Machine Learning

Source: Datacamp Courses

library(peoplesciRR)
Are there statistical differences in outcomes of HR processes?

method = multinomial
(target outcome = PE scores)

method = continuous
(target outcome = percent of target bonus)

method = binomial
(target outcome = promotion T/F)

adjusted results include features: radford level, job profile tenure, function
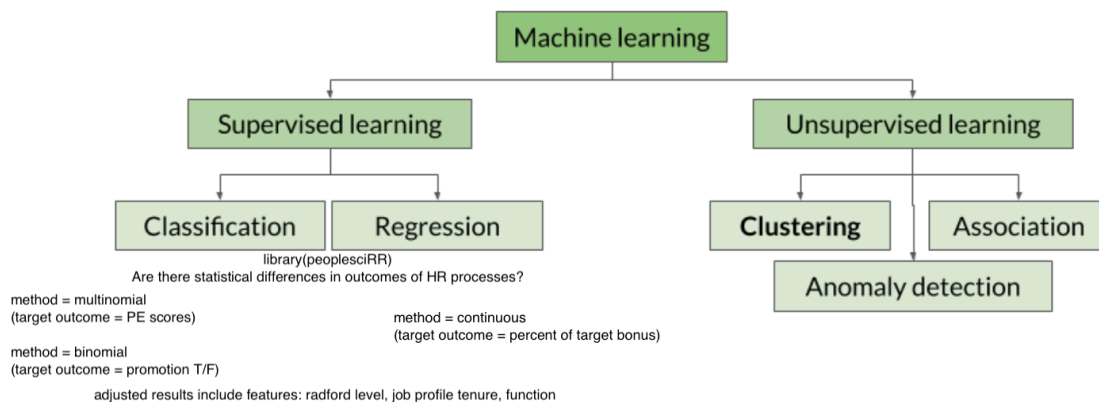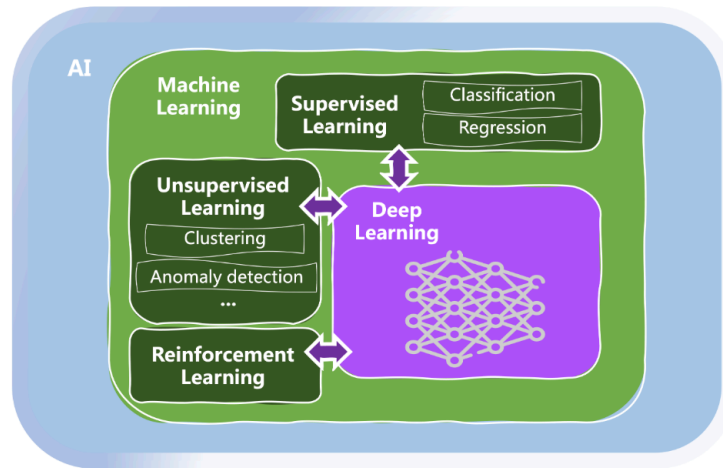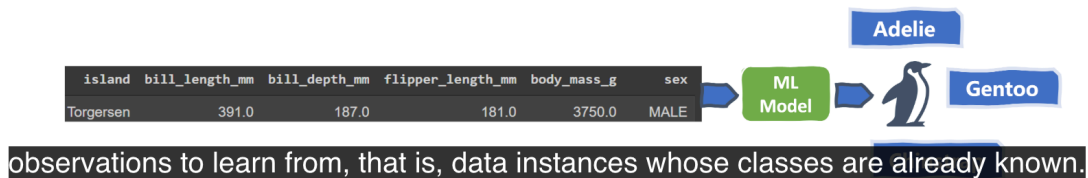
"Predicting the outcome based on gender is the same as seeing how groups differ on the outcome"

Supervised Learning: classification

Classification: assign each data observation the category (*class*) it may belong to

- **Binary classification:** two classes, e.g. positive/negative, male/female, etc.
- **Multi-class classification:** several mutually exclusive classes, e.g. multiple species

> **Supervised learning:** *Data annotation* (getting labelled observations with *known class a priori*) needed to learn/train a **model** capable of making inference

island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Torgersen	391.0	187.0	181.0	3750.0	MALE

observations to learn from, that is, data instances whose classes are already known.

1:50

Supervised Learning: regression and forecasting

**Regression:** assign each data observation a numerical output or *label* based on its inputs

House attributes → Regression Model → Price
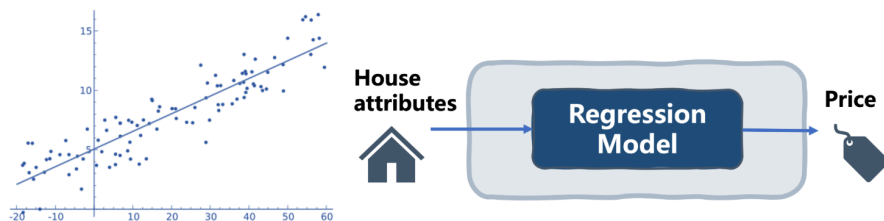
Time series forecasting: predict future values of variable, based on its past behavior

Daily bus Passengers Jan-Jun 2023 → Forecasting model → Daily bus passengers July 2023

For instance, forecasting the number of daily bus passengers for next month, based on historical data.
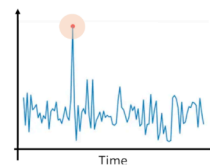
Unsupervised and reinforcement learning

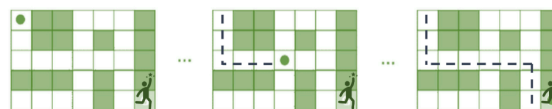**Clustering:** find subgroups of data with *similar* characteristics (e.g. *k-means* algorithm)

Anomaly detection: detecting *abnormal* data observations e.g. unusual card transactions
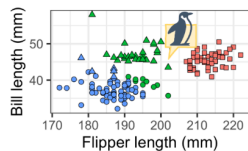
Association rule discovery: find common co-occurrences of items in transaction data

Reinforcement learning: learn by *experience* (trial and error) to master a complex task

And in reinforcement learning, an AI agent is trained to solve complex

Clusting package in R: https://cran.r-project.org/web/packages/mclust/vignettes/mclust.html