Global Initiative of Academic Networks (GIAN)

BRINGING SYNERGY ACROSS DIFFERENT TRANSIT MODES IN INDIA BY ADDRESSING CHALLENGES FOR SUSTAINABLE TRANSPORT MODES

JUNE 23 - 27, WARANGAL, INDIA

MODULE 4

Instructors

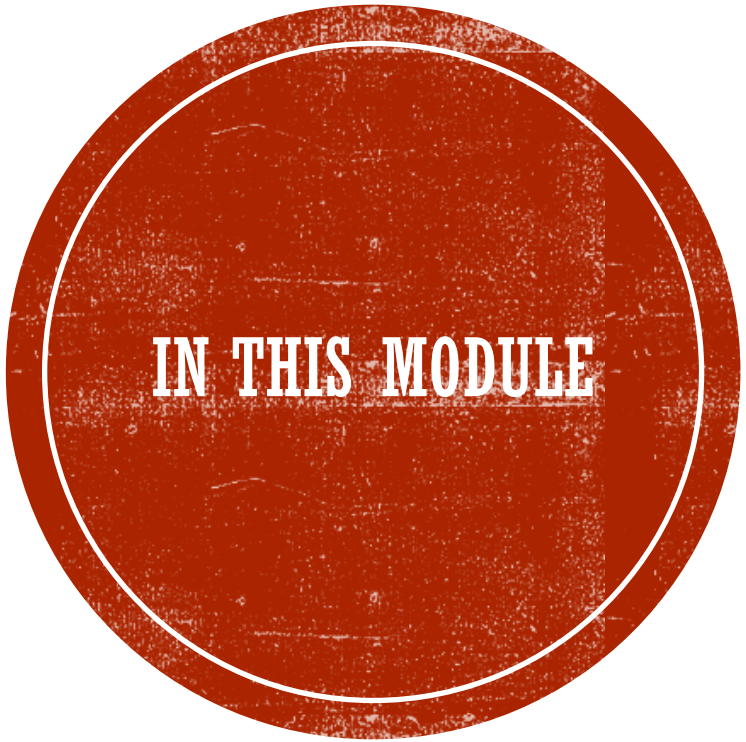
Naveen Eluru, University of Central Florida

Raghuram Kadali, NIT, Warangal

CSRK Prasad, NIT, Warangal

COURSE MODULES

Introduction	• Public Transportation – An Introduction
Public transport data	• Background on data components useful for public transportation system analysis, their compilation and consistency analysis
Modeling approaches for public transit analysis	• Introduce traditional frameworks for public transit analysis – linear regression, discrete choice models (such as multinomial logit, ordered logit, and count models)
Emerging models for public transit data analysis	• Flexible discrete choice models (NL, ML, discrete continuous models) and machine learning models (KNN, RF, SVM, Decision Tress and Gradient Boost)
Integrating emerging modes with public transit	• Bringing it all together to leverage emerging modes and data analytics to improve public transportation across India

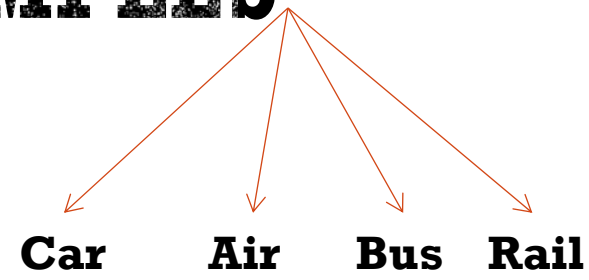


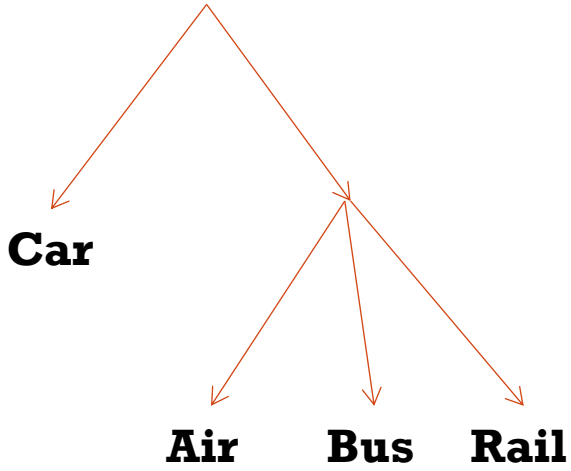
I will build on the basic choice modeling approaches for data analysis and introduce Nested Logit models, GEV models, Mixed Logit model, latent class models, discrete-continuous models and multiple discrete extreme value models

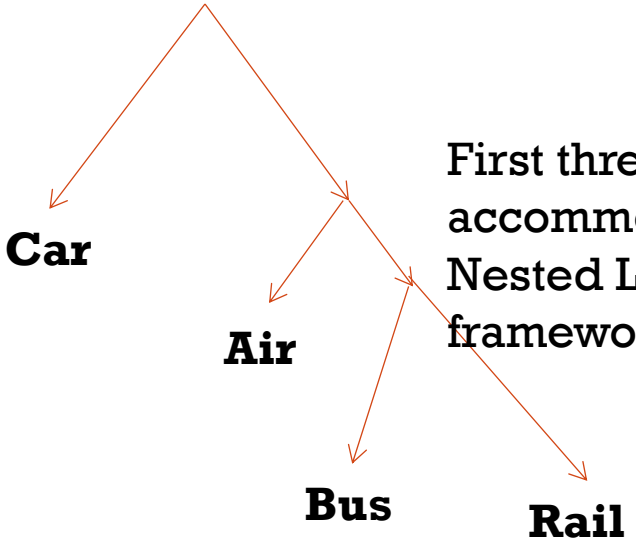
NESTED LOGIT MODEL

- MNL - $P_{in} = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$
- Independent errors
  - Consider mode choice model, with 3 alternatives car, bus and metro. A person whose personality prefers transit modes will assign a higher value to both bus and metro. Neglecting this might have implications for what we are trying to do
    - $V_{in} + \varepsilon_{in}$
- There is a "stickiness" associated to a set of alternatives i.e. the behavior of the alternatives in the "set" is different from the alternatives not in the set
- Within a set however, the behavior is again similar to that in MNL

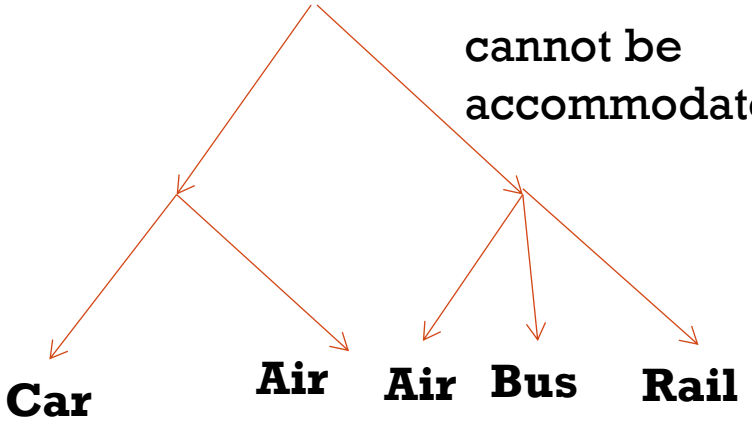
EXAMPLES







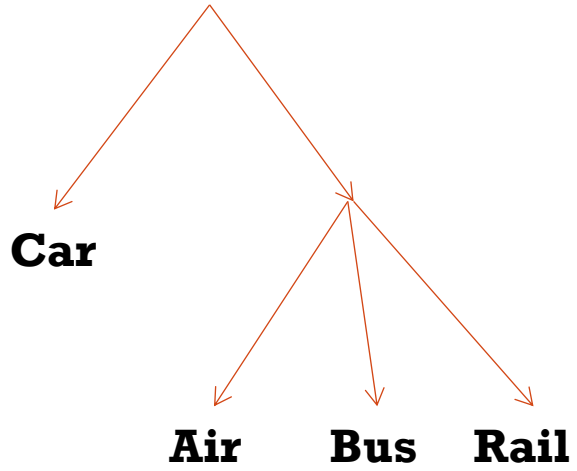
First three can be accommodated in the Nested Logit framework



cannot be accommodated

FORMULATION

- Consider the 4 alternatives: Car, Air, Bus Rail



- $U_C = V_C + \varepsilon_{Car}$
- $U_A = V_A + \varepsilon_A + \varepsilon_{cc}$
- $U_B = V_B + \varepsilon_B + \varepsilon_{cc}$
- $U_R = V_R + \varepsilon_R + \varepsilon_{cc}$
  - where ε_{cc} represents the common error term for the common carriers
- Overall error is still identical i.e. $\varepsilon_{Air} = \varepsilon_A + \varepsilon_{cc}$, $\varepsilon_{Bus} = \varepsilon_B + \varepsilon_{cc}$, $\varepsilon_{Rail} = \varepsilon_R + \varepsilon_{cc}$
- ε_{Car}, ε_{Air}, ε_{Bus}, and ε_{Rail} are distributed G(0,1)
- Now lets say ε_A, ε_B, and ε_R are Gumbel G(0,θ)
  - ($0 < \theta \leq 1$)

FORMULATION

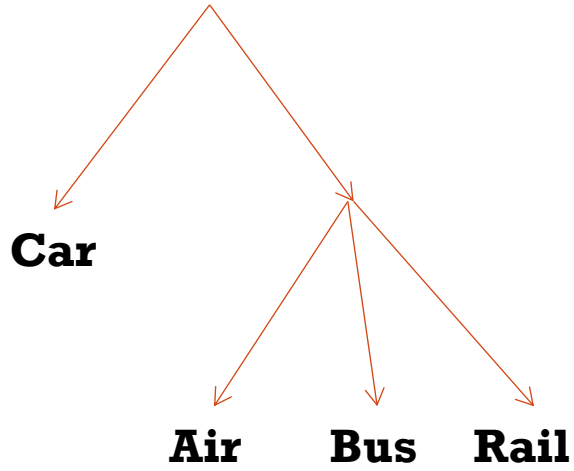
- Assuming each pair of the error terms in ε_{Air}, ε_{Bus}, and ε_{Rail} are independent we can compute $Var(\varepsilon_{cc})$ as
  - $\frac{\Pi^2}{6} - \frac{\Pi^2\theta^2}{6}$
- Correlation (U_A, U_B) = Correlation (U_A, U_R) = Correlation (U_B, U_R)
- $\text{Correlation}(a,b) = \frac{\text{covariance}(a,b)}{[\text{var}(a) * \text{var}(b)]^{1/2}}$

FORMULATION

- In our case, covariance(U_A,U_B) = Var(ε_{cc}) ;
- Var(U_A) = Var(U_B)= Var(U_R) = $\frac{\Pi^2}{6}$
- So Correlation (U_A,U_B) = Correlation (U_A,U_R) = Correlation (U_B,U_R)
  - = { $\frac{\Pi^2}{6}$ - $\frac{\Pi^2\theta^2}{6}$ } / $\frac{\Pi^2}{6}$ = (1- θ^2)
- Correlation is (1- θ^2)
- Hence when we test our hypothesis (which is to see if correlation exists), we do not test if θ is different from 0, but if θ is different from 1
  - Null hypothesis is θ=1

FORMULATION

- Now how do we generate the probability expressions
- Now lets consider the nest
- $P_{air}|cc = \frac{\exp(\frac{V_A}{\theta})}{\exp(\frac{V_A}{\theta})+\exp(\frac{V_B}{\theta})+\exp(\frac{V_R}{\theta})}$



- Now when we need to generate the probability for the car or cc we somehow need to compute a net utility for the cc
- Now the choice between car and cc is determined as Ucar > Max(U_A,U_B,U_R)
- For a gumbel distribution G(V_1,θ), G(V_2,θ), G(V_3,θ)
- Max (V1, V2, V3) = G[θln(exp $\left(\frac{V_1}{\theta}\right)$ + exp $\left(\frac{V_2}{\theta}\right)$ + exp$(\frac{V_3}{\theta})$), θ)

FORMULATION

- In our case
- Max (U1, U2, U3)
- $= \theta\{ \ln(\exp\left(\frac{V_A}{\theta}\right) + \exp\left(\frac{V_B}{\theta}\right) + \exp(\frac{V_R}{\theta}))\} + \varepsilon^*$
- This is effectively the composite nest utility
- $\Gamma = \{ \ln(\exp\left(\frac{V_A}{\theta}\right) + \exp\left(\frac{V_B}{\theta}\right) + \exp(\frac{V_R}{\theta}))\}$
- P_{Car} = Prob [U_c > Max (U_A, U_B, U_R)]
- = Prob [$V_c + \varepsilon_{Car} >$ Max (U_A, U_B, U_R)]
- = Prob [$V_c + \varepsilon_{Car} > \theta\Gamma + \varepsilon^* + \varepsilon_{cc}$]

$G(0,1)$

FORMULATION

- $P_{Car} = \frac{\exp(V_C)}{\exp(V_C)+\exp(\theta\Gamma)}$
- $P_{cc} = \frac{\exp(\theta\Gamma)}{\exp(V_C)+\exp(\theta\Gamma)}$
  - θ – log–sum parameter
  - Γ – log-sum variable
- $P_{air}|cc = \frac{\exp(\frac{V_A}{\theta})}{\exp(\frac{V_A}{\theta})+\exp(\frac{V_B}{\theta})+\exp(\frac{V_B}{\theta})}$

- $P_{air} = P_{air}|cc * P_{cc}$
- Similar to P_{Bus}, P_{Rail}
- To get MNL from NL set $\theta = 1$
- Test it now

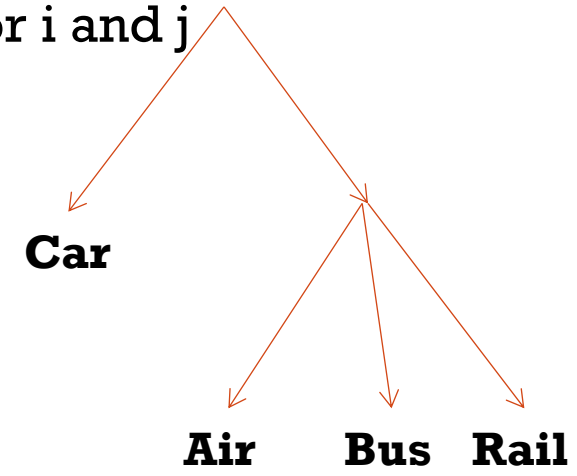
FORMULATION

- $P_{air} = P_{air|cc} * P_{cc}$
- $\theta = 1$
- $= \frac{\exp(\frac{V_A}{\theta})}{\exp(\frac{V_A}{\theta})+\exp(\frac{V_B}{\theta})+\exp(\frac{V_B}{\theta})} * \frac{\exp(\theta\Gamma)}{\exp(V_C)+\exp(\theta\Gamma)}$

- $= \frac{\exp(V_A)}{\exp(V_A)+\exp(V_B)+\exp(V_R)} * \frac{\exp(\{\ln(\exp(V_A)+\exp(V_B)+\exp(VR)\})}{\exp(V_C)+\exp(\{\ln(\exp(\frac{V_A}{\theta})+\exp(\frac{V_B}{\theta})+\exp(\frac{V_B}{\theta}))\})}$
- $= \frac{\exp(VA)}{\exp(V_C)+\exp(V_A)+\exp(V_B)+\exp(VR)}$ -> MNL

IIA PROPERTY

- **MNL** - Consider the ratio of alternative probabilities for i and j.
- $P_i / P_j = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)} \Big/ \frac{\exp(Vj)}{\sum_{\forall j} \exp(Vj)} = \frac{\exp(Vi)}{\exp(Vj)} = \exp(V_i - V_j)$
- **NL** - Consider the ratio of alternative probabilities for i and j
- $P_A/P_R = [P_{air}|cc * P_{cc}] / [P_{rail}|cc * P_{cc}]$
- $= [P_{air}|cc] / [P_{rail}|cc]$
  - Simplifies exactly like the MNL
- $P_A/P_C = [P_{air}|cc * P_{cc}] / [P_{car}]$
  - Does not simplify
- Alternatives within the nest still act as if they are part of the MNL structure
- Alternatives outside the next exhibit different substitution patterns

