Regression and Machine Learning

Akanksha S. Kashikar

Department of Statistics
Savitribai Phule Pune University
Pune, India
akanksha.kashikar@gmail.com

Outline

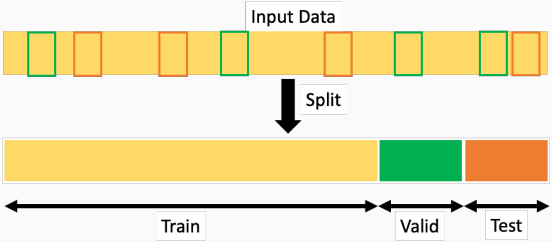
1. Introduction
2. Modifications
3. Variants in ML

Introduction

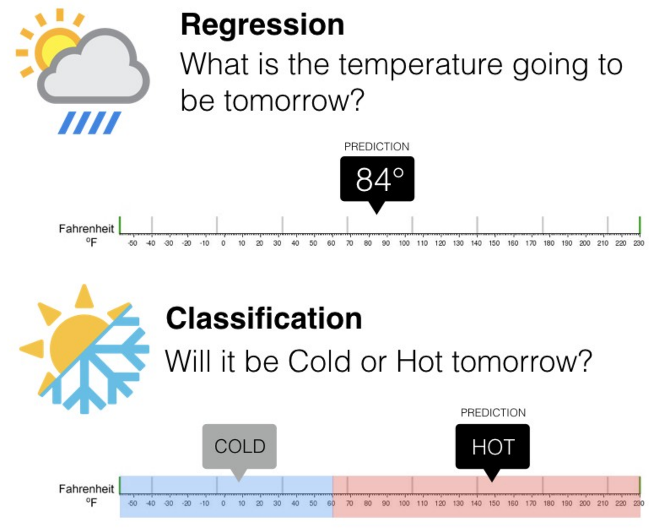
Differences in ML and Classical Statistics

- Train Error, Test Error, Validation Error
- Cross-validation
- For parameter tuning and/or model choice
- Similar to PRESS in classical regression

Train-Test-Validation Sets



WHAT DOES REGRESSION MEAN IN ML?



(Source: https://www.springboard.com/blog/data-science/regression-vs-classification/)

REGRESSION MODELS STUDIED SO FAR

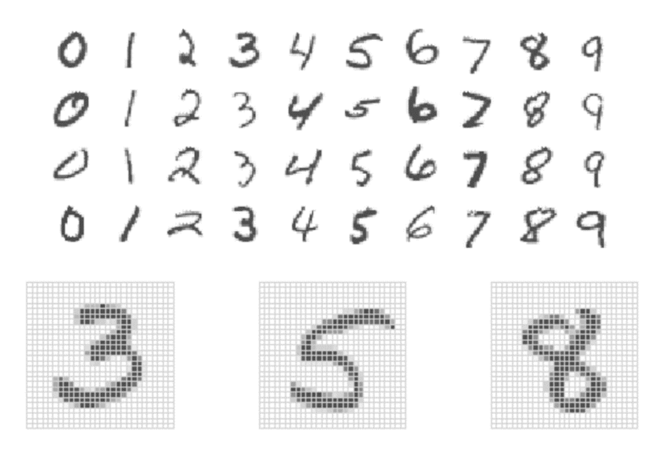
- Simple Linear Regression
- Multiple Linear Regression
- Binomial Logistic Regression
- Multinomial Logistic Regression
- Poisson Regression

Common Variants of Regression in ML

- k-nearest neighbours
- Decision Trees ($+$ bagging, boosting)
- Support Vector Machines
- Artificial Neural Networks

Modifications

HIGH DIMENSIONALITY



(Source: ISLR)

