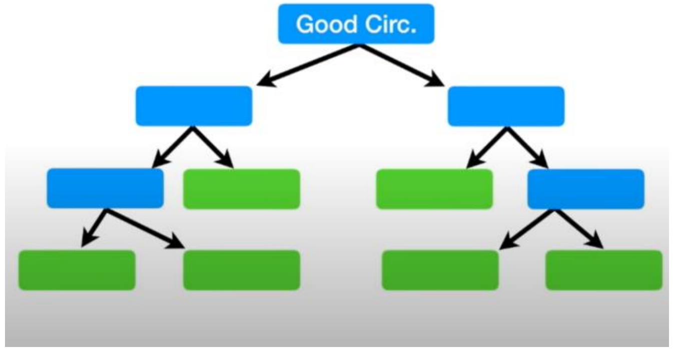
Decision Tree

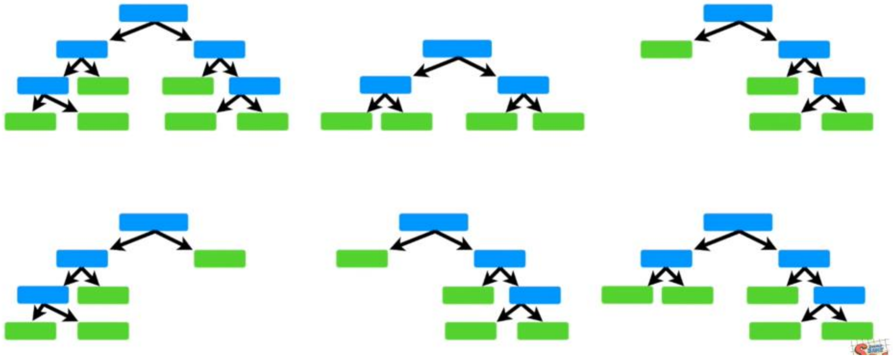
Decision Trees are easy to build, easy to use and easy to interpret…



Random Forest

1) Using a bootstrapped dataset
2) Only considering a random a subset of variables at each step.



