AGENDA

- What is Statistical Learning?
- Why estimate Function?
- How to estimate a Function?
- The Trade-off between Prediction Accuracy and Model Interpretability
- Supervised vs Unsupervised Learning
- Regression vs Classification Problems

WHAT IS STATISTICAL LEARNING?

STATISTICAL LEARNING TOOLS

Definition: *Statistical learning* refers to a vast set of tools for *understanding data*.

SUPERVISED LEARNING

Definition: This involves building a statistical model for predicting, or estimating, an *output* based on one or more *inputs*.

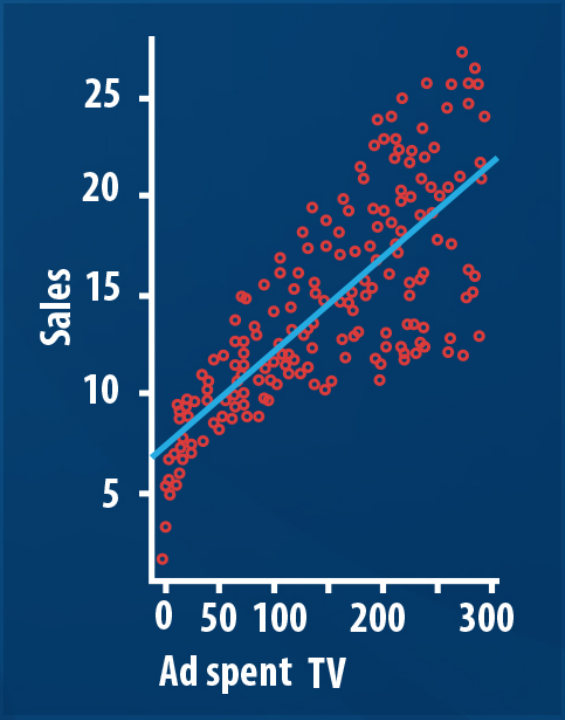
UNSUPERVISED LEARNING

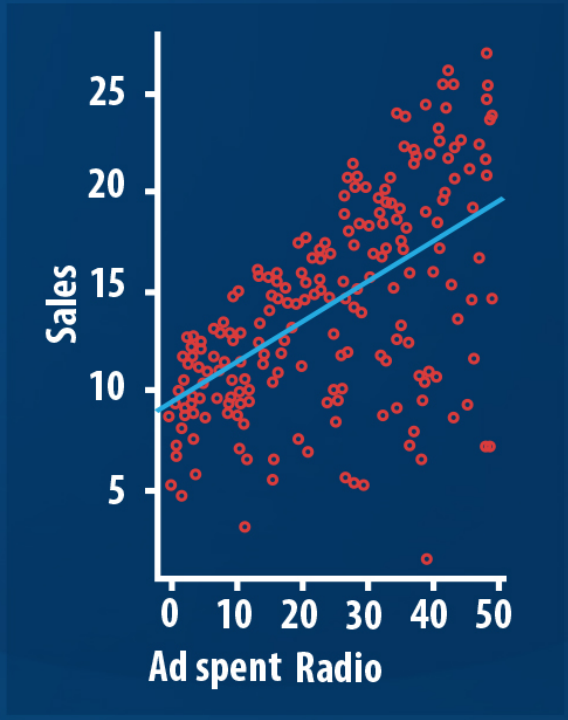
Definition: This involves inputs but no supervising output; nevertheless we can learn relationships and structure from such data.

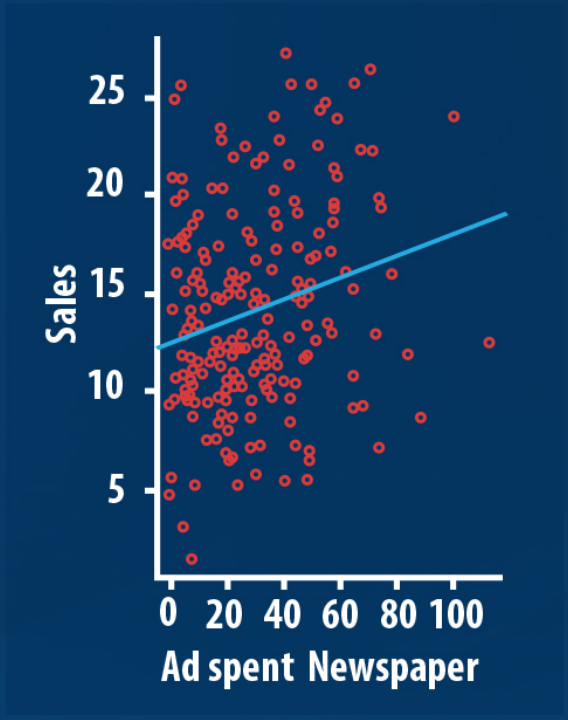
WHAT IS STATISTICAL LEARNING?

SUPERVISED LEARNING

Sales vs Ad Budgets in TV, Radio and Newspaper







Can we fit a model of the following form

$$Y = f(X) + \epsilon$$

Independent factors $X = (X_1, X_2, \ldots, X_p)$ ε = Error term

Quantitative response

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

WHY ESTIMATE F?