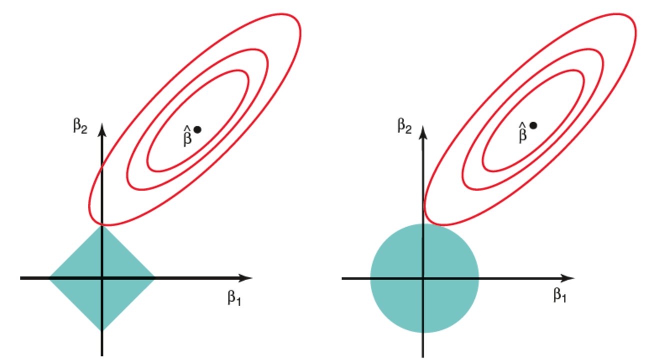
Lasso and Ridge

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \le s$$

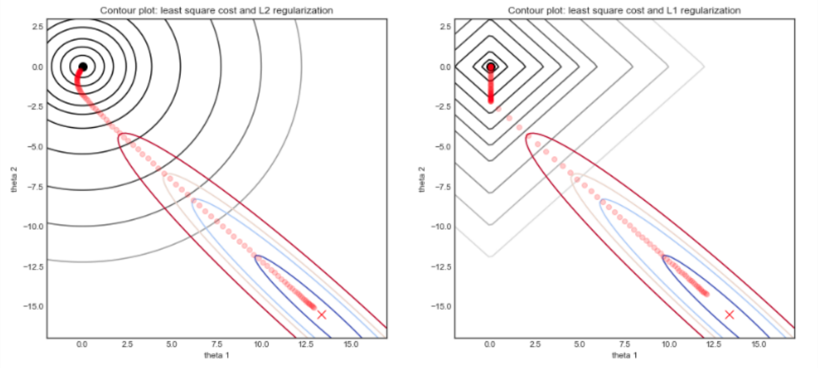
and

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \le s,$$

Lasso vs Ridge



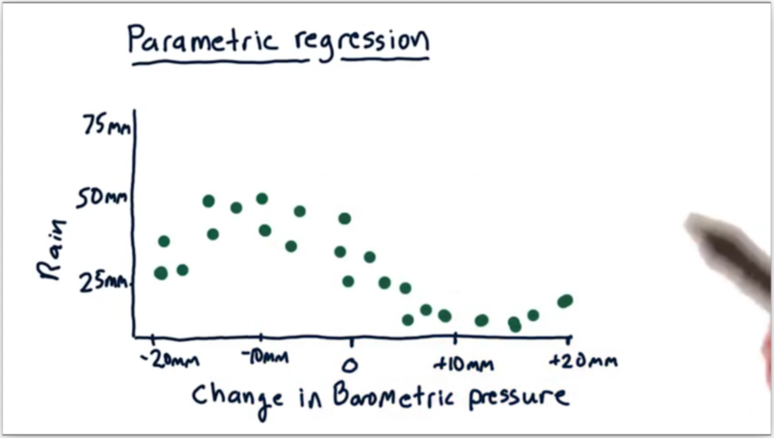
Lasso vs Ridge



(Source: https://stats.stackexchange.com/questions/348308/graphical-interpretation-of-lasso)