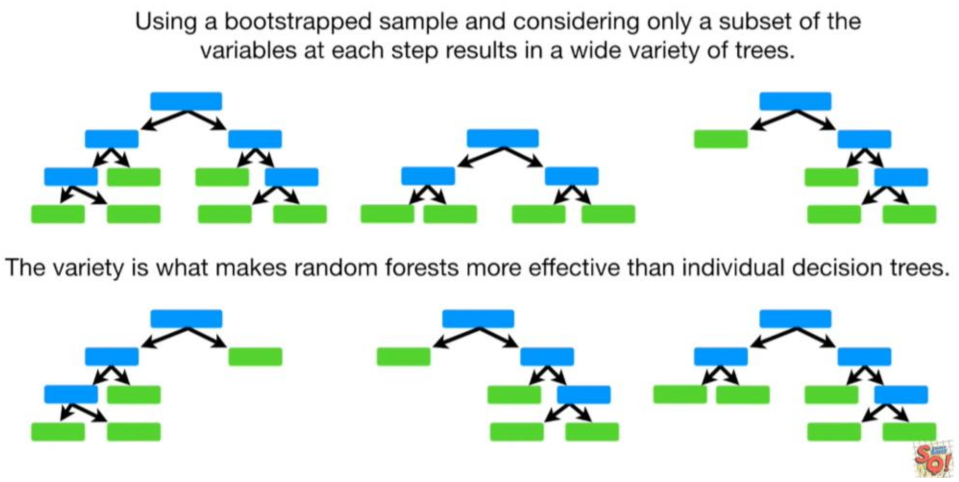


Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	No	No	168	YES

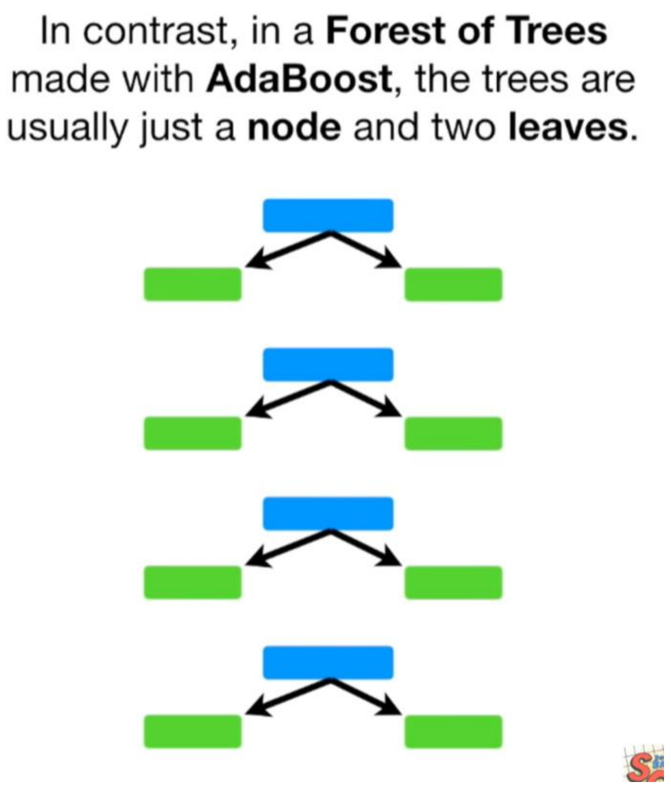
ML Algorithm: ADA Boost

IDARE®

Random Forest



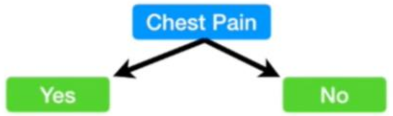
ADA Boost



ML Algorithm: ADA Boost

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
Yes	No	Yes	167	Yes

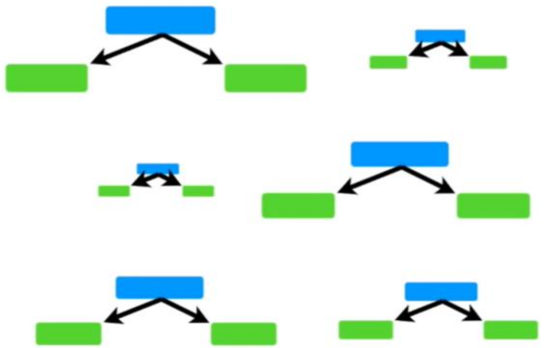
…but a **Stump** can only use one variable to make a decision.



Thus, **Stumps** are technically "weak learners".

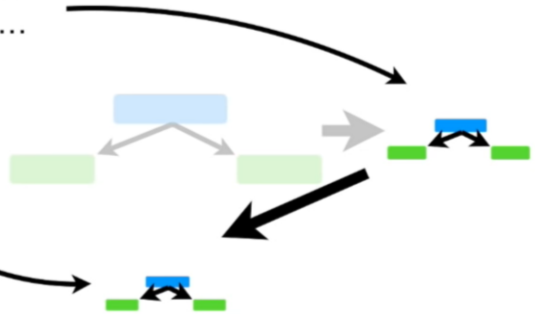
However, that's the way **AdaBoost** likes it, and it's one of the reasons why they are so commonly combined.

In contrast, in a **Forest of Stumps** made with **AdaBoost**, some stumps get more say in the final classification than others.



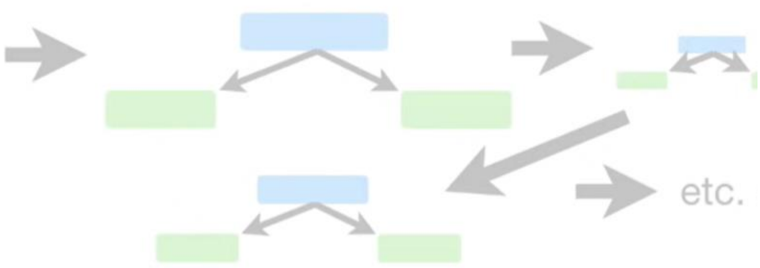
…and the errors that the second stump makes…

