**Lesson 1**

Regression

(graphs drawn using <https://www.desmos.com/calculator>)

Linear Regression is a topic in statistics that is used as a predictive analysis tool.

Regression can be non-linear.

It aims to form relationships between data points and:

* determining the strength of predictors,
* forecasting an effect,
* trend forecasting.

--------------------------------------------------------------------------

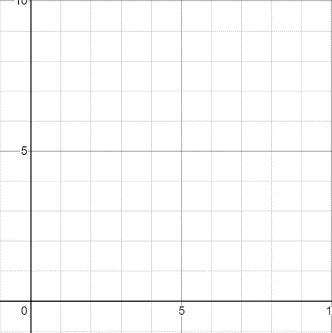
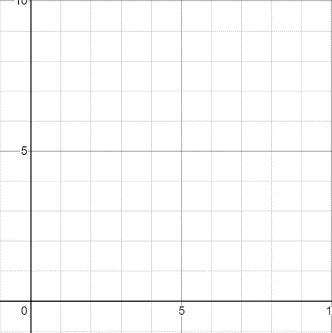
0 - Linear Regression Analysis

x axis : Independent variable

y axis : Dependent variable

graph 1 : positive relationship (m)

graph 2 : positive relationship (-m)



Examples with the dataset

1 - Least-squares regression (line of best fit)

This aims to place a line across as many datapoints as possible, by using formulae to calculate the gradient and intersect with the dataset.

Gradient

Intersect

Dataset:

x = [1, 1, 2, 3, 4, 3, 4, 6, 4]

y = [2, 1, 0.5, 1, 3, 3, 2, 5, 4]





**Lesson 2**

2- Linear Regression with h(x)

Machine Learning = Construct a Hypothesis/Model

* This topic recycles concepts learned from High-school mathematics and rebrands concepts using terms found in data science.
* Our goal with h(x) is to choose parameters so that our input (x) can predict the corresponding output (y) for our training set.

Number of plot points

Number of samples in dataset

(Input)

Parameters

Hypothesis/Model

(Output)

\* A model is a dataset that represents reality

**Lesson 3**

Training models

First, let’s understand what overfitting is.

Overfitting

* happens when your model matches the dataset too closely
* This happens because your model is trying too hard to capture the noise in your training dataset
* Noise is the data points that don’t really represent the true properties of your data, but random chance
* To avoid overfitting, we can use regularization

Training Dataset (60%)

Contains outcomes to train a machine

Multiple Prediction Algorithms are applied

Adjusts weight in neural network

Cross Validation Dataset (20%)

compare the performances of the prediction algorithms based on training set (Using Cost Functions)

Aim to minimize over fitting

Original dataset

Test Dataset (20%)

Apply chosen algorithm on real-world data

Training data vs. Testing data

* A dataset (training data) should be 10x it’s dimension (testing data) and independent of the model used
* This is also to say, throwing more data at a problem won’t always get better results

**Lesson 4**

3 - Regression with Regularization – How to increase accuracy of learning model

Regularization

* This technique is used for tuning the function by adding an additional penalty term in the error function.
* this technique discourages learning a more complex or flexible model, as to avoid the risk of overfitting.

**Lesson 5**

Cost Functions – How to measure the precision of my algorithm

* This function minimizes parameters over the dataset.
* It measures the performance of a Machine Learning model for given data.
* It quantifies the error between predicted values and expected values and presents it in the form of a single real number.

Here are examples of popular cost functions:

Mean Absolute Error (MAE) measures the difference between the estimator (the dataset) and the estimated value (the prediction).

It calculates the average squared difference between the predictions and expected results.

MAE doesn’t add any additional weight to the distance between points — the error growth is linear.

i - index of sample,

ŷ - predicted value,

y - expected value,

m - number of samples in dataset.

i - index of sample,

ŷ - predicted value,

y - expected value,

m - number of samples in dataset.

Mean Squared Error (MSE) measures the difference between the estimator (the dataset) and the estimated value (the prediction).

It calculates the average squared difference between the predictions and expected results.

MSE errors grow exponentially with larger values of distance. It’s a metric that adds a massive penalty to points which are far away and a minimal penalty for points which are close to the expected result

A picture containing sky, map, table, indoor

Description automatically generated

Example - MSE



= estimated value. From using line of best fit

Example – MAE



TODO

* Bias vs. Variance
* Linear Models (Linear Regression, Multiple Regression, GLMs t-tests ANOVA, Design Matrices)
* How many correlations/Cost functions should I use to measure accuracy?
* How do I increase accuracy of my model? >> regularization?

https://stackoverflow.com/questions/47577168/how-can-i-increase-the-accuracy-of-my-linear-regression-modelmachine-learning

* Add [interaction terms](https://en.wikipedia.org/wiki/Interaction_(statistics)#In_regression) to model how two or more independent variables together impact the target variable
* Add [polynomial terms](https://en.wikipedia.org/wiki/Polynomial_regression) to model the nonlinear relationship between an independent variable and the target variable
* Add [spines](http://people.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2/TrinityWorkshop/Workshop-handouts/TW-04-Intro-splines.pdf) to approximate piecewise linear models
* Fit [isotonic regression](https://en.wikipedia.org/wiki/Isotonic_regression) to remove any assumption of the target function form
* Fit non-parametric models, such as [MARS](https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines)

What does any of this mean? ☹