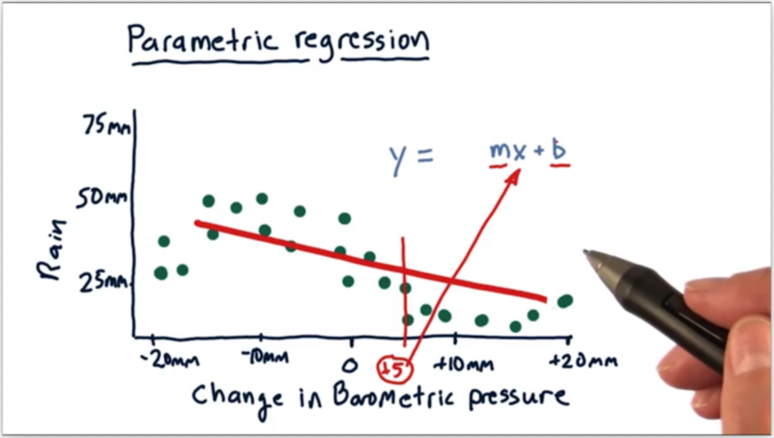
Variants in ML

NEED FOR NONLINEAR MODELLING



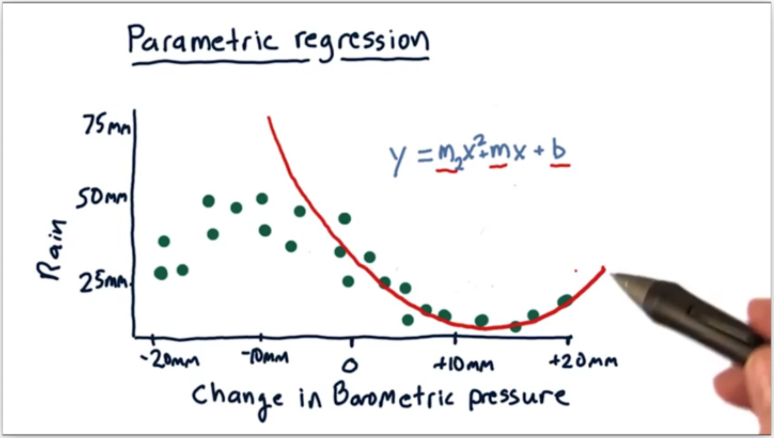
(Source: https://www.omscs-notes.com/machine-learning-trading/regression/)

Outcome with Linear Model



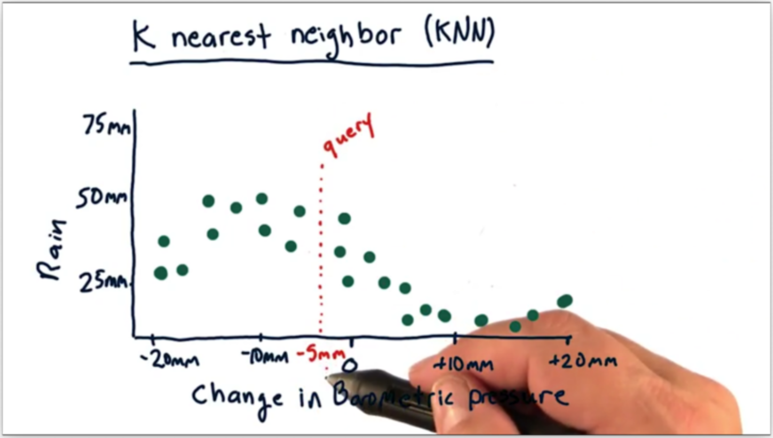
(Source: https://www.omscs-notes.com/machine-learning-trading/regression/)

OUTCOME WITH QUADRATIC MODEL



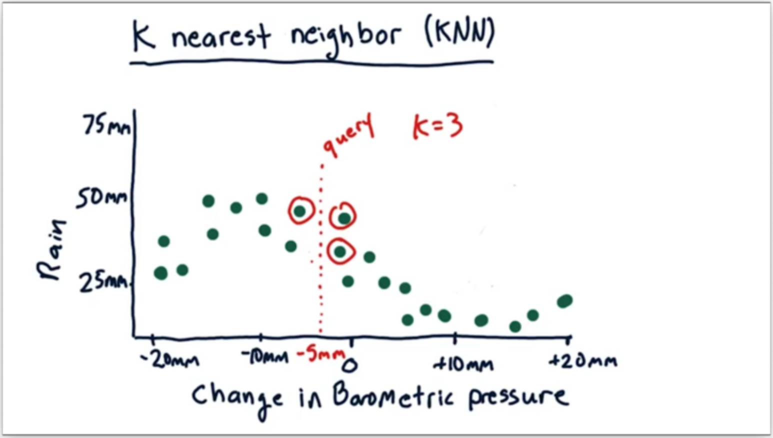
(Source: https://www.omscs-notes.com/machine-learning-trading/regression/)

k NN Regression



(Source: https://www.omscs-notes.com/machine-learning-trading/regression/)