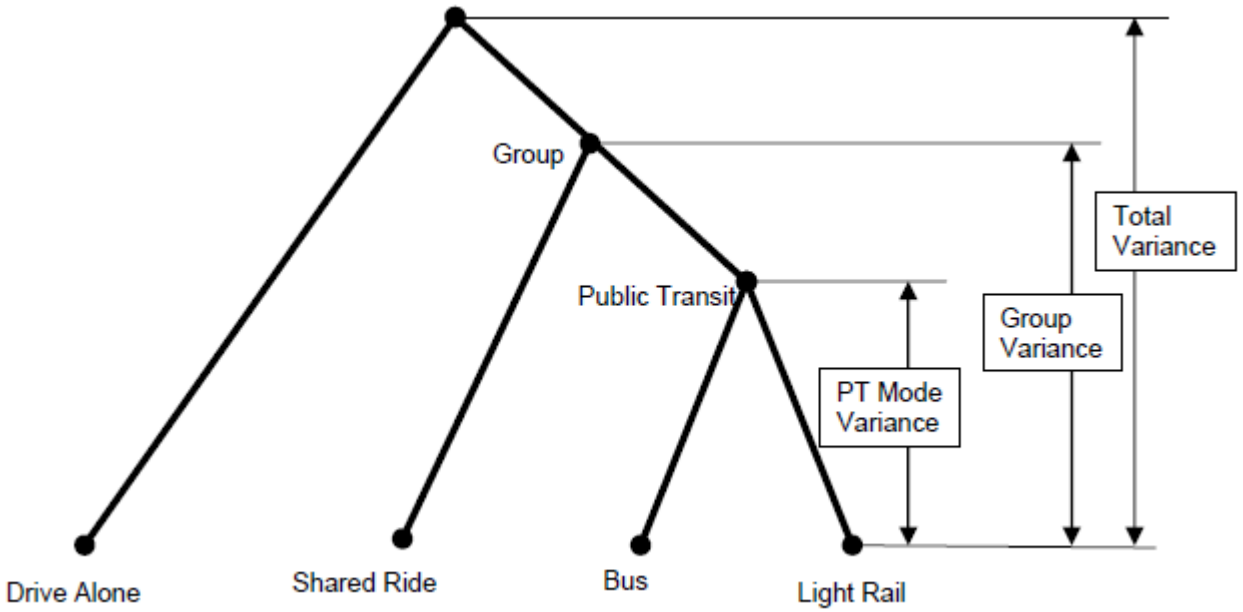


ELASTICITY

- Self-Elasticity
  - For Non-nested alternatives (same as MNL)
    - $\beta_k[1\text{-}P_{in}]\ x_{ik}$
  - For Nested alternatives
    - $\beta_k\{[1\text{-}P_{in}] + \frac{1-\theta}{\theta}(1\text{-}\ P_{in}|N)\}\ x_{ik}$

- Cross-Elasticity
  - Effect on Non-nested alts for change in Non-nested alts
    - $-\beta_k[P_{in}]\ x_{ik}$
  - Effect on Non-nested alts for change in Nested alts
    - $-\beta_k[P_{in}]\ x_{ik}$
  - Effect on Nested alts for change in Non-nested alts
    - $-\beta_k[P_{in}]\ x_{ik}$
  - Effect on Nested alts for change in Nested alts
    - $-\beta_k\{[P_{in}] + \frac{1-\theta}{\theta}(P_{in}|N)\}\ x_{ik}$

ANOTHER EXAMPLE

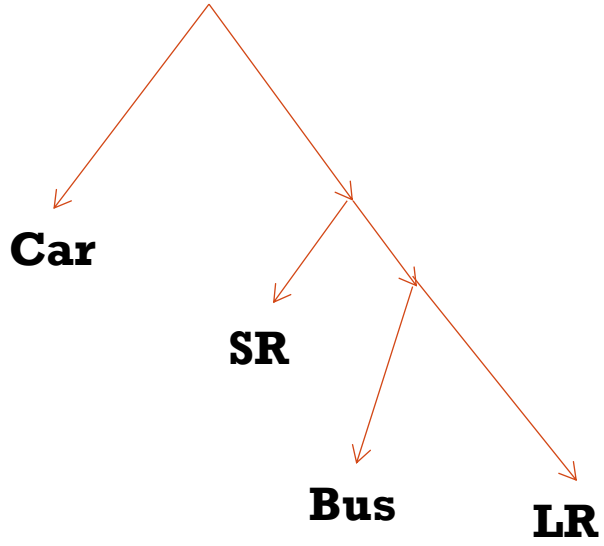


$$U_{DA} = V_{DA} + \varepsilon_{DA}$$
$$U_{SR} = V_{GRP} + V_{SR} + \varepsilon_{GRP} + \varepsilon_{SR}$$
$$U_{Bus} = V_{GRP} + V_{PT} + V_{Bus} + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{Bus}$$
$$U_{LTR} = V_{GRP} + V_{PT} + V_{LTR} + \varepsilon_{GRP} + \varepsilon_{PT} + \varepsilon_{LTR}$$

G(0,1)

ANOTHER EXAMPLE

- ε_{SR} - G(0,θ)
- $\varepsilon_{B,}$ ε_{LR} G(0,η)
- 0< η< θ <1



- Same approach as the previous case to generate the probabilities

- Read Section 8.3 of Bhat and Koppelman 2006 for exact probability expressions
  - Koppelman, F.S. and Bhat, C., 2006. A self instructing course in mode choice modeling: multinomial and nested logit models

REMARKS

- Multinomial logit model needs to be estimated first
- Account for systematic effects
- Then attempt to incorporate correlation
- Estimate NL model and if $\theta = 1$ it indicates MNL is good enough
- In case $\theta > 1$ then the formulation is not consistent with utility framework

MIXED MULTINOMIAL LOGIT MODEL

- Intuition
  - In the MNL model we estimate a single parameter to determine the influence of an exogenous variable on the choice process
    - For example, we claim that the influence of income is the same for the entire population. However, based on whether the respondent is lavish or conservative with money the influence varies. But, we cannot accommodate for such taste variations in the MNL
- In a MMNL model we allow the coefficients to vary across different individuals
- We accommodate for correlation across the error terms for different alternatives (relaxing the independence assumption)
- We incorporate different error variances (relaxing the identical assumption)

MIXED MULTINOMIAL LOGIT MODEL

- The MMNL model involves the integration of the MNL formulation over the unobserved parameters

$$P_{qi}(\theta) = \int_{-\infty}^{+\infty} L_{qi}(\beta) f(\beta \mid \theta)\, d(\beta),$$

where

$$L_{qi}(\beta) = \frac{e^{\beta' x_{qi}}}{\sum_j e^{\beta' x_{qi}}}$$

- The MMNL model can be formulated from two unique but equivalent formulations:
  - Error components
  - Random coefficients

ERROR COMPONENTS

- Consider utility of person *q* for alternative *i*

$$U_{qi} = \gamma' y_{qi} + \zeta_{qi}$$

$$= \gamma' y_{qi} + \mu' z_{qi} + \varepsilon_{qi}$$

Systematic component

Random component

Induces heteroscedasticity and correlation across error based on how we specify Z_{qi}

Example of Z_{qi}

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

RANDOM COMPONENTS

- Consider utility of person *q* for alternative *i*

$$U_{qi} = \beta'_q x_{qi} + \varepsilon_{qi}$$

Random coefficients

$$\beta_{qk} \sim N(\mu_k, \sigma_k)$$

$$U_{qi} = \alpha_{qi} + \sum_{k=1}^{K} \beta_{qk} x_{qik} + \varepsilon_{qi}$$

$$\beta_{qk} = \mu_k + \sigma_k s_{qk}$$

Fixed coefficients

$$V_{qi} = \alpha_{qi} + \sum_k \mu_k x_{qik}$$

$$P_{qi} = \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi} + \sum_k \sigma_k s_{qk} x_{qik}}}{\sum_j e^{V_{qj} + \sum_k \sigma_k s_{qk} x_{qjk}}} d\Phi(s_{q1}) d\Phi(s_{q2}) \ldots d\Phi(s_{qK}) \right\}$$

MIXED MULTINOMIAL LOGIT MODEL

- Estimation

$$P_{qi} = \left\{ \int_{s_{q1}=-\infty}^{s_{q1}=+\infty} \int_{s_{q2}=-\infty}^{s_{q2}=+\infty} \cdots \int_{s_{qK}=-\infty}^{s_{qK}=+\infty} \frac{e^{V_{qi}+\sum_k \sigma_k s_{qk} x_{qik}}}{\sum_j e^{V_{qj}+\sum_k \sigma_k s_{qk} x_{qjk}}} d\Phi(s_{q1}) d\Phi(s_{q2}) \ldots d\Phi(s_{qK}) \right\}$$

- The probabilities are approximated through simulation
- For any given value of σ (1,2,..K), draw a S_q (1,2,..K) and compute P_{qi}. Repeat this multiple times and average the P_{qi}.

$$\tilde{P}_{qi} = \frac{1}{R} \sum_1^R P_{qi}$$

MIXED MULTINOMIAL LOGIT MODEL

- Log-likelihood function

$$SL = \sum_{1}^{Q} \sum_{i=1}^{I} d_{qi} \tilde{P}_{qi}$$

d_{qi} is 1 if q chose i; else 0

- Now that we have the LL function, we undertake Maximum Likelihood to get our estimates!

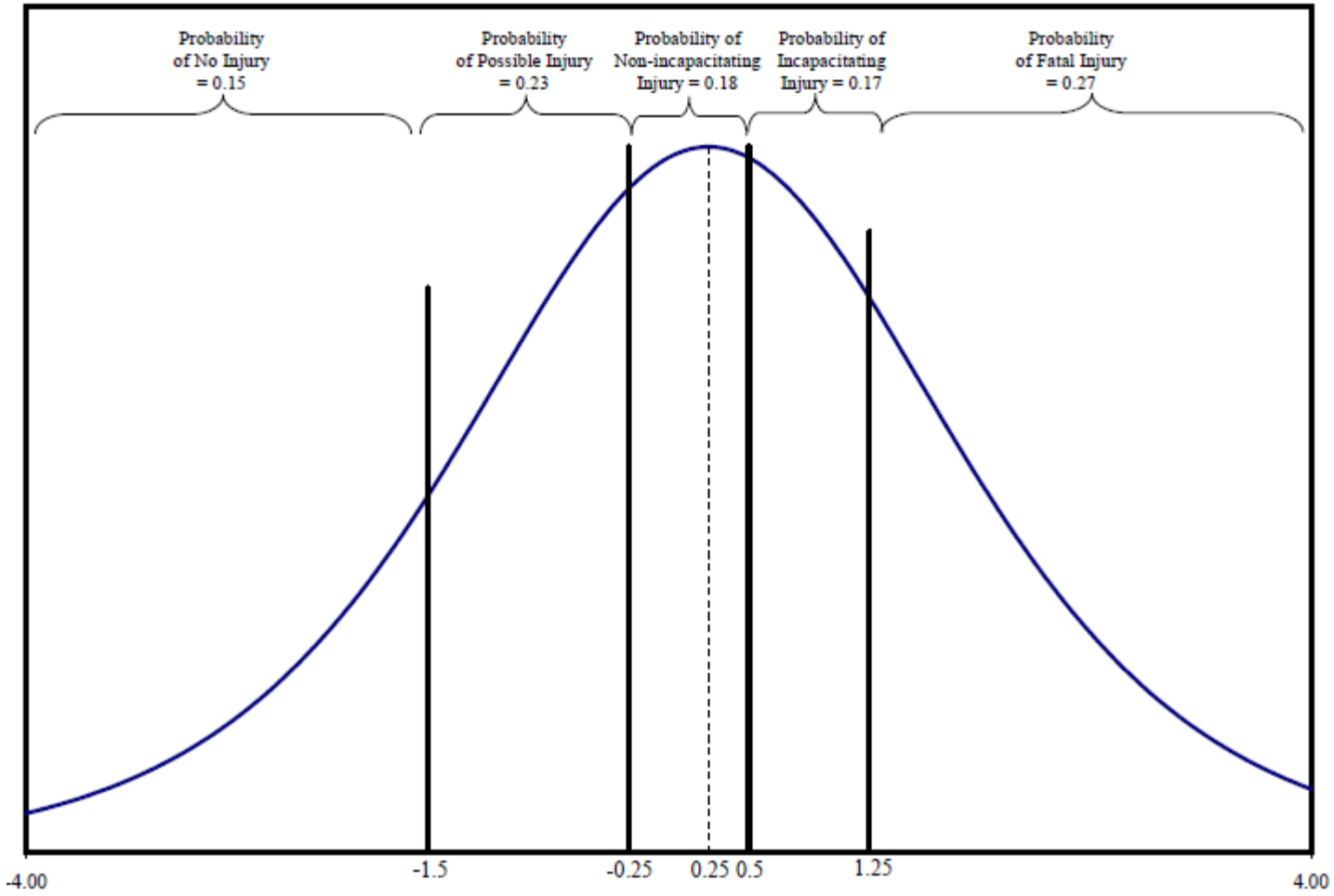
ADVANCED OR MODELS

ADVANCED OR MODELS

- As we discussed earlier, the OR models do not allow for alternative specific effects of various exogenous variables
- Lets consider the following example
- Non-motorist injury severity due to traffic collisions is reported as a five level ordinal variable
  - No injury
  - Possible Injury
  - Non-incapacitating injury
  - Incapacitating injury
  - Fatal injury
- Now we estimated a model and found that
  - A motorist being intoxicated has a coefficient of 0.25 (+ive so increases probability of fatal injury)
  - Coefficient for being hit head-on versus sideways is 0.25
  - Thresholds ψ_i = (-1.5, -0.25, 0.5, 1.25)
  - Let us assume these are the only variables affecting injury severity

ADVANCED OR MODELS

- Now consider two crashes
  - involving a drunk motorist and sideways crash
  - Involving a sober motorist and head-on crash
- Based on our OL model
  - Latent propensity for both crashes is 0.25
- So the probability will be (for standard logistic)
  - No injury (0.15)
  - Possible injury (0.23)
  - Non-incapacitating injury (0.18)
  - Incapacitating injury (0.17)
  - Fatal injury (0.27).



Source: *Eluru, N., C.R. Bhat, and D.A. Hensher (2008), "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes ," Accident Analysis and Prevention, Vol. 40, No. 3, pp. 1033-1054*

ORDERED LOGIT MODELS

- Standard ordered response model

$$y_q^* = \beta' x_q + \varepsilon_q$$

$$\text{where } y_q = k, \text{ if } \psi_{k-1} < y_q^* < \psi_k, \forall k = 1, 2..K$$

  - K represents the alternatives
  - y_q corresponds to the latent propensity for DM q
  - x_q is an (L x 1)-column vector of attributes (excluding a constant) associated with the DM q
  - β is a corresponding (L x 1)-column vector of variable effects
  - ψ_k corresponds to thresholds (ψ_0 = -∞ and ψ_K = + ∞)
  - ε_q represents the idiosyncratic error term distributed as a logistic

MIXED ORDERED LOGIT MODELS

- In the MGORL model we allow the thresholds to vary across DMs based on the variables
- $y_q^* = (\boldsymbol{\beta} + \boldsymbol{\alpha}_q)\boldsymbol{X}_q + \varepsilon_q$
  - $\tau_{q,k} = \tau_{q,k-1} + exp[(\boldsymbol{\delta_j} + \boldsymbol{\gamma_{q,k}})\, \boldsymbol{Z}_{q,k}]$
- $Pr(y_q = k|\boldsymbol{\alpha}_q, \boldsymbol{\gamma}_{qk}) = \Lambda[(\boldsymbol{\delta_k} + \boldsymbol{\gamma_{q,k}})\, \boldsymbol{Z_{q,k}} - (\boldsymbol{\beta} + \boldsymbol{\alpha}_q)\boldsymbol{X_q}] - \Lambda[(\boldsymbol{\delta_{k-1}} + \boldsymbol{\gamma}_{q,k-1})\, \boldsymbol{Z_{q,k}} - (\boldsymbol{\beta} + \boldsymbol{\alpha_q})\boldsymbol{X_q}]$
- $P_{qk} = \int_{\boldsymbol{\alpha_q}, \boldsymbol{\gamma_{qk}}} [Pr(y_q = k|\boldsymbol{\alpha_q}, \boldsymbol{\gamma_{qk}})] * \boldsymbol{dF}(\boldsymbol{\alpha_q}, \boldsymbol{\gamma}_{qk})\boldsymbol{d}(\boldsymbol{\alpha_q}, \boldsymbol{\gamma_{qk}})$
- Simulation approach is same as the MMNL

EXAMINING PEDESTRIAN AND BICYCLIST INJURY SEVERITY LEVEL IN TRAFFIC CRASHES – A MIXED GENERALIZED ORDERED RESPONSE MODEL

SOURCE

- Eluru, N., C.R. Bhat, and D.A. Hensher (2008), "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes", Accident Analysis and Prevention, Vol. 40, No.3, pp. 1033-1054
- *Listed in the Top 50 papers published in Accident Analysis Prevention* - Zou, X., Vu, H.L. and Huang, H., 2020. Fifty years of Accident Analysis & Prevention: a bibliometric and scientometric overview. Accident Analysis & Prevention, 144, p.105568.