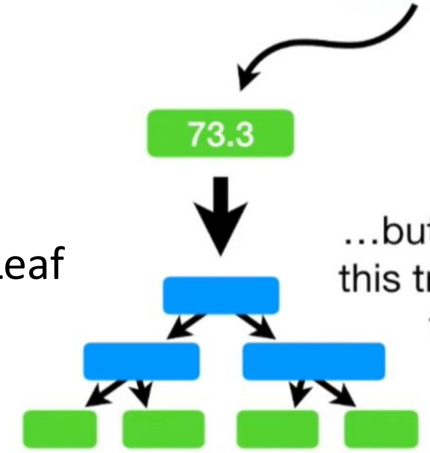
…influence how the third stump is made.



ML Algorithm: Gradient Boost

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc...	etc...	etc...	etc...



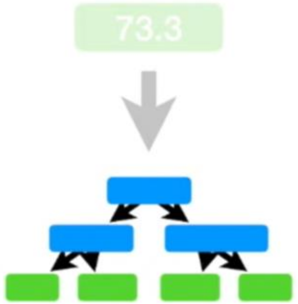


Leaf

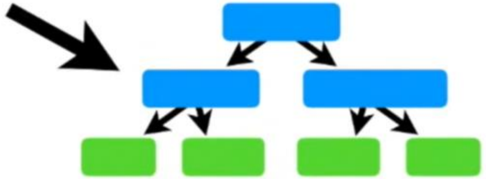
...but unlike **AdaBoost**, this tree is usually larger than a stump.

That said, **Gradient Boost** still restricts the size of the tree.

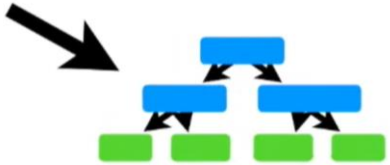
However, in practice, people often set the maximum number of leaves to be between **8** and **32**.

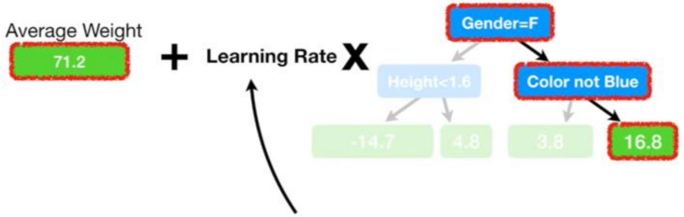


Then **Gradient Boost** builds another tree based on the errors made by the previous tree...