Typically two reasons : Prediction and Inference

Prediction

In many situations, a set of inputs *X* are readily available, but the output *Y* cannot be easily obtained. In this setting, since the error term averages to zero, we can predict *Y* using

$$Y = f(X)$$

Inference

Where f represents our estimate for f, and *Y* represents the resulting prediction for Y. In this case f is often treated as a black box, in the sense that one is not typically concerned with the exact form of f, provided that it yields accurate predictions for Y

Of Additional interest!

Which predictors are associated with the response?

What is the relationship between the response and each predictor?

Can the relationship between Y and each predictor be adequately summarized using a linear equation, or is the relationship more complicated?

HOW DO WE ESTIMATE F?

Two standard methods are used typically used : Parametric and Non Parametric methods

Parametric Methods Typically employ a two step process

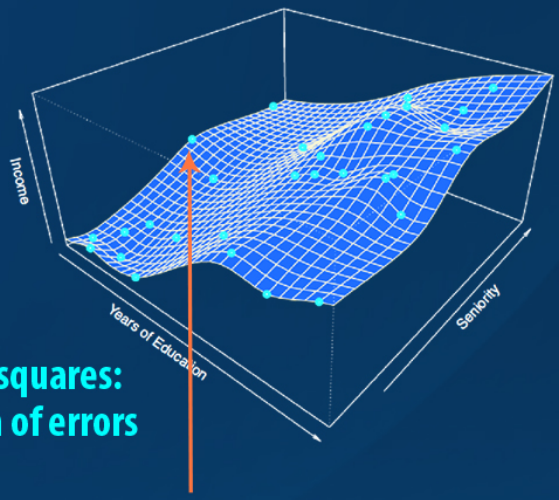
We assume a linear model:
Y - linear combination of different Xs

Step 1 Make assumption of functional form or shape of *f*

Example: $f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$

Step 2 Select a procedure to fit or train model

Example: $Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$ ← We use ordinary least squares: Minimise squared sum of errors



A thin plate spline is used to fit the data points

Non Parametric Methods

Makes an estimate of *f* without an explicit functional form

PREDICTION ACCURACY vs MODEL INTERPRETABILITY

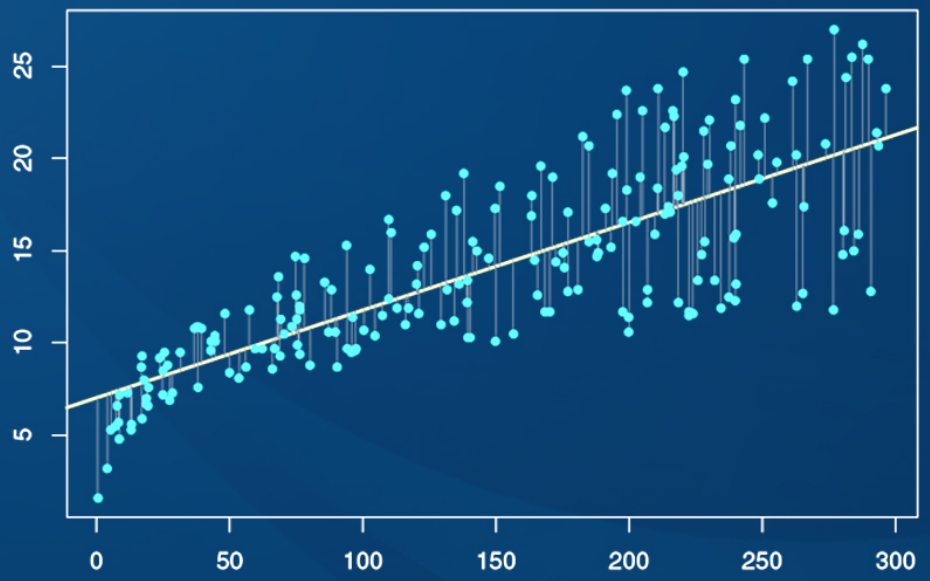
Thumb rule!

In general, as the flexibility of a method increases, its interpretability decreases.