MOTIVATION

- Increased personal vehicle dependency in the US leads to
  - Increasing traffic congestion
  - Air quality problems
- Metropolitan organizations encourage non-motorized travel
  - Walking and bicycling for short distance trips
- Safety of non-motorists (pedestrians and bicyclists) in the US
  - Worse record in the US compared to other developed countries
  - Controlling for exposure in terms of miles traveled, US pedestrians are 3 times likely to get killed compared to German pedestrians, and over 6 times more likely compared to Dutch pedestrians (the corresponding numbers for cyclists are 2 and 3)

MOTIVATION

- In terms of absolute numbers, in 2005
  - 4881pedestrian and 784 bicyclist fatalities
  - 110,000 non-motorists are injured
- To put these numbers into perspective
  - A non-motorist is killed every 93 minutes and one is injured every 5 minutes in traffic accidents in the US
- High risk of non-motorists has attracted a lot of attention in the past decade
- Researchers examined the crashes involving non-motorists to:
  - Improve motorized vehicle and roadway design,
  - Enhance control strategies at conflict locations
  - Design good bicycle and pedestrian facilities
  - Formulate driver and non-motorized user education programs

MOTIVATION

- The host of factors that could potentially influence non-motorist injury severity include
  - Pedestrian/bicyclist characteristics (such as age, gender, helmet use, alcohol consumption)
  - Motorized vehicle driver characteristics (such as state of soberness and age)
  - Motorized vehicle attributes (such as vehicle type and speed)
  - Roadway characteristics (such as speed limit and whether the highway is divided or not)
  - Environmental factors (such as time of day, day of week, and weather conditions)
  - Crash characteristics (such as the direction of impact and motorist/non-motorist maneuver type at impact).

EARLIER RESEARCH

- A vastly researched area in the recent decade
- Research classified into two categories
  - Descriptive analyses at an aggregate level
    - A common association across the entire sample is arrived at through frequency analysis or cross-tabulation
  - Multivariate analyses at individual level of accidents
    - A host of factors influencing non-motorist injury severity are examined
- Remarks on earlier studies
  - The more recent studies have employed multivariate analyses
  - In cases where a binary dependent variable is employed (fatal vs non-fatal) logistic regression methods are predominant

EARLIER RESEARCH

- Remarks on earlier studies
  - In cases with ordered injury categories (such as property damage only, no visible injury but pain, non-incapacitating injury, incapacitating injury, and fatal injury) an ordered response model is employed
  - Studies have examined pedestrian or bicyclist injury severity separately
    - It is important from a policy perspective to compare the similarities and differences in the factors, and the magnitude of the impact of factors, affecting injury severity between the two non-motorist user groups
  - Earlier studies have very often, failed to recognize the need to consider motorist vehicle characteristics in the analysis

EARLIER RESEARCH

- Important findings
  - Pedestrians
    - Male, intoxicated, very young and elderly are prone to severe injuries
    - Alcohol-intoxicated driver, non-sedan and high speed vehicles cause severe injuries
  - Bicyclists
    - Similar to pedestrians
    - Accidents at high speed limit, low traffic volume and curved/non-flat roadway locations
    - Conditions of darkness with no lighting, in inclement weather (fog, rain and snow) and accidents in the morning peak period lead to severe injuries

EARLIER RESEARCH

- Current research in perspective
  - Employ a generalized version of the ordered logit model
  - Undertake the analysis for pedestrians and bicyclists
  - Consider the factors from all the six categories identified earlier
  - Allow for the presence of unobserved attributes to influence injury severity
    - For instance, the slower reaction time of being intoxicated may be exacerbated by the use of a walkman. But accident reports may not record or may miss information on walkman use and so walkman use may be unobserved
  - To summarize, develop a generalized model with a comprehensive set of variable to examine injury severity determinants

NOTATION

- Standard ordered response model

$$y_q^* = \beta' x_q + \varepsilon_q$$

$$\text{where } y_q = k, \text{ if } \psi_{k-1} < y_q^* < \psi_k, \forall k = 1, 2..K$$

K represents the number of injury categories

y_q corresponds to the latent injury risk propensity for non-motorist q in the crash she or he was involved in

x_q is an (L x 1)-column vector of attributes (excluding a constant) associated with the non-motorist, driver, vehicle, roadway, environment, and crash characteristics of the crash involving individual q

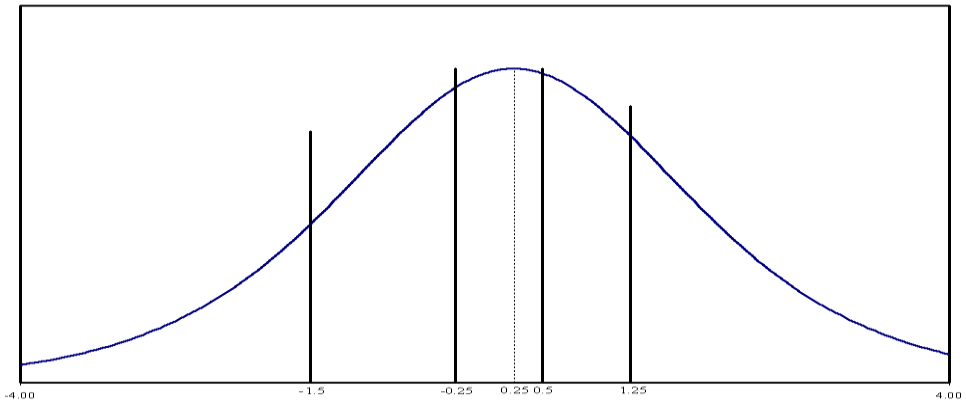
β is a corresponding (L x 1)-column vector of variable effects

ψ_k corresponds to thresholds ($\psi_0 = -\infty$ and $\psi_K = +\infty$)

ε_q represents the idiosyncratic error term distributed as a logistic

EXAMPLE OF AN ORDERED RESPONSE LOGIT MODEL (ORL)

- For K = 5 injury categories, $\psi_0 = -\infty$ $\psi_1 = -1.5$, $\psi_2 = -0.25$, $\psi_3 = 0.5$, $\psi_4 = 1.25$ and $\psi_K = +\infty$
- Propensity ($\beta'x$) = 0.25
- Probability of injury severity in a particular category is the area under the curve between the corresponding thresholds



- Potential limitations of ORL model
  - The thresholds remain constant across individual accidents

MIXED GENERALIZED ORDERED RESPONSE LOGIT (MGORL) MODEL

- In the MGORL model we allow the thresholds to vary across individual accidents based on the variables

$$y_q^* = \beta_q' x_q + \varepsilon_q,\ y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}$$

Next, we adopt a specific parametric form for the thresholds to guarantee the ordering conditions $(-\infty < \psi_{q,1} < \psi_{q,2} < \ldots < \psi_{q,K-1} < \infty)$ for each crash q.

To do so, we write:

$$\psi_{q,k} = \psi_{q,k-1} + \exp(\alpha_{qk} + \gamma'_{qk} z_{qk}),$$

DATA SOURCE

- 2004 General Estimates System (GES)
  - National Highway Traffic Safety Administration's National Center for Statistics and Analysis
  - Data compiled from a sample of police-reported accidents
  - The injury severity is collected on a five point ordinal scale: (1) No injury, (2) Possible injury, (3) Non-incapacitating injury, (4) Incapacitating injury, and (5) Fatal injury
  - Categories 1 and 2 are collapsed into a single category
- Sample preparation
  - Accidents involving pedestrians and bicyclists
  - Accidents involving a single vehicle and single non-motorist are chosen

DATA SOURCE

- Sample characteristics
- Distribution of non-motorist injury severity by non-motorist type

Injury severity category	Pedestrian	Bicyclist	All Non-motorists
No injury	135 (7.8%)	89 (7.3%)	224 (7.6%)
Non-incapacitating injury	951 (55.3%)	863 (70.6%)	1814 (61.6%)
Incapacitating injury	541 (31.4%)	250 (20.4%)	791 (26.9%)
Fatal injury	94 (5.5%)	21 (1.7%)	115 (3.9%)
Total	1223 (100.0%)	1721 (100.0%)	2944 (100.0%)

DATA SOURCE

- Distribution of Non-Motorist Injury Severity by Non-Motorist Alcohol Intoxication

Injury severity category	Non-motorist was alcohol intoxicated?		All Non-motorists
	No	Yes	
No injury	217 (8.0%)	7 (2.8%)	224 (7.6%)
Non-incapacitating injury	1688 (62.6%)	126 (51.2%)	1814 (61.6%)
Incapacitating injury	699 (25.9%)	92 (37.4%)	791 (26.9%)
Fatal injury	94 (3.5%)	21 (8.5%)	115 (3.9%)
Total	2698 (100.0%)	246 (100.0%)	2944 (100.0%)

EMPIRICAL ANALYSIS

- Results based on the estimation of the MGORL model for variables from all the six categories of variables identified earlier

- Non-motorist characteristics
  - Age is an important consideration. Non-motorists >60 years are prone to severe (even fatal) injuries
  - Gender effect is marginal
  - Alcohol intoxication increases likelihood of injury
  - Pedestrians are more likely to be severely injured

EMPIRICAL ANALYSIS

- Motorist characteristics
  - Alcohol intoxication leads to higher loading of severe injuries
- Motorized vehicle attributes
  - Non-sedan vehicle increases potential injury to non-motorist
- Roadway characteristics
  - Crashes on roads with high speed limits result in severe crashes
  - Signalized intersection reduce the severity of a crash for non-motorist

EMPIRICAL ANALYSIS

- Environment factors
  - Crashes occurring between 6PM – 12AM result in more severe injuries
  - Interestingly, presence of snow reduces the probability of fatality

- Crash characteristics
  - Direction of crash impacts the injury severity
    - Frontal impacts result in more severe crashes

RESULTS

Variables	Latent Propensity	Threshold between Non-incapacitating and Incapacitating injury	Threshold between Incapacitating and Fatal injury
Constant	1.846 (12.94)	1.305 (36.26)	1.645 (11.49)
Non-motorist Characteristics			
Pedestrian (Bicyclist is the base)	---	-0.103 (-2.67)	---
Male	0.159 (1.85)	---	---
Age Variables (age ≤ 60 years is base)			
> 60 years	0.667 (5.26)	---	-0.536 (-4.61)
Under the influence of alcohol	0.455 (3.47)	---	---
Motorized Vehicle Driver Characteristics			
Under the influence of alcohol	0.837 (2.14)	0.271 (2.87)	-0.250 (-1.53)
Motorized Vehicle Attributes			
Sports utility vehicle	0.364 (3.15)	---	---
Pick-up truck	---	-0.070 (-2.18)	-0.197 (-1.98)
Van	---	---	-0.237 (-1.70)

RESULTS

Variables	Latent Propensity	Threshold between Non-incapacitating and Incapacitating injury	Threshold between Incapac and Fatal injury
Roadway Design Characteristics			
Speed Limit			
25-50mph	0.218 (1.97)	---	-0.225 (-2.01)
>50 mph	0.605 (3.06)	---	-0.679 (-3.93)
Speed limit > 25mph * pedestrian	---	-0.117 (-2.61)	---
Accident Location			
Signalized Intersection	-0.300 (-3.32)	---	0.387 (3.43)
Environmental Factors			
6pm - 12am	0.297 (3.43)	---	-0.352 (-3.82)
12am - 6am	---	-0.304 (-4.66)	-0.365 (-2.59)
Snow	---	---	0.538 (1.60)
Crash Characteristics			
Direction of Impact (sideways impact is the base)			
Frontal Impact	0.447 (3.20)	0.072 (1.64)	-0.226 (-2.38)
Other directions of impact	-0.734 (-2.91)	---	-0.603 (-2.23)

VALIDATION EXERCISE

- Comparing the proposed (MGORL) model vs standard (ORL) model

Injury Categories/ Measures of fit	Pedestrians			Bicyclists		
	Actual shares	ORL predictions	MGORL predictions	Actual shares	ORL predictions	MGORL predictions
No injury	7.84	6.04	7.44	7.28	9.89	7.93
Non-incapacitating injury	55.26	57.70	55.55	70.56	65.90	70.07
Incapacitating injury	31.44	31.38	31.73	20.44	21.59	20.28
Fatal injury	5.46	4.94	5.29	1.72	2.62	1.72
Number of observations	1721	1721	1721	1223	1223	1223
Root mean square error (RMSE)	---	1.54	0.30	---	2.77	0.42
Mean absolute percentage error (MAPE)	---	9.28	2.46	---	25.14	2.62

IMPLICATIONS FROM RESEARCH

- Education and training
  - The results reinforce the need to educate non-motorists and motorists of the risks of driving under influence. It is necessary to underscore that alcohol combined with night driving is deadly
  - Encouraging non-motorists to wear "reflector" gear to improve visibility
- Traffic regulation and control
  - Signs need to be posted to communicate to non-motorists information regarding heavy traffic on roadways
  - Restricting speed limits to < 25 mph on roadways with heavy pedestrian and bicycle traffic
  - Good street lighting and illumination, and additional traffic signal installation might alleviate non-motorist injury severity
- Planning and design of pedestrian/bicyclist facilities
  - On roadways with high speed limits bicycle facility need to be separated from roadway. Further bicycle facilities need to be chosen based on roadway speed limits, vehicular mix and presence of lighting

CONCLUSIONS

- The current study addresses the safety of non-motorists
- An advanced econometric framework to address the ordinal category of the reported injury severity is developed. The proposed model generalizes the standard ORL model
- The MGORL model developed is employed on 2004 General Estimates System (GES) database
- The standard ORL model employed produces inconsistent estimates
- It is very interesting to note that the general pattern and relative magnitude of elasticity effects of injury severity determinants are similar for pedestrians and bicyclists

CONCLUSIONS

- Pedestrians are more likely to be injured in the event of a crash
- The most important variables influencing the injury severity are:
  - Non-motorist age
  - Speed limit of the roadway
  - Location of the crash (if a signalized intersection or not)
  - Time of day (evening time being more riskier)
- Important implications for education and training, traffic regulation and control, and planning of pedestrian/bicycle facilities

MACHINE LEARNING APPROACHES

INTRODUCTION

- A branch of computer science
- Grew out of artificial intelligence
- Main idea: teaches computers how to solve a problem by example
  - We have a set of images of cars, we want to know if a new image of any random object will be a car
- Used for
  - Classification – Predict answers to yes/no questions
  - Regression – Predict real values
  - Clustering – Find patterns of similar objects

CLASSIFICATION

- Given: a *training set* of observations (e.g., labeled images) and a *test set* for evaluation only
- Each observation is represented by a set of numbers (features).

Intersection is represented as: [1 500 60 12 1 0 ....]

If it has a severe crash or not
Average traffic flow per hour
Average cycle length
Average queue length
Adaptive?
Right turn allowed?

Training set is from 2015 and before

Testing set is from 2016 to 2017

CLASSIFICATION

- Given: a *training set* of observations (e.g., labeled images) and a *test set* for evaluation only
- Each observation is represented by a set of numbers (features).

Intersection is represented as: [1 500 60 12 1 0 ....]

If it has a severe crash or not
Average traffic flow per hour
Average cycle length
Average queue length
Adaptive?
Right turn allowed?