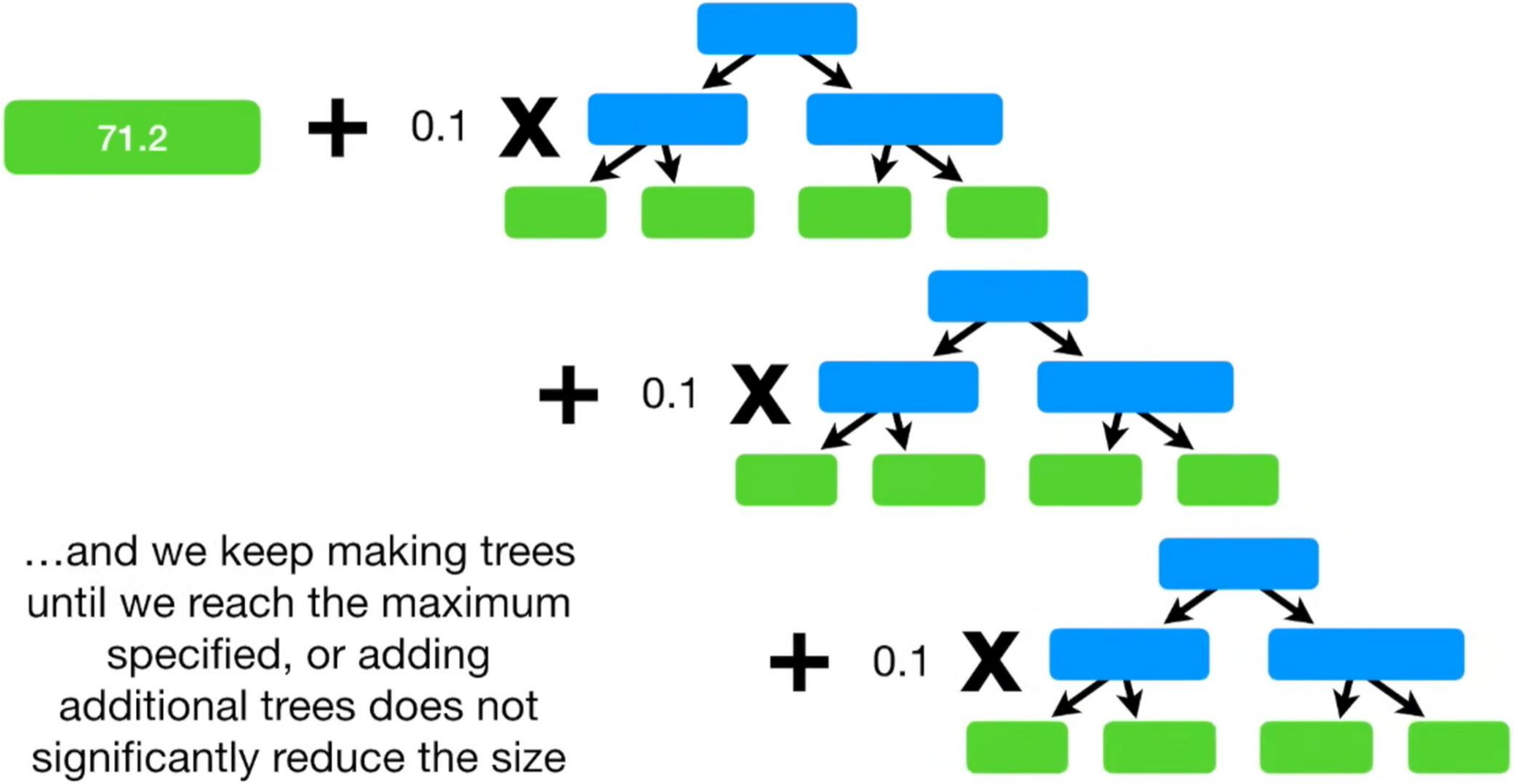
...and then it scales the tree...





Gradient Boost deals with this problem by using a **Learning Rate** to scale the contribution from the new tree.

ML Algorithm: Gradient Boost



...and we keep making trees until we reach the maximum specified, or adding additional trees does not significantly reduce the size of the **Residuals**.

...etc...etc...etc...

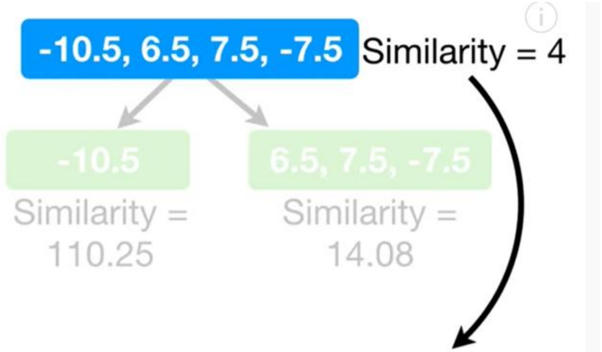
ML Algorithm: XG Boost

Similar to Gradient boost, ADA Boost

XG boost Initial Guess is 0.5

$$\text{Similarity Score} = \frac{\text{Sum of Residuals, Squared}}{\text{Number of Residuals} + \lambda}$$

Remember **λ (lambda)** is a **Regularization Parameter**, which means that it is intended to reduce the prediction's sensitivity to individual observations.



$$\text{Gain} = \text{Left}_{\text{Similarity}} + \text{Right}_{\text{Similarity}} - \text{Root}_{\text{Similarity}}$$