3 NN Regression

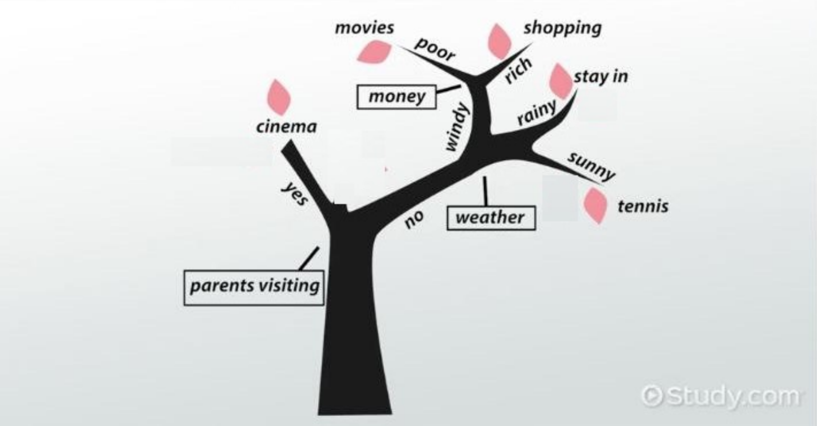


(Source: https://www.omscs-notes.com/machine-learning-trading/regression/)

k Nearest Neighbours Regression

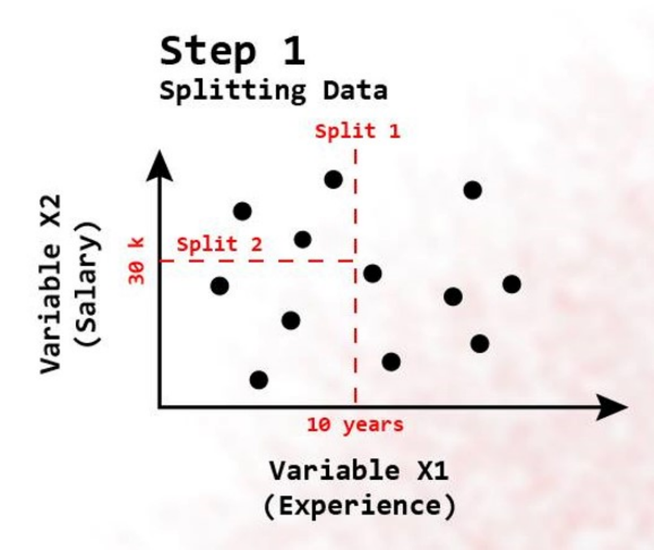
- Nonparametric
- Can capture nonlinear patterns
- Local nature

Decision Trees



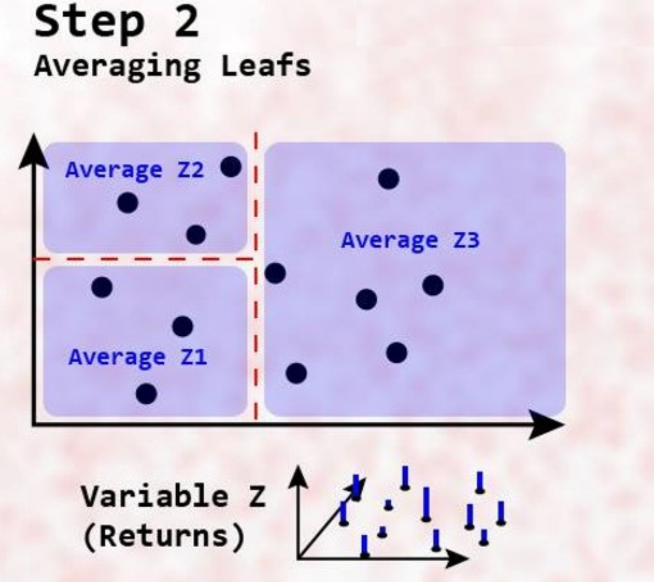
Source: https://study.com/academy/lesson/what-is-a-decision-tree-examples-advantages-role-in-management.html