Predict if a severe crash will happen in a random intersection in 2018

CLASSIFICATION

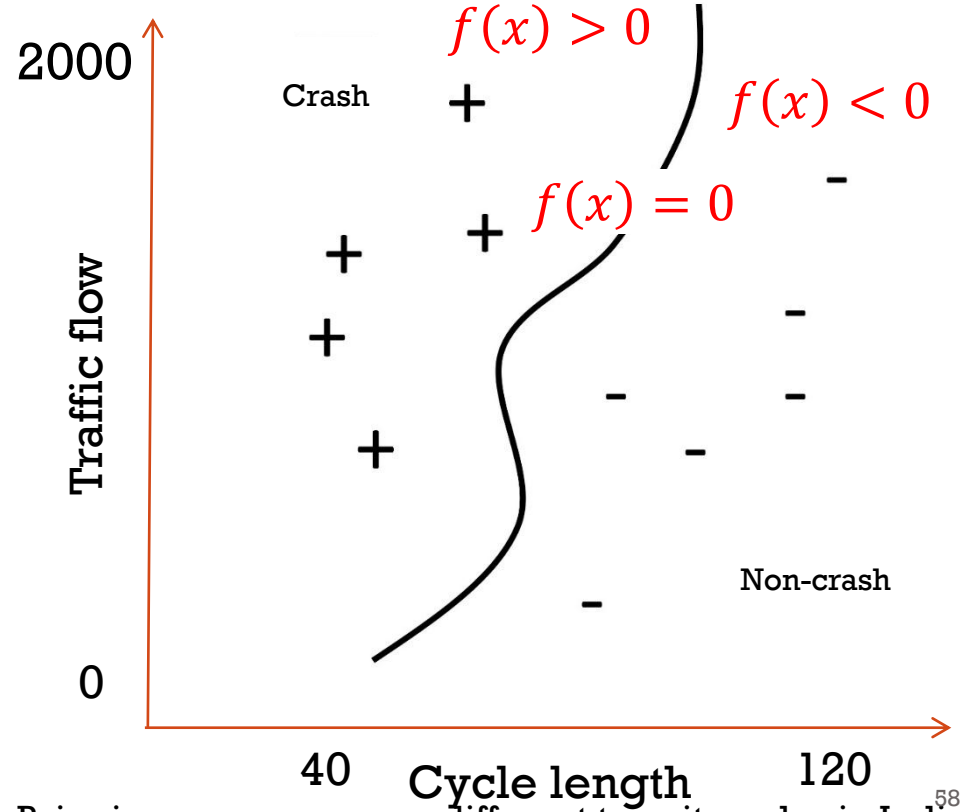
- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

Intersection: [1 500 60]

If it has a severe crash or not

Average traffic flow per hour

Average cycle length



$f(x)$ = function (traffic flow, cycle length)

Machine learning algorithm will create the function f

CLASSIFICATION

- Well-known classification algorithms
  - Logistic Regression
  - Decision Trees
  - Support Vector Machines
  - Random Forests
  - Neural Networks

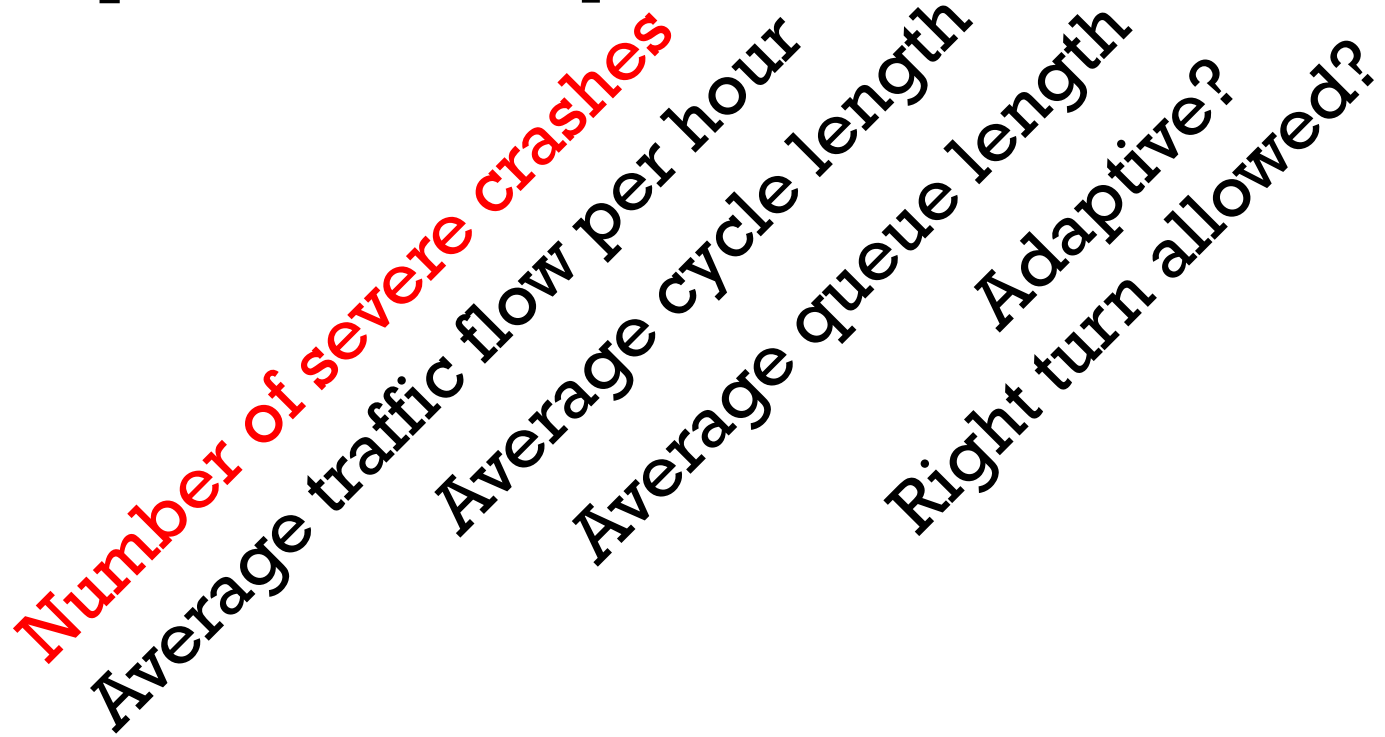
REGRESSION

- For predicting real-valued outcomes:
  - How many pedestrian crashes will occur in a given intersection?
  - How much traffic will move in a given freeway segment?
  - How many cars will park at a given time of the day?
  - How many people will ride bus from a given stop?

REGRESSION

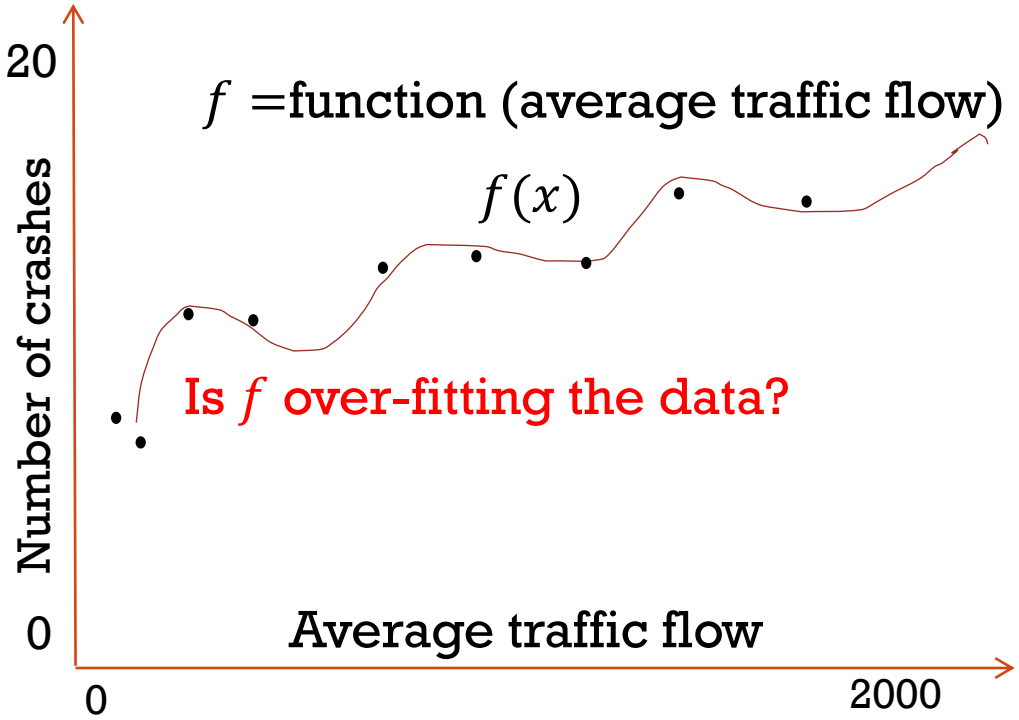
- Each observation is represented by a set of numbers.

Intersection is represented as: [5 500 60 12 1 0 ....]

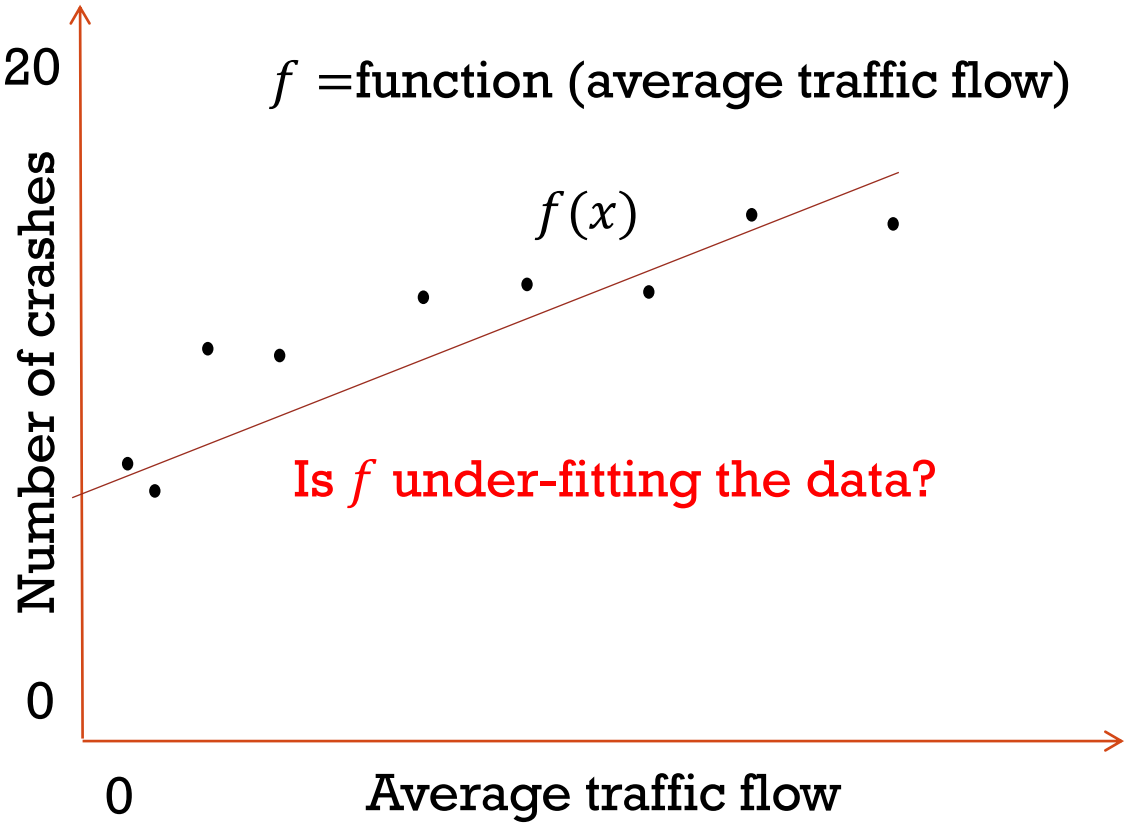


REGRESSION

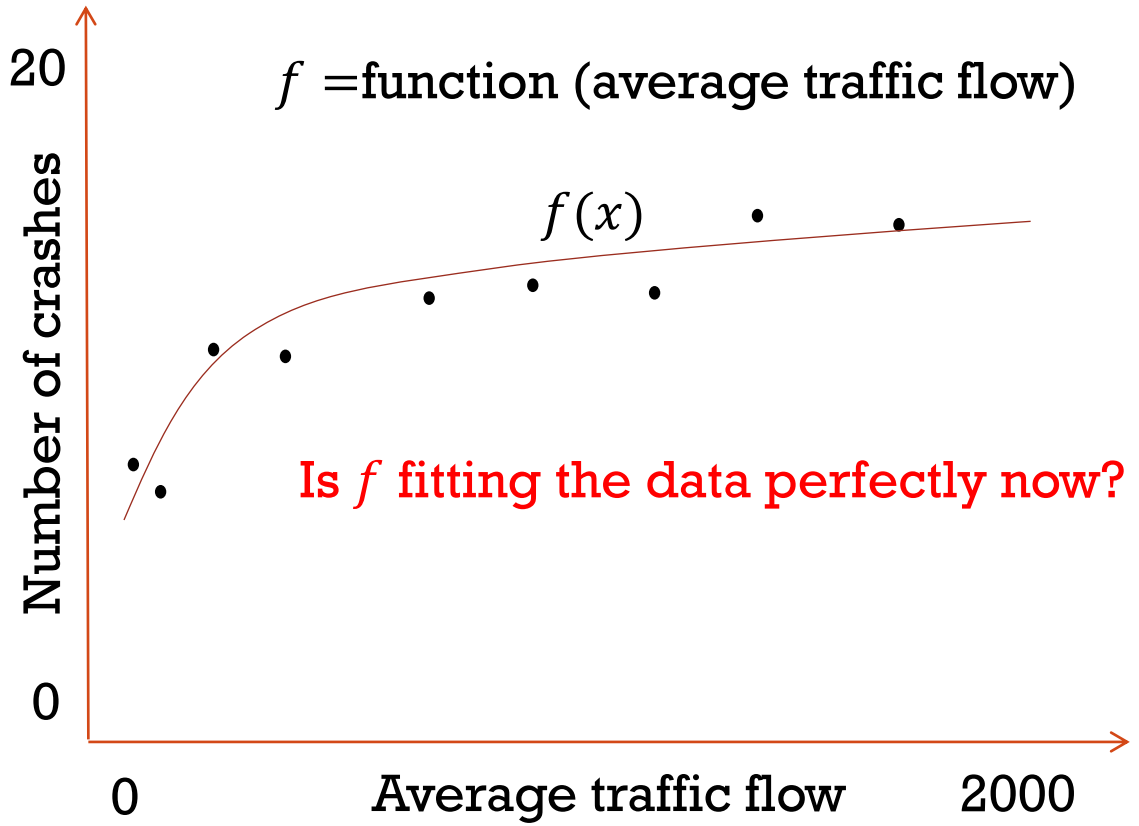
- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.



REGRESSION


20
0
Number of crashes
f =function (average traffic flow)
$f(x)$
Is f under-fitting the data?
0
Average traffic flow

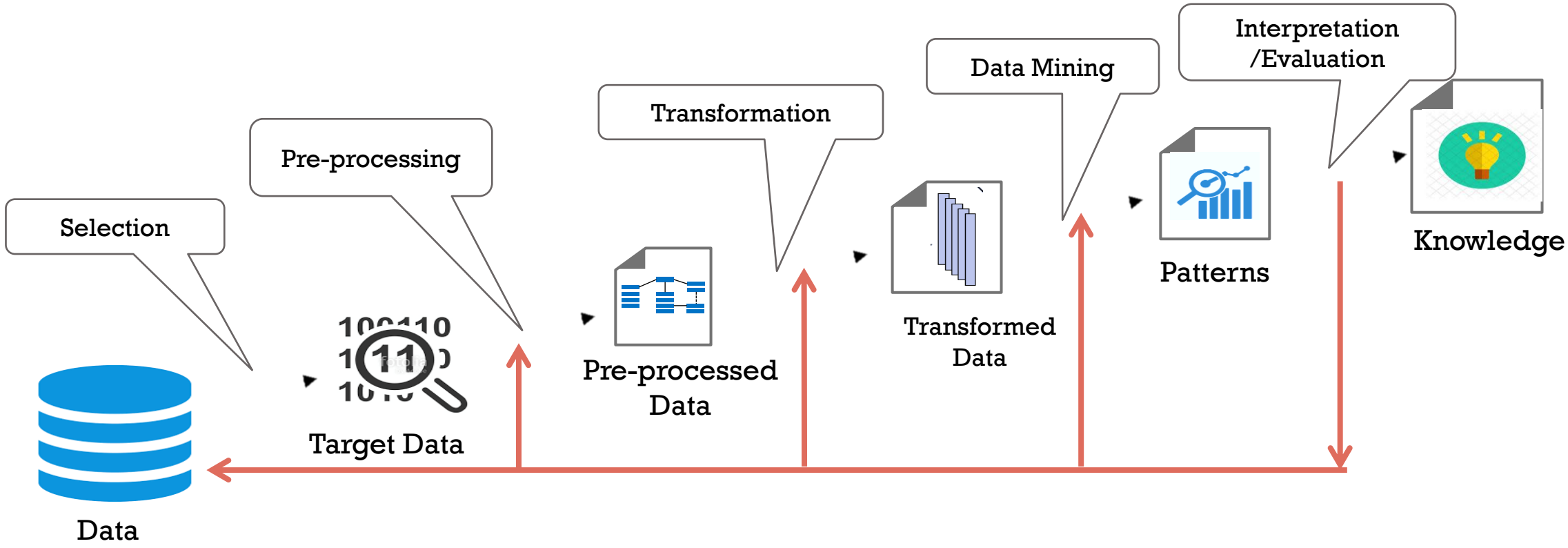
REGRESSION


20
0
Number of crashes
f =function (average traffic flow)
$f(x)$
Is f fitting the data perfectly now?
0
Average traffic flow
2000

SUPERVISED LEARNING

- "Supervised" means that the training data has ground truth labels to learn from.
  - Classification and Regression are supervised learning problems.
- (Supervised) classification often has +1 or -1 labels.
- (Supervised) regression has numerical labels.
- Supervised learning algorithms are much easier to evaluate than unsupervised ones, why?