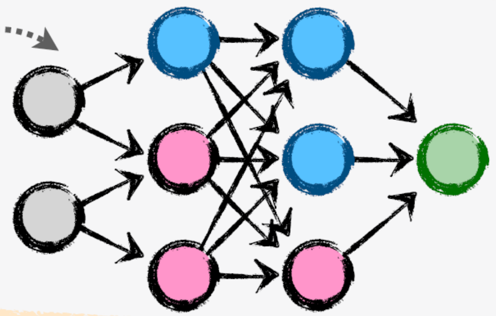
Can Change the tree structure by cutting branches

Added new parameter such as Gain, lambda gamma

Neural Network

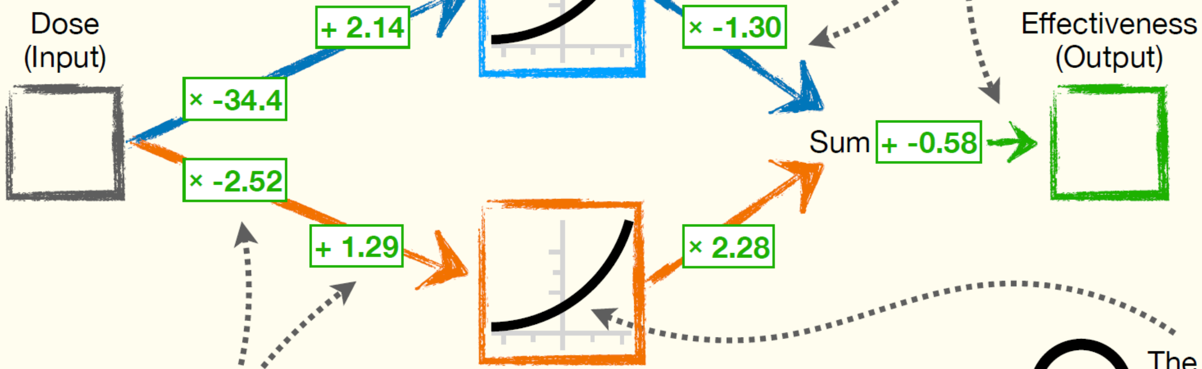
(Oh no! It's the dreaded **Terminology Alert!!!**)

1 Although **Neural Networks** usually look like super complicated groups of neurons connected by synapses, which is where the name **Neural Network** comes from, they're all made from the same simple parts.



2 **Neural Networks** consist of **Nodes**, the square boxes…

3 …and connections between **Nodes**, the arrows.



4 The bent or curved lines inside some of the **Nodes** are called **Activation Functions**, and they make **Neural Networks** flexible and able to fit just about any data.

5 The numbers along the connections represent parameter values that were estimated when this **Neural Network** was fit to data using a process called **Backpropagation**. In this chapter, we'll see exactly what the parameters do and how they're estimated, step-by-step.

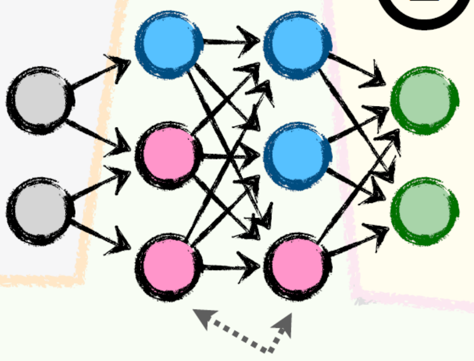
IDARE®

Terminology Alert!!! Layers

(Oh no! *A Double* **Terminology Alert!!!**)

1 **Neural Networks** are organized in **Layers**. Usually, a **Neural Network** has multiple **Input Nodes** that form an **Input Layer**…

2 …and usually there are multiple **Output Nodes** that form an **Output Layer**…



3 …and the **Layers** of **Nodes** between the **Input** and **Output Layers** are called **Hidden Layers**. Part of the *art* of **Neural Networks** is deciding how many **Hidden Layers** to use and how many **Nodes** should be in each one. Generally speaking, the more **Layers** and **Nodes**, the more complicated the shape that can be fit to the data.

4 In the example we'll use, we have a single **Input Node**…

…a single **Hidden Layer** with **2 Nodes** in it…

…and a single **Output Node**.