Division of Predictor/Regressor Space

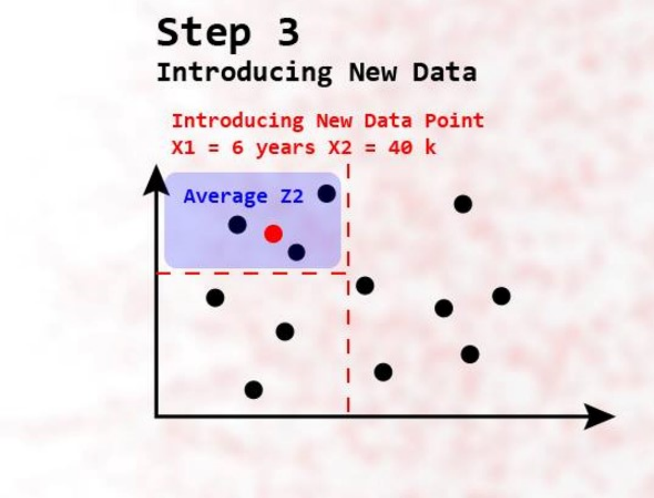


Source: artificialintelligence.digest

Fitting the Regression Tree

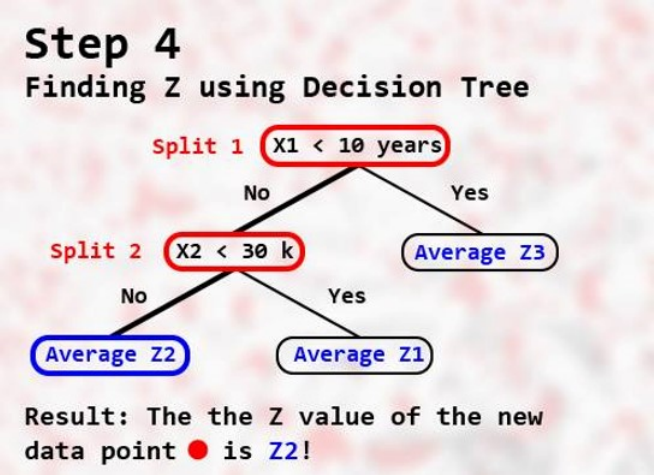


Prediction for the new observation



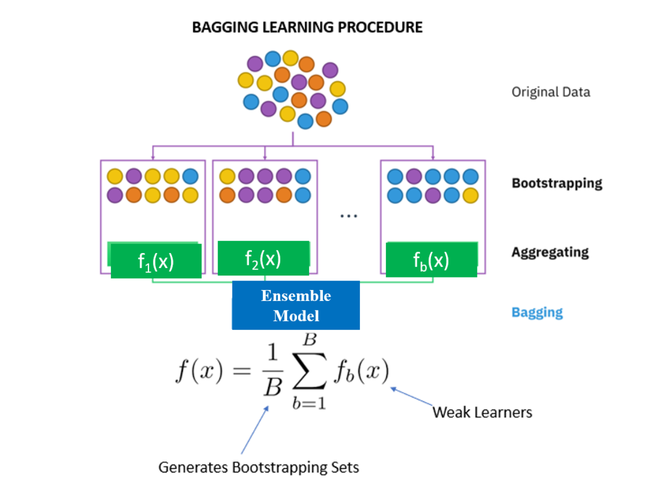
Source: artificialintelligence.digest

The Final Regression Tree



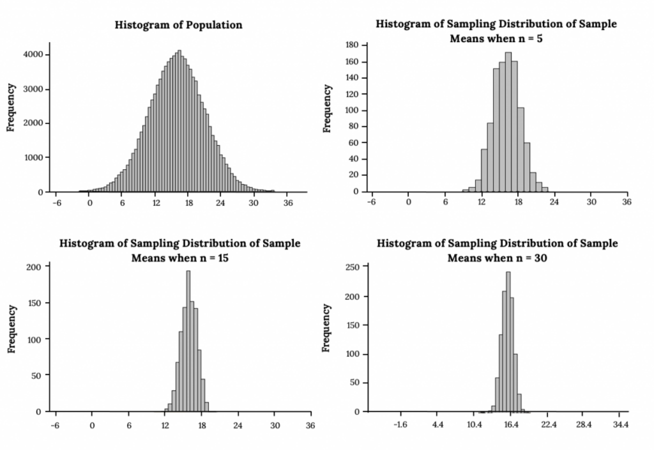
Source: artificialintelligence.digest

BAGGING



$$f(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x)$$

Bagging - Rationale



https://pressbooks.lib.vt.edu/introstatistics/chapter/the-central-limit-theorem-for-sample-means-averages/

Random Forest

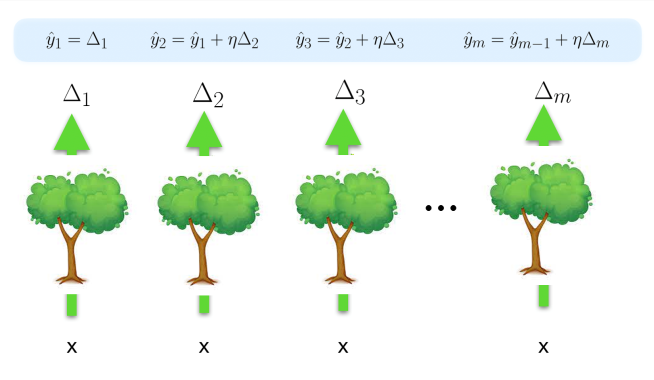
Randomly chosen predictors at each point to avoid highly correlated models

Boosting

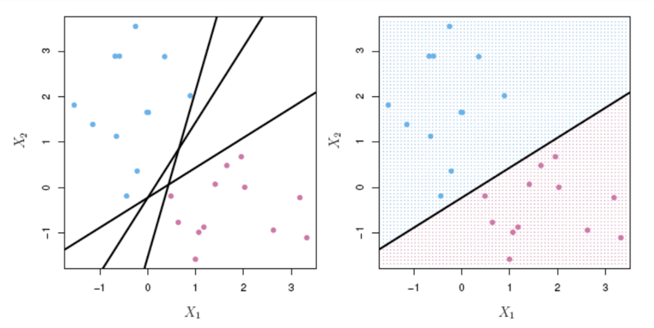
1. Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For $b = 1, 2, \ldots, B$, repeat:
   (a) Fit a tree $\hat{f}^b$ with d splits ($d+1$ terminal nodes) to the training data (X, r).
   (b) Update $\hat{f}$ by adding in a shrunken version of the new tree:
   $$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x).$$
   (c) Update the residuals,
   $$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i).$$
3. Output the boosted model,
   $$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x).$$

Source: ISLR

Boosting

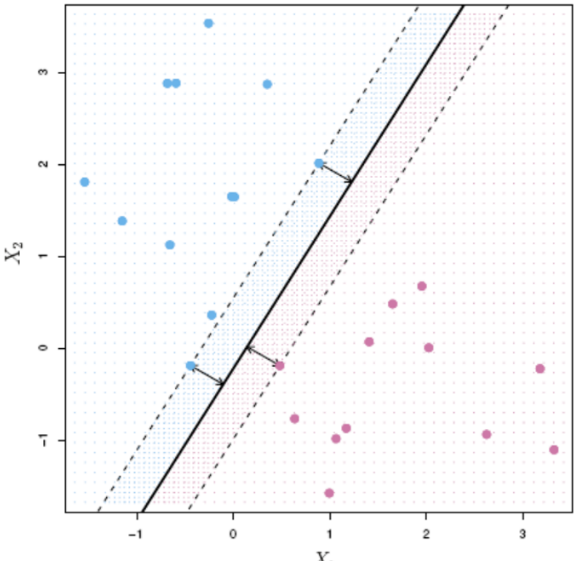

$\hat{y}_1 = \Delta_1$
$\hat{y}_2 = \hat{y}_1 + \eta\Delta_2$
$\hat{y}_3 = \hat{y}_2 + \eta\Delta_3$
$\hat{y}_m = \hat{y}_{m-1} + \eta\Delta_m$
Δ_1
Δ_2
Δ_3
Δ_m
x
x
x
x

Maximal Margin Classifier

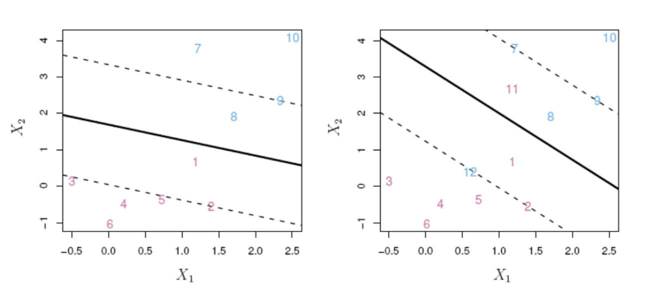


Source: ISLR

Maximal Margin Classifier

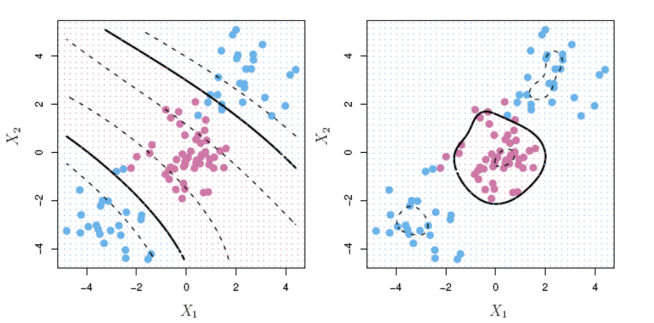


Support Vector Classifier



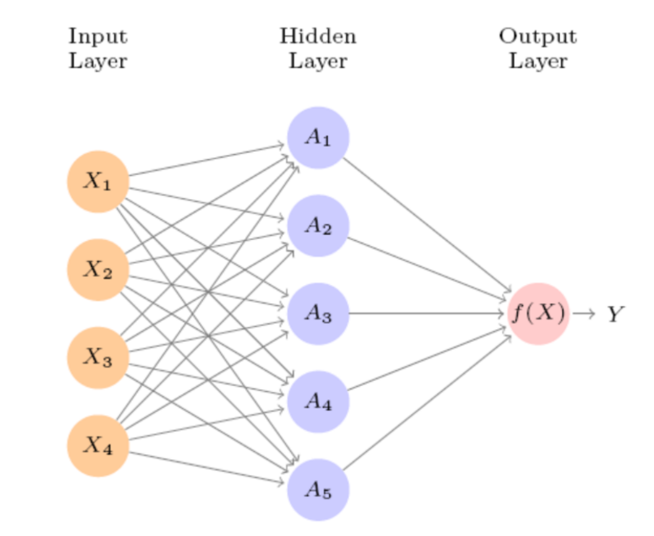
Source: ISLR

Support Vector Machines



Source: ISLR

Artificial Neural Networks



Source: ISLR

Artificial Neural Networks

$$\begin{aligned} f(X) &= \beta_0 + \sum_{k=1}^{K} \beta_k h_k(X) \\ &= \beta_0 + \sum_{k=1}^{K} \beta_k g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j). \end{aligned}$$

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k,$$

$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$

Source: ISLR


Thank
You