KEY TAKE HOME MESSAGE

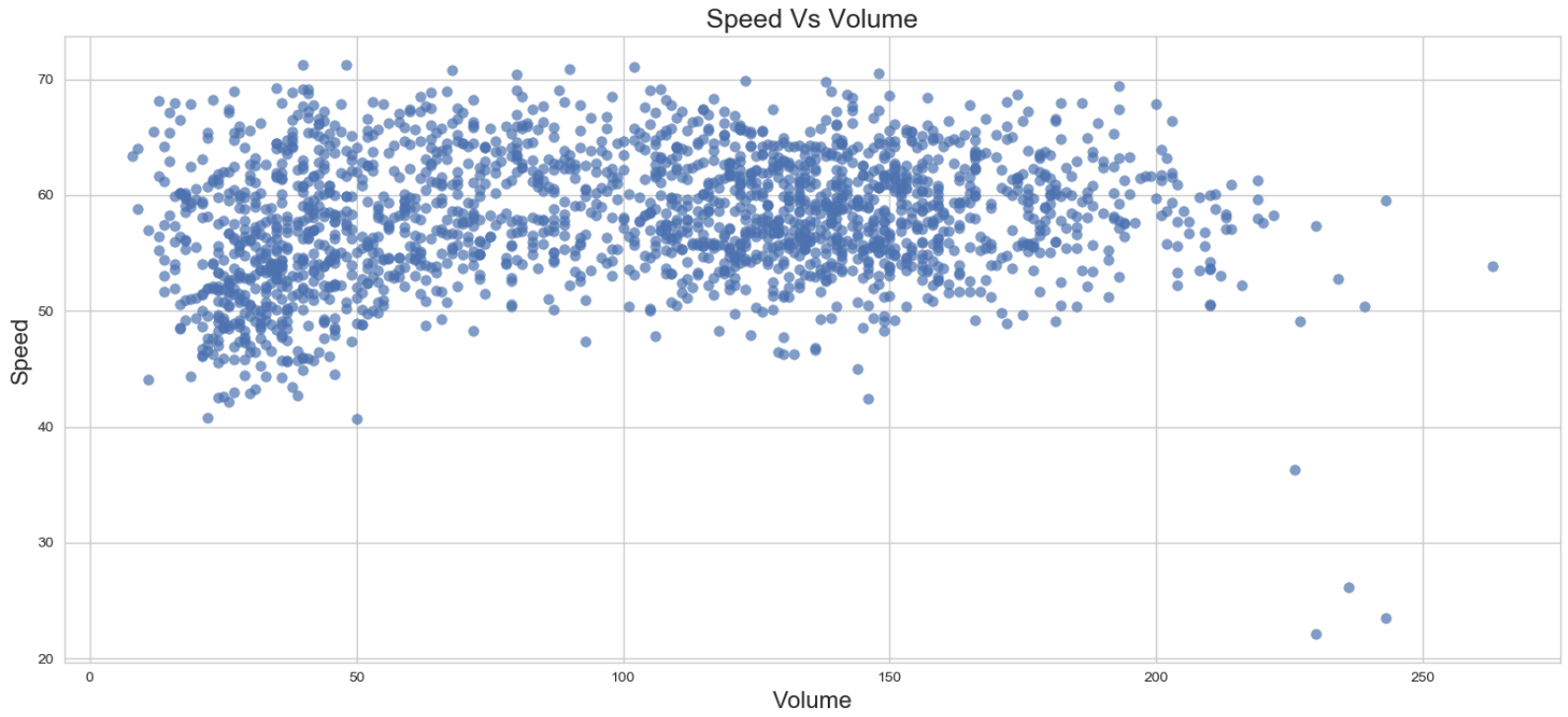


Based on "From Data Mining to Knowledge Discovery in Databases", AI Magazine, Vol 17, No. 3 (1996)
http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230

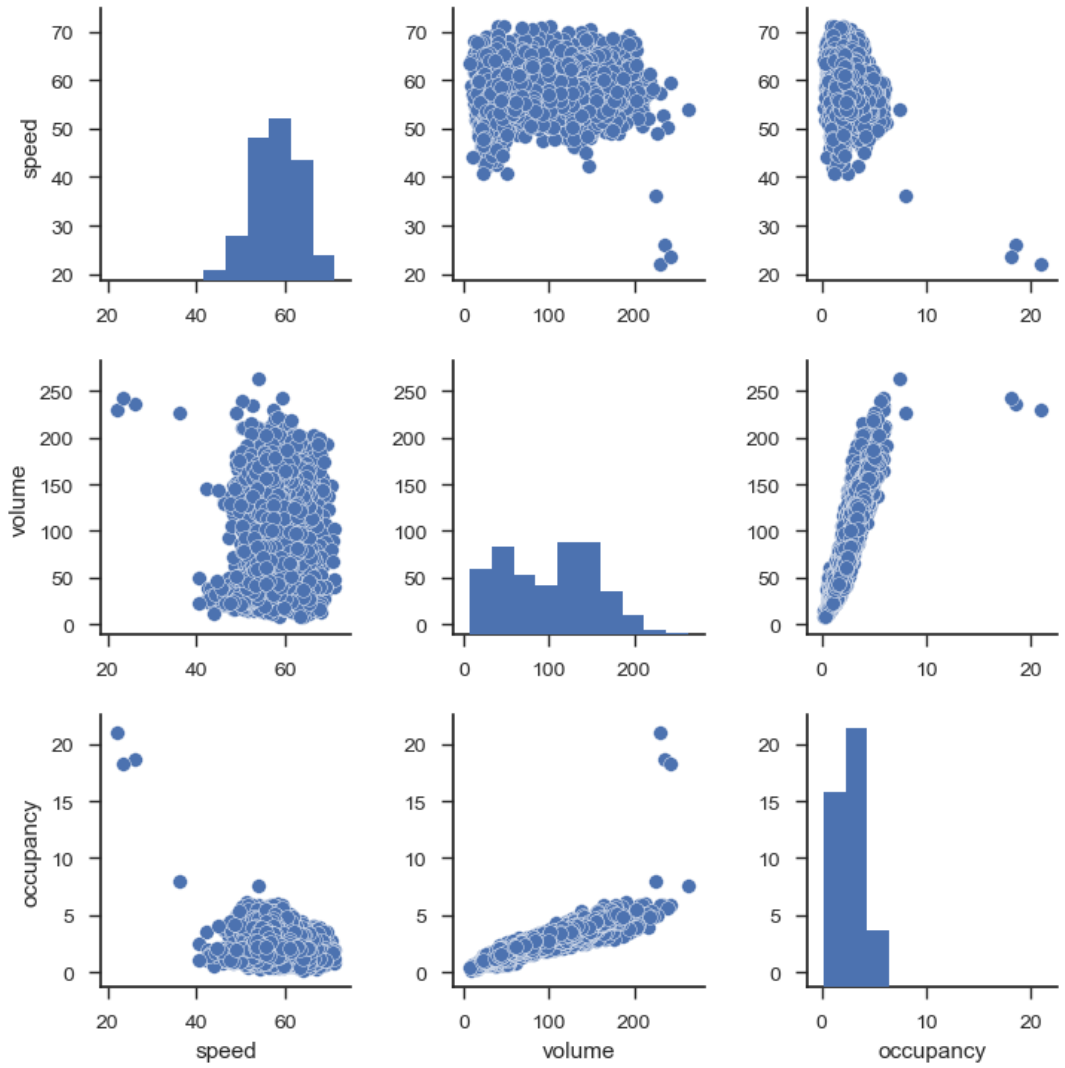
DIFFERENT PLOT FOR DIFFERENT VIEWS

- Scatter
- Scatter plot matrix
- Line plots
- Bar plots
- Histograms
- Box plots
- Violin plots
- Q-Q plots