SUPERVISED vs UNSUPERVISED LEARNING

Supervised

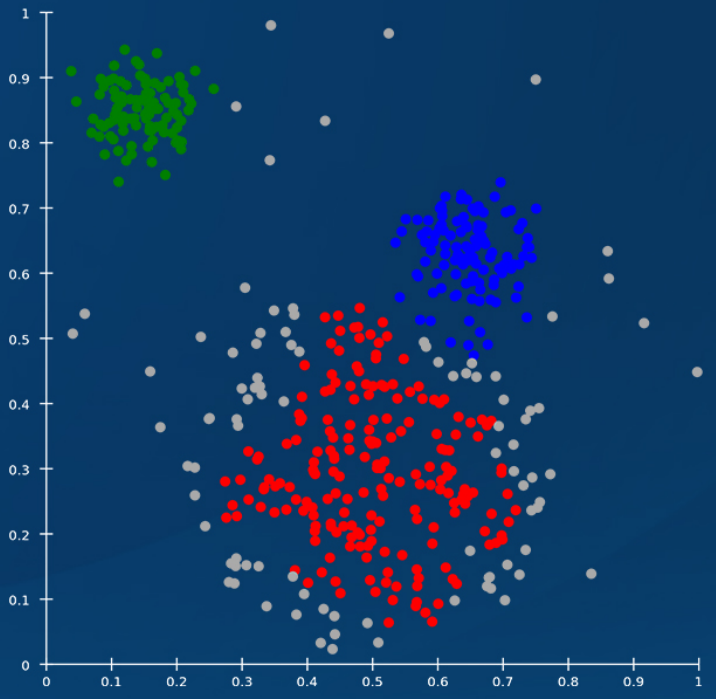
For each observation of the inputs there is an associated output measurement.



Linear Regression

Unsupervised

In unsupervised learning: for every observation we observe inputs but no associated output is observed.



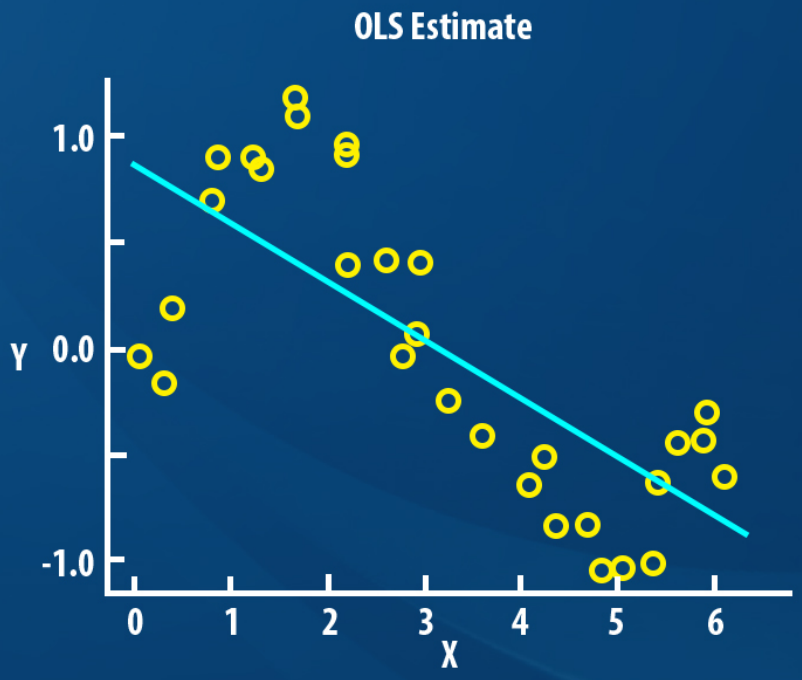
Clustering

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

REGRESSION vs CLASSIFICATION

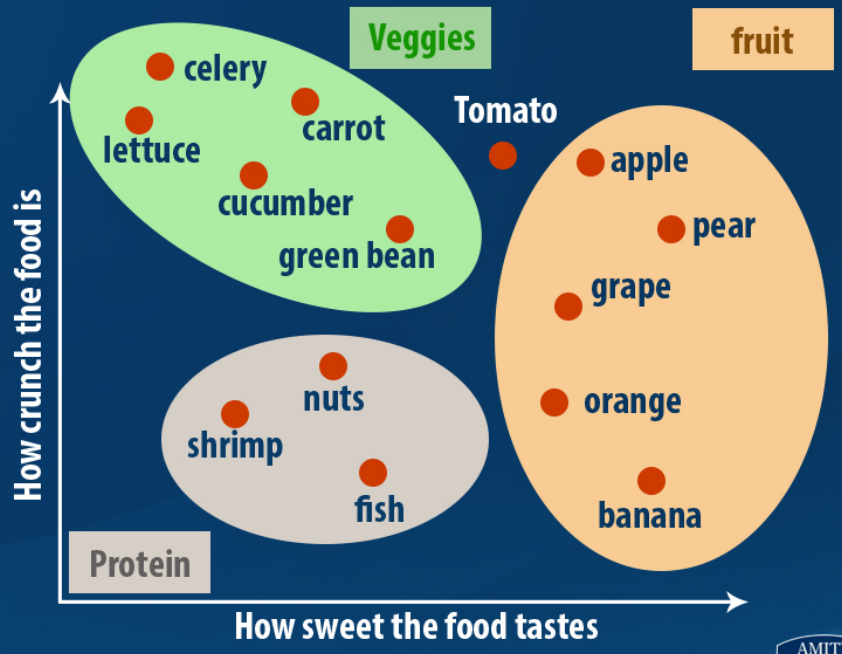
Regression

$$y = \alpha + \beta_1 x1 + \beta_2 x_2 + \ldots + \beta_k x_k$$



Classification

In unsupervised learning: for every observation we observe inputs but no associated output is observed.



AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW