



DECODED

Timeline

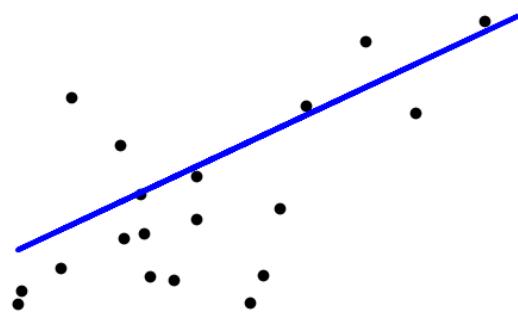


FROM MINING
TO PREDICTION

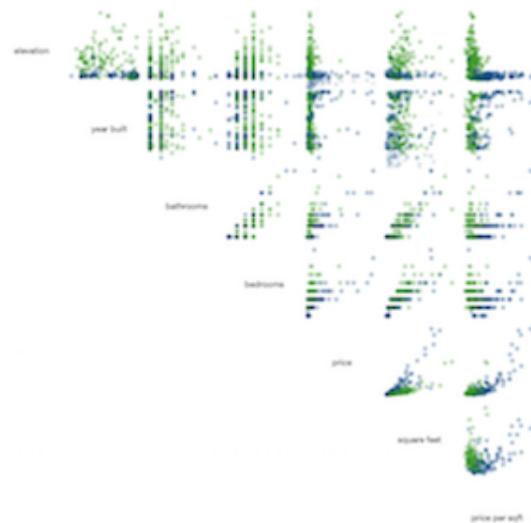
SUPERVISED LEARNING

Types of predictive models

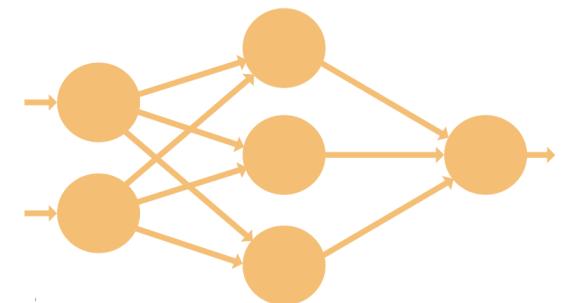
Supervised learning



Regression



Classification



Neural networks



REGRESSION

Regression

Alcohol levels in wine 

Statistical Testing

What is regression?

A technique for finding the relationship between a numerical output and a series of inputs (numerical or categorical)

A technique for finding the relationship between a numerical **output** and **a series of inputs** (numerical or categorical)

A technique for finding the
relationship between one
dependent numerical
variable and a series of
independent variables

A technique for finding **the**
relationship between one
dependent numerical
variable and a series of
independent variables

$$Y = \beta_0 + \beta_1 X_1$$

$$\text{Dependent variable} = \beta_0 + \beta_1 X_1$$

Dependent
variable

=

The point
that the line
of best fit
hits the Y
axis

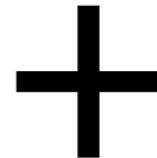
+

$\beta_1 X_1$

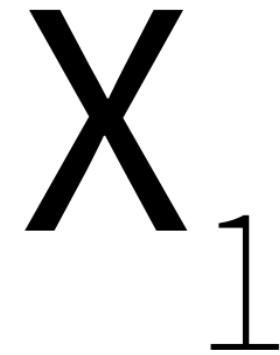
Dependent
variable



The point
that the line
of best fit
hits the Y
axis



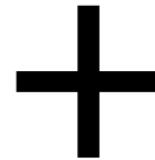
Slope
coefficient



Dependent
variable



The point
that the line
of best fit
hits the Y
axis



Slope
coefficient



Independent
variable

WINE ALCOHOL LEVEL



BRIEF

What variables can determine the alcohol level of a bottle of red wine?



BRIEF



Use regression and the Red Wine
Quality dataset to estimate wine
alcohol level!

Let's open up a new Jupyter Notebook and find the data in Kaggle

BRIEF



Explore the data and possible
correlations

Use some of the visualisation techniques that you learnt last sprint.

CORRELATION COEFFICIENT



BRIEF

Predict wine alcohol level using one physicochemical variable

Build a simple model in Python that predicts wine alcohol level using a single independent variable. This is called univariate regression.

HOW DID YOUR
MODEL PERFORM?

C O E F F I C I E N T S

R²

ROOT MEAN
SQUARED ERROR



Predict the wine alcohol level using multiple variables

Build a more complex model in Python that predicts wine alcohol level using multiple independent variables. This is called multivariate regression.



READY TO BUY
A BOTTLE OF WINE?



WHAT'S NEXT?

More complex regression techniques

Lasso and Ridge

BRIEF (5 MINUTES)

Give us your feedback

surveylink