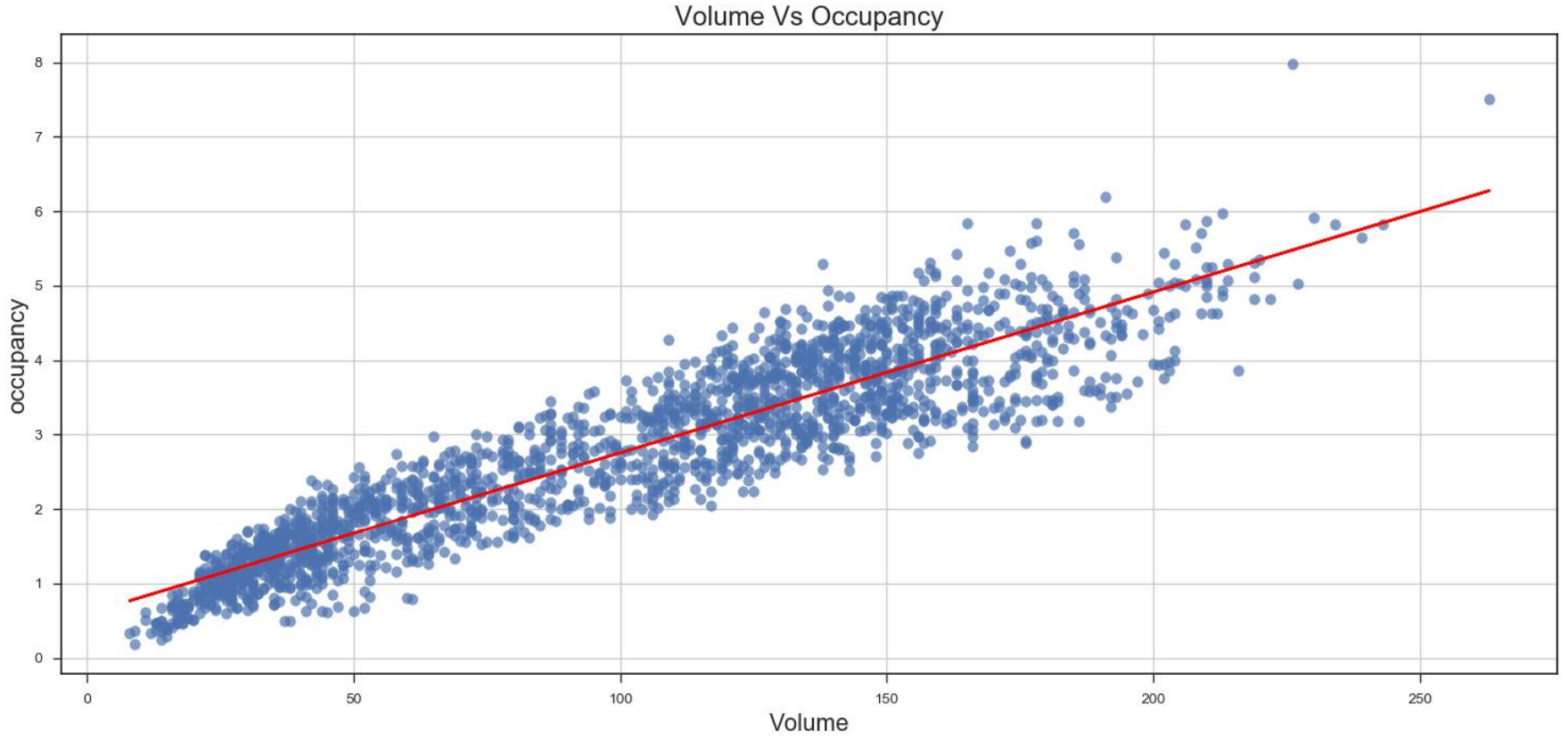
SCATTER PLOT



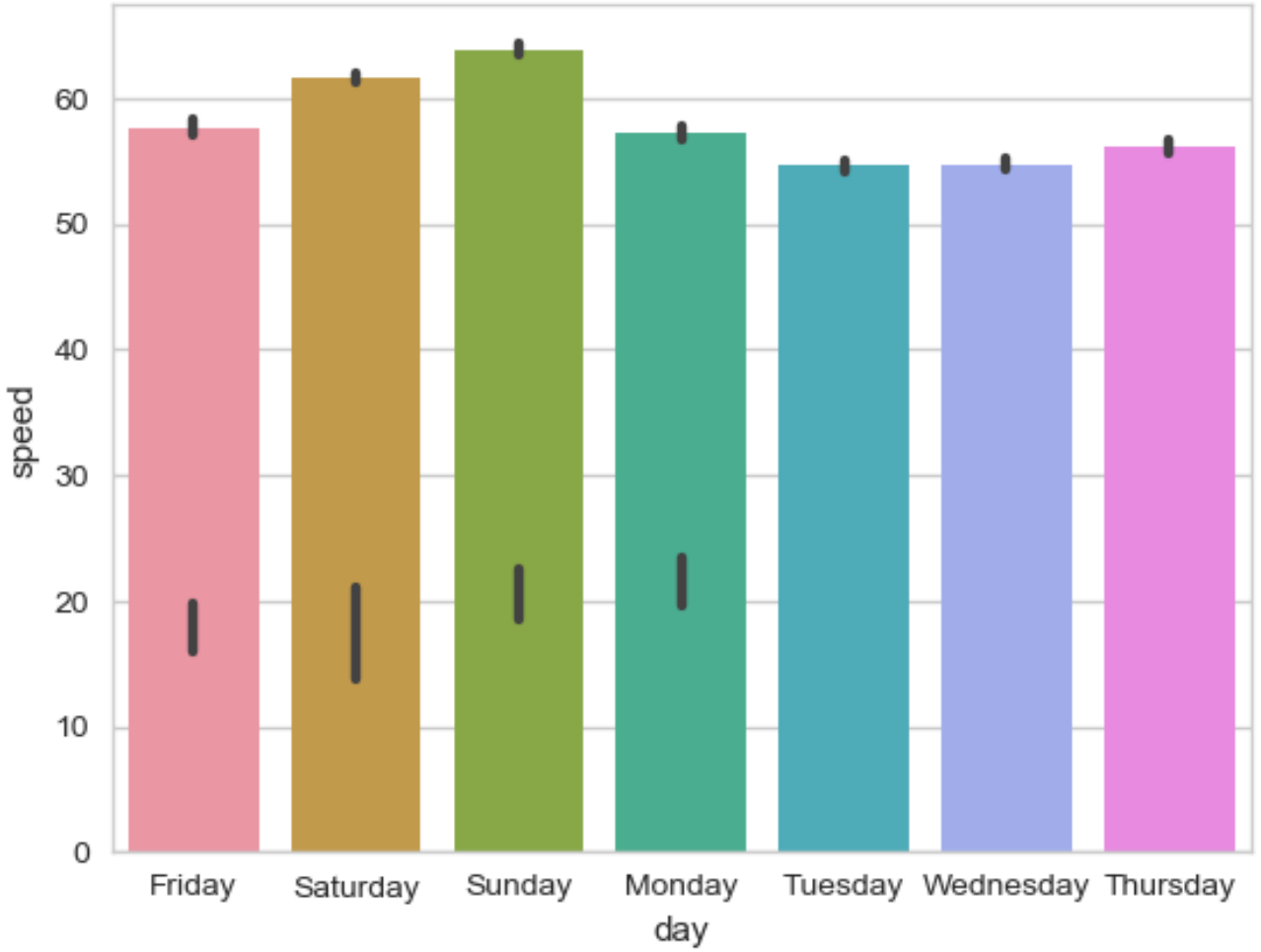
SCATTER PLOT MATRIX



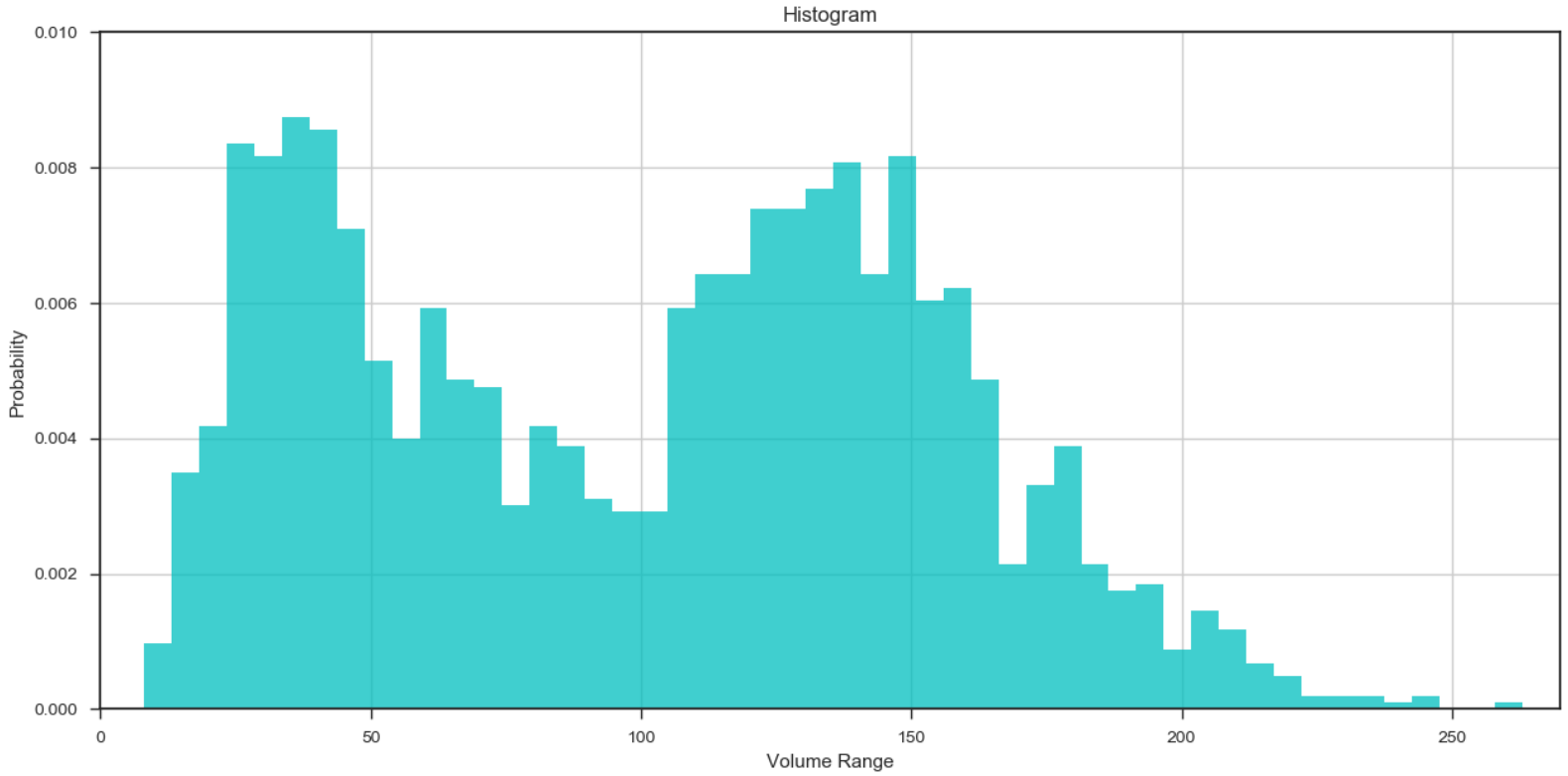
LINE PLOTS


Volume Vs Occupancy
occupancy
Volume
0
1
2
3
4
5
6
7
8
0
50
100
150
200
250

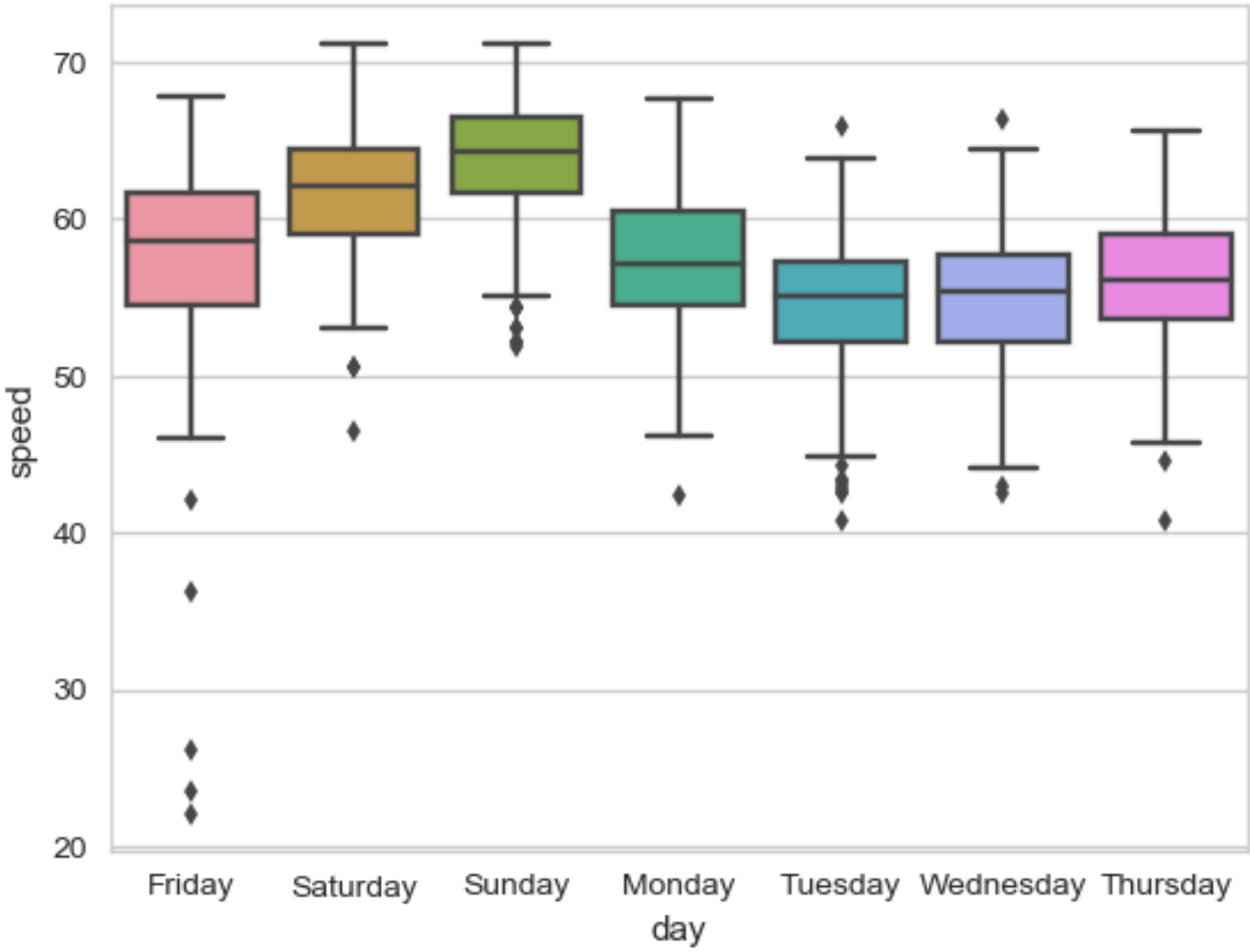
BAR PLOTS


speed
60
50
40
30
20
10
0
Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
day

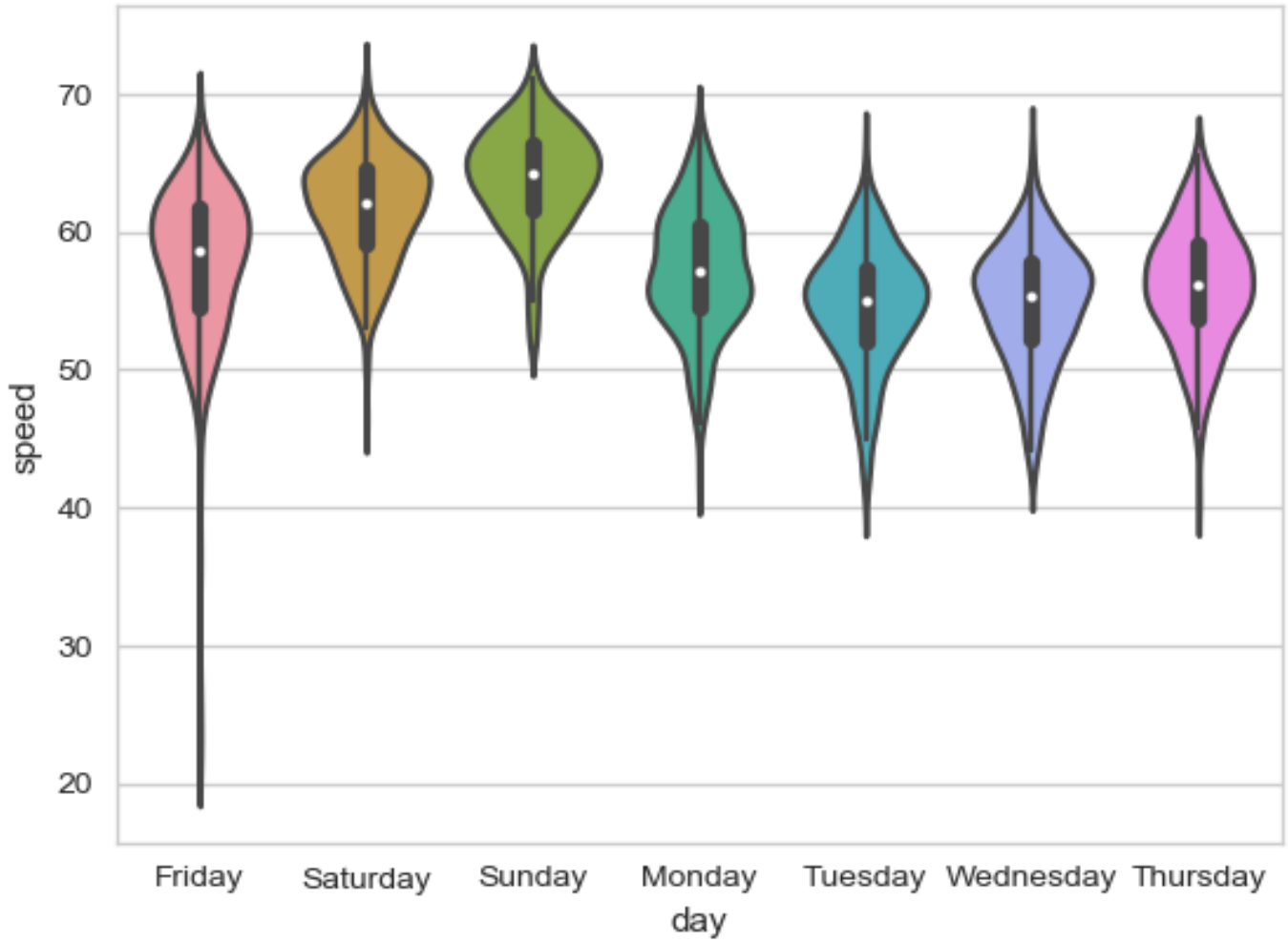
HISTOGRAMS


Histogram
0.010
0.008
0.006
0.004
0.002
0.000
Probability
0
50
100
150
200
250
Volume Range

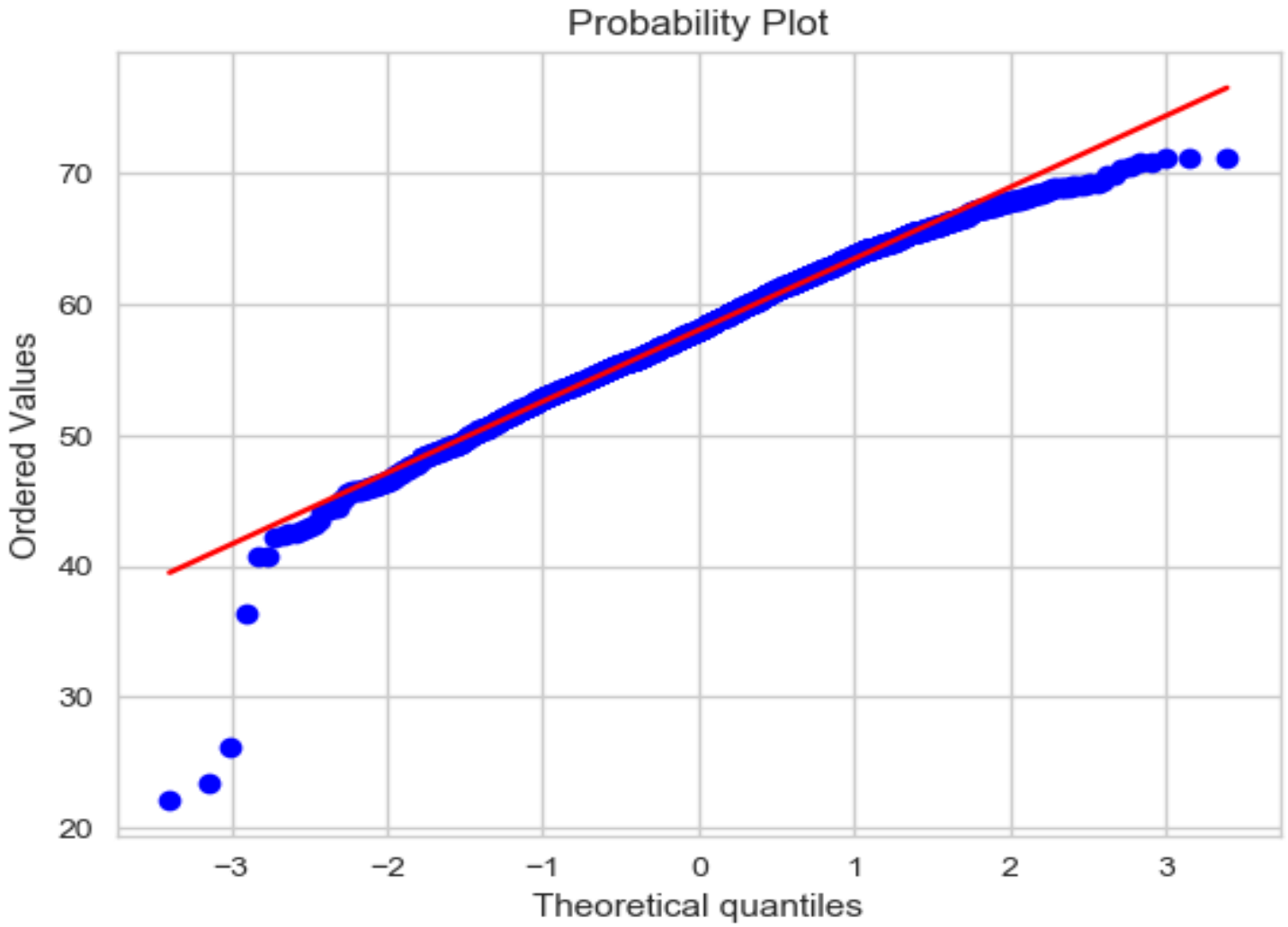
BOX PLOTS

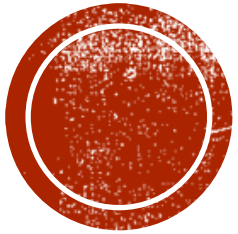

70
60
50
40
30
20
speed
Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
day

VIOLIN PLOTS



Q-Q PLOTS


Probability Plot
Ordered Values
70
60
50
40
30
20
−3
−2
−1
0
1
2
3
Theoretical quantiles



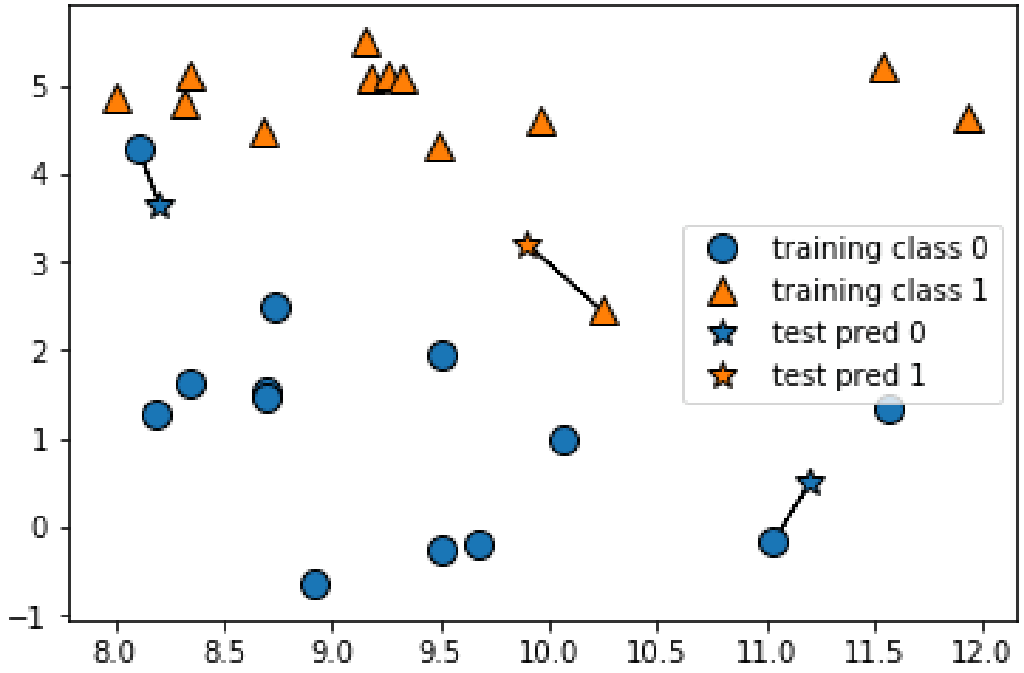
Machine Learning Techniques

CLASSIFICATION

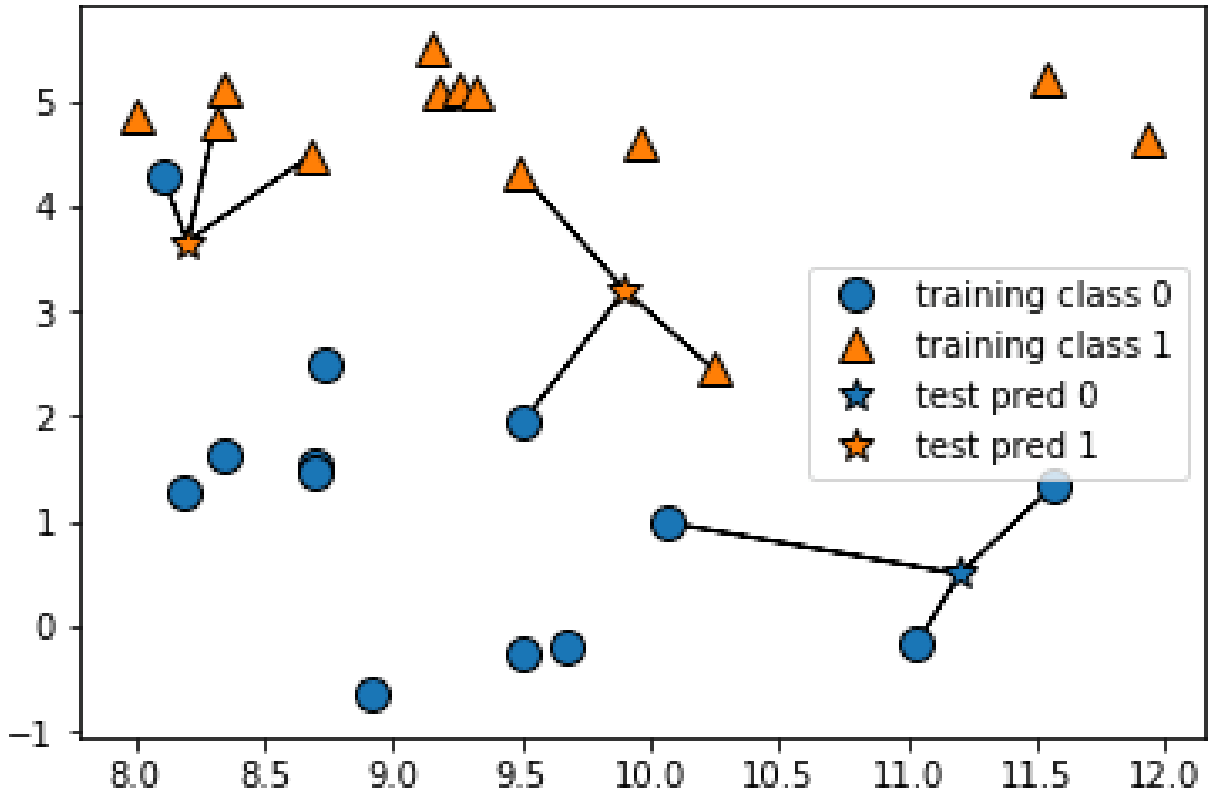
- Well-known classification algorithms
  - *k*-nearest neighbor
  - Logistic Regression
  - Decision Trees
  - Random Forests
  - Support Vector Machines
  - Neural Networks

NEAREST NEIGHBOR: BASIC IDEA

- Assigns the label of its nearest neighbor to an observation x
- We need to implement a distance measure $d(x_i, x)$ between pairs of observations



K-NEAREST NEIGHBOR (K=3)



K-NEAREST NEIGHBOR ALGORITHM

Classify $(\boldsymbol{X}, \boldsymbol{Y}, x)$ {reads data $\boldsymbol{X}$, labels $\boldsymbol{Y}$ and query x}

for $i = 1$ to m do

Compute distance $d(x_i, x)$

end for

Compute set I containing indices for k smallest distances $d(x_i, x)$

return majority label of $\{y_i \; where \; i \in I\}$

DISTANCE COMPUTATION

- Distance calculation $d(x_i, x)$ for all observations can become extremely costly when
  - the number of observations is very large
  - x_i has high number of dimensions
- Solution
  - Random projection

THE LINEAR PROBABILITY MODEL

In the linear regression:

$Y = \beta_0 + \beta_1 X + e$; where $Y = (0, 1)$

- The error terms are heteroskedastic
- e is not normally distributed because Y takes on only two values
- The predicted probabilities can be greater than 1 or less than 0

LOGISTIC REGRESSION

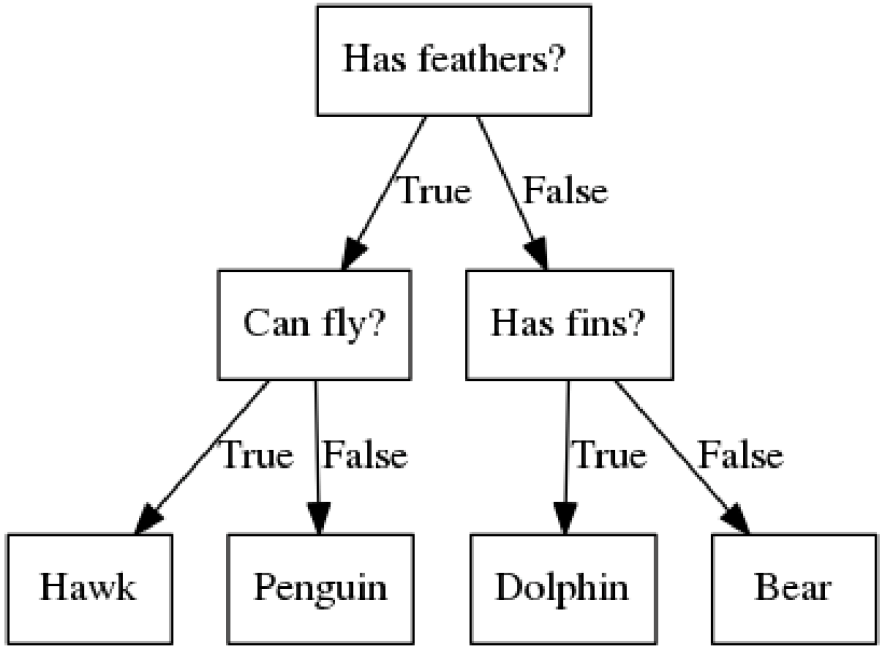
The "logit" model:

$$\ln[p/(1-p)] = \beta_0 + \beta_1 X$$

- p is the probability that the event Y occurs, p(Y=1)
  - [range=0 to 1]
- p/(1-p) is the "odds ratio"
  - [range=0 to ∞]
- ln[p/(1-p)]: log odds ratio, or "logit"
  - [range=-∞ to +∞]

DECISION TREE

- Learn a hierarchy of if-else questions
  - leading to a decision

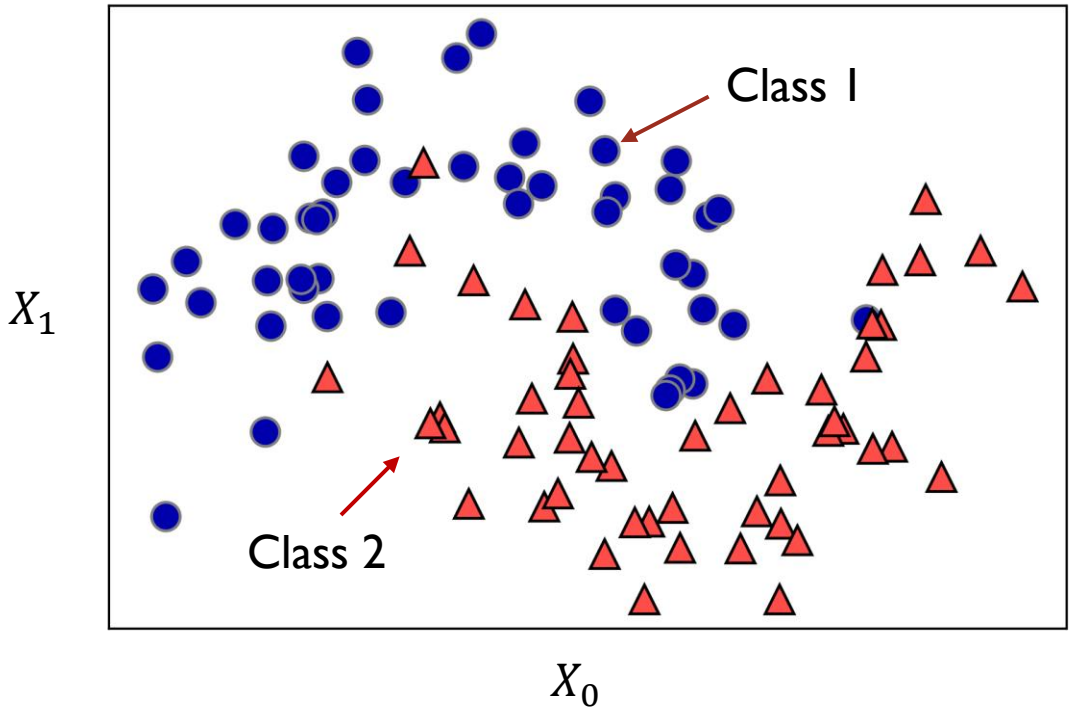


Each internal node: test one attribute X_i

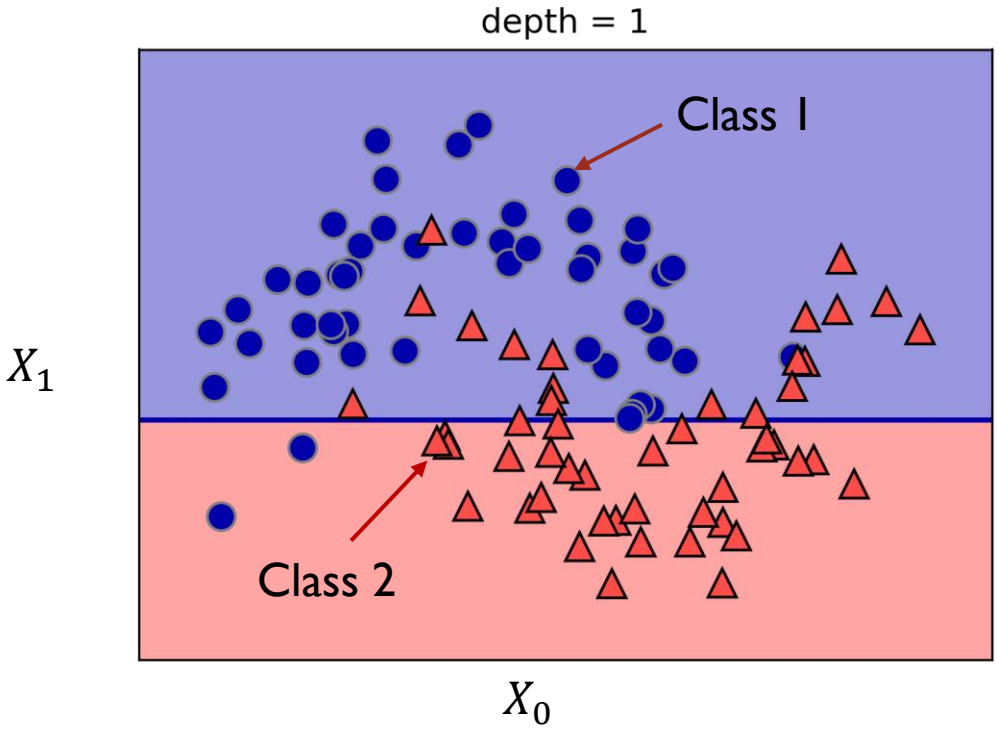
Each branch from a node: selects one value for X_i

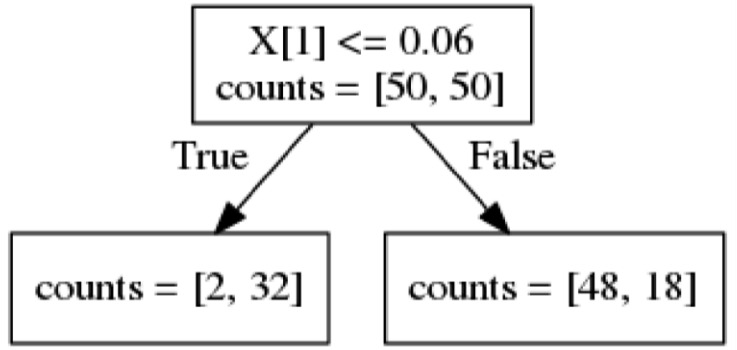
Each leaf node: predicts Y (or P(Y | X ∈ leaf))

BUILDING DECISION TREES

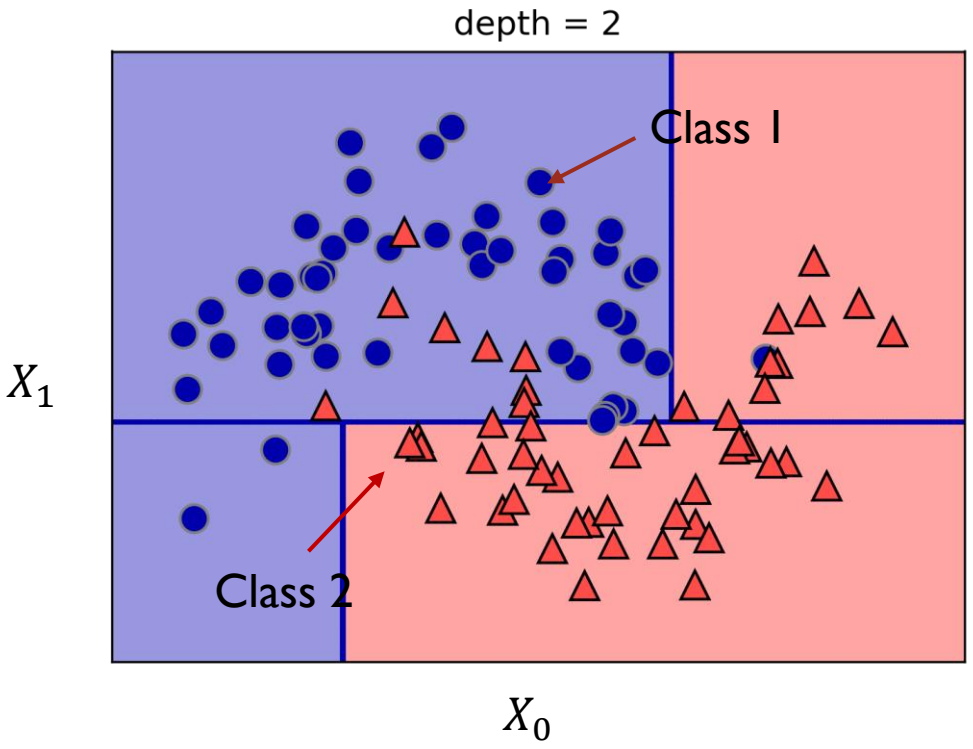


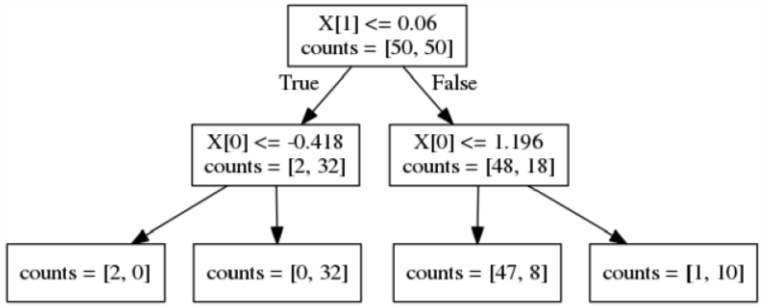
BUILDING DECISION TREES



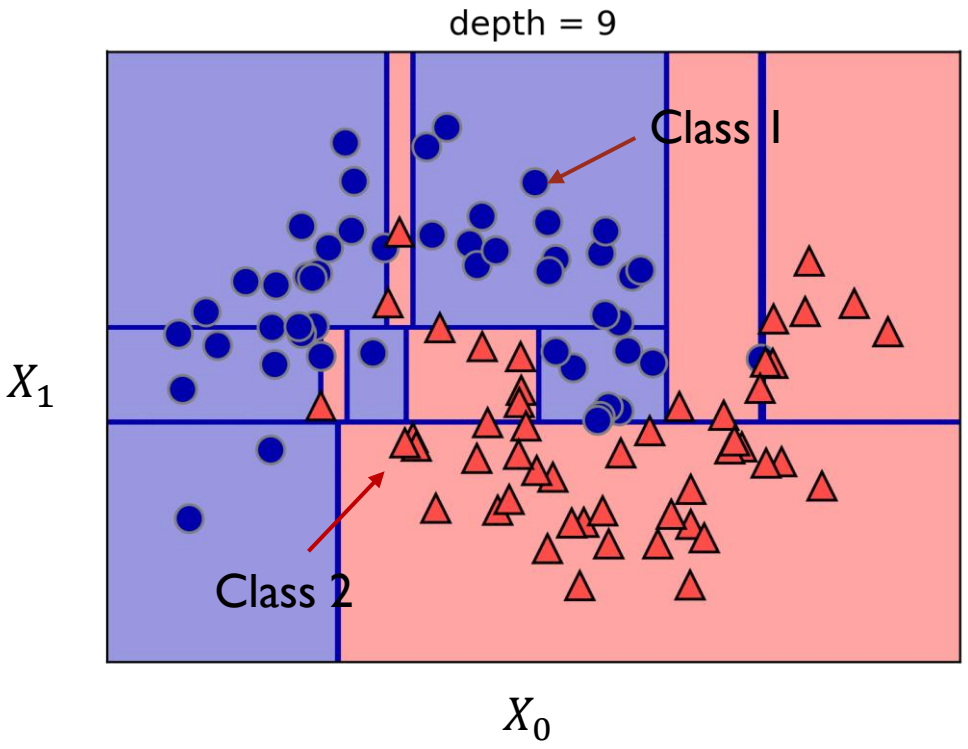


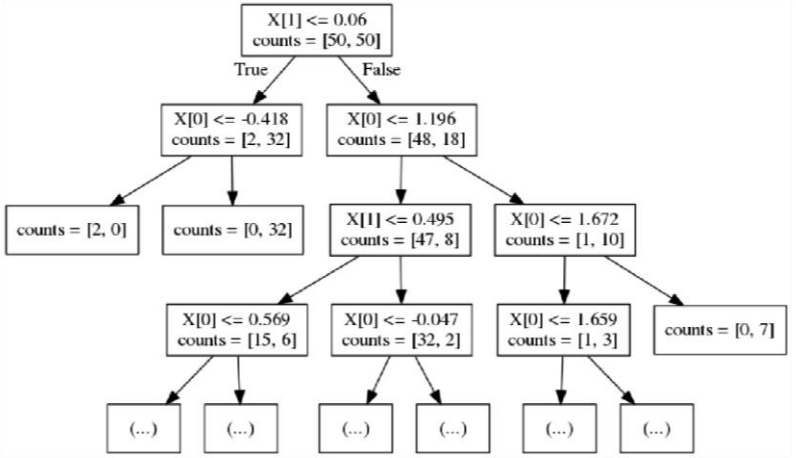
BUILDING DECISION TREES





BUILDING DECISION TREES





OVERFITTING

- Error of model m over:
  - training data: $error_{train}(m)$
  - entire distribution $\mathbb{D}$ of the data: $error_{\mathbb{D}}(m)$
- Model $\mathrm{m} \in M$ overfits the training data if there is an alternative hypothesis $m' \in M$ so that

$$error_{train}(m) < error_{train}(m')$$

and

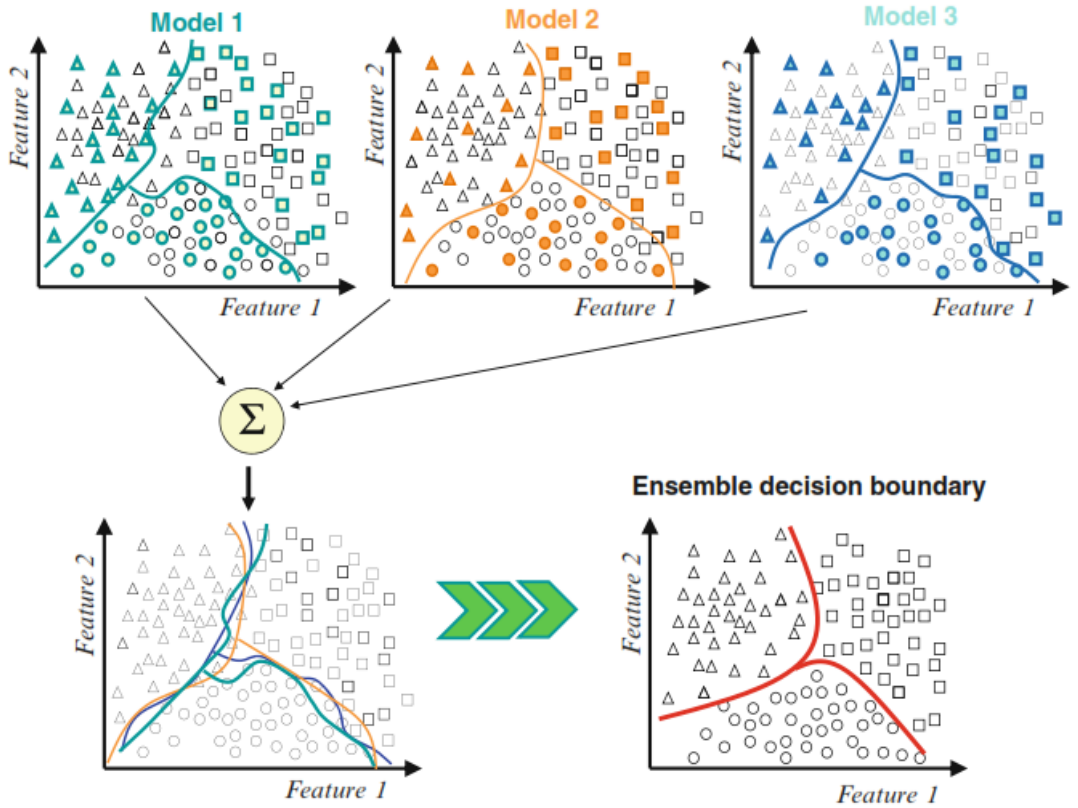
$$error_{\mathbb{D}}(m) > error_{\mathbb{D}}(m')$$

AVOIDING OVERFITTING

- How can we avoid overfitting?
  - Stop growing when the data split is not statistically significant
  - Grow full tree then post-prune
- How can we select "best" tree
  - Measure performance over training data
  - Measure performance over a separate validation dataset
  - Apply a statistical test to estimate whether pruning or expanding a particular node is likely to produce an improvement beyond the training set

ENSEMBLES

- Combines multiple machine learning models



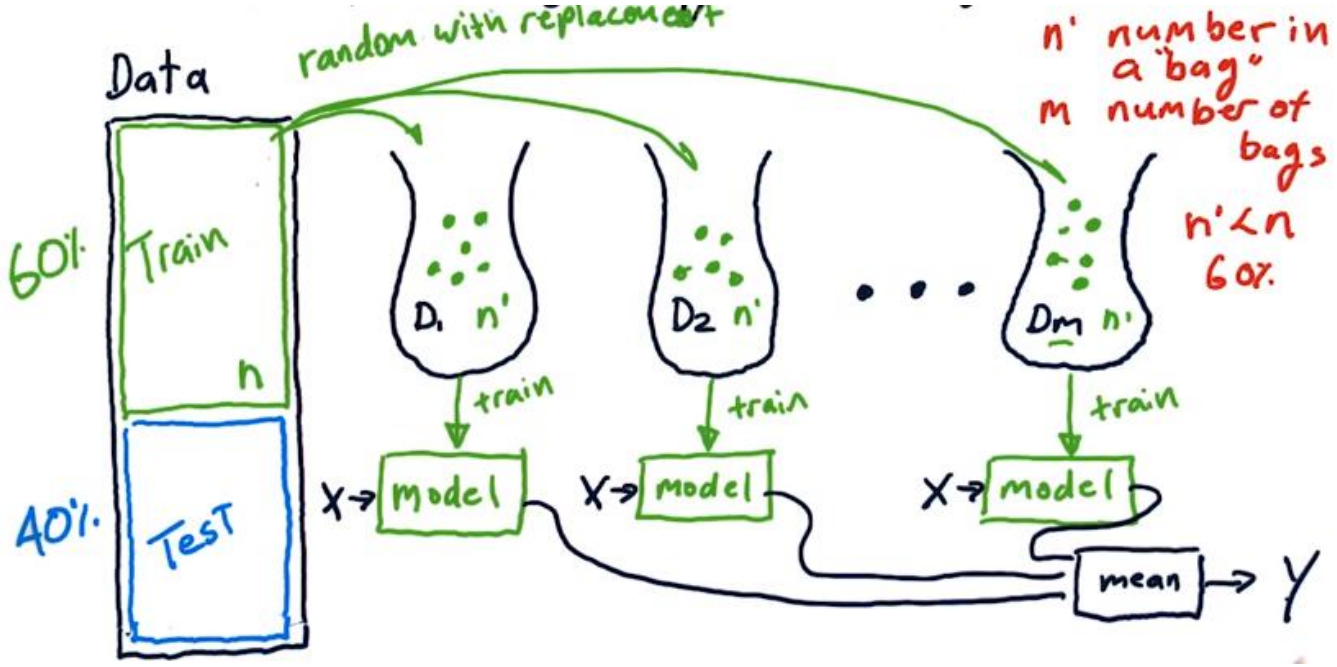
ENSEMBLES

Trees can be simple but often produce noisy predictions

- Bagging or Bootstrap Aggregation
  - Fit many large trees to bootstrapped resampled data and classify by majority vote
- Random Forest
  - Advanced version of bagging
- Boosting
  - Fit many large trees to reweighted data and classify by weighted majority vote

BAGGING OR BOOTSTRAP AGGREGATION

- Train different models with different bags of data
  - train m models using data sets $D_1, D_2, \ldots, D_m$ for each model
- Combine the outputs
  - 'voting' for Classification
  - 'mean' for Regression



BAGGING OR BOOTSTRAP AGGREGATION

- Bagging averages many trees and produces smoother decision boundaries

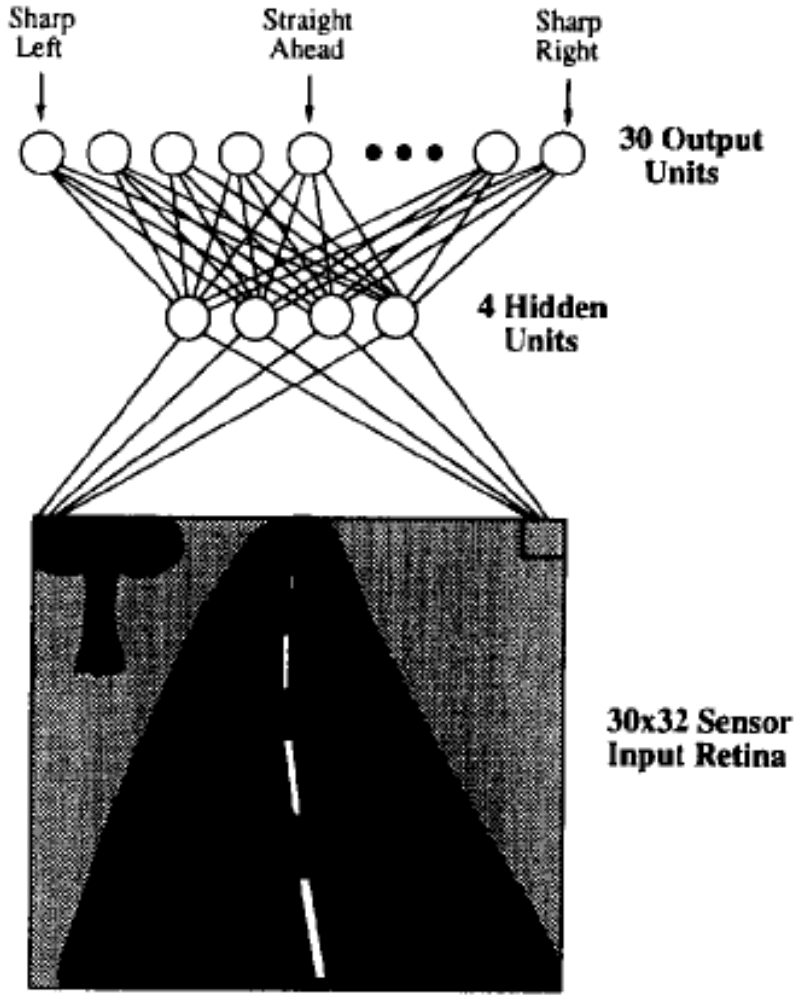
RANDOM FOREST

- Refinement of bagged trees
- Trees are grown over bootstrapped sample
- At each tree split, a random sample of m features is drawn as candidates for splitting
- Out of bag error rate
- Tries to improve on bagging by de-correlating the trees and reduce the variance

BOOSTING

- Average many trees, each grown to re-weighted versions of the training data
- Weighting de-correlates the trees by focusing on regions missed by past trees
- Final classifier is a weighted average of classifiers

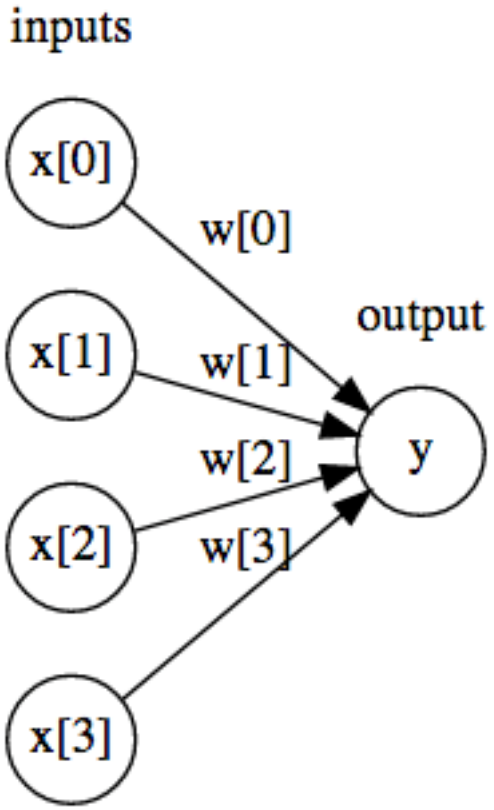
ARTIFICIAL NEURAL NETWORKS: EXAMPLE



960 input units

Source: ALVINN (Autonomous Land Vehicle in a Neural Network)
Pomerleau (1993)

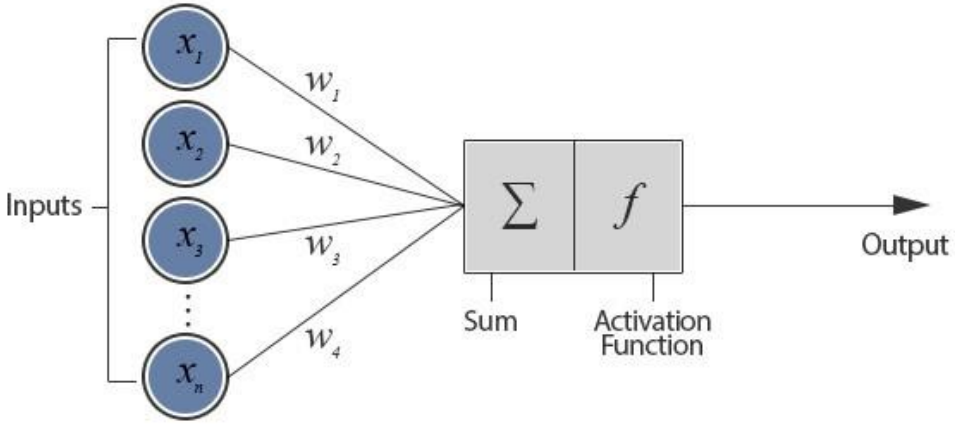
LINEAR MODEL



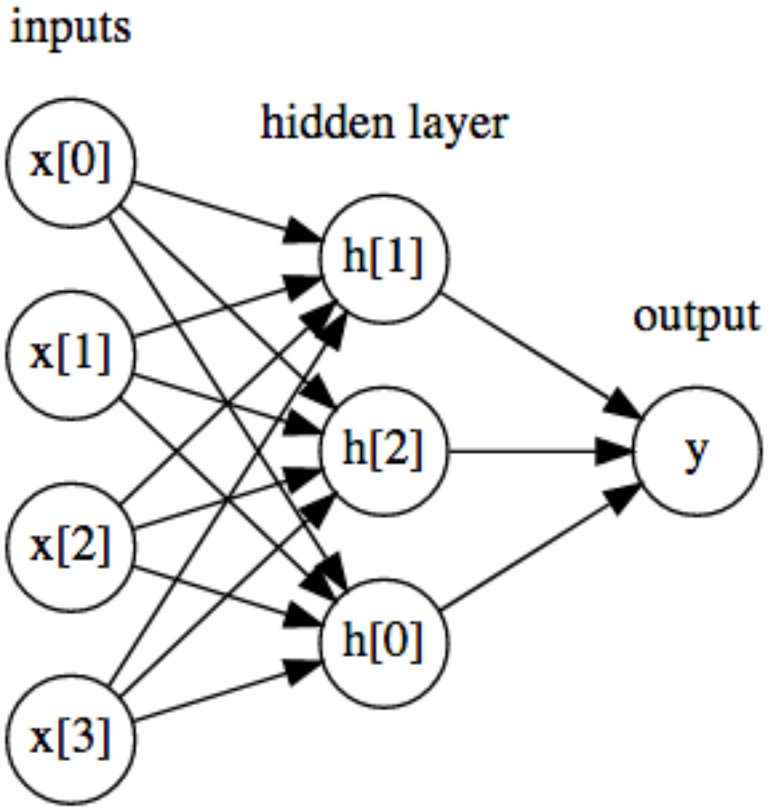
$$\hat{y} = \boldsymbol{w}^T \boldsymbol{x}$$

NEURAL NETS

- **Neural network** is a computational graph whose nodes are computing units and whose directed edges transmit numerical information from node to node.
- Each computing unit (neuron) is capable of evaluating a single primitive function (activation function) of its input.
- The network represents a chain of function compositions which transform an input to an output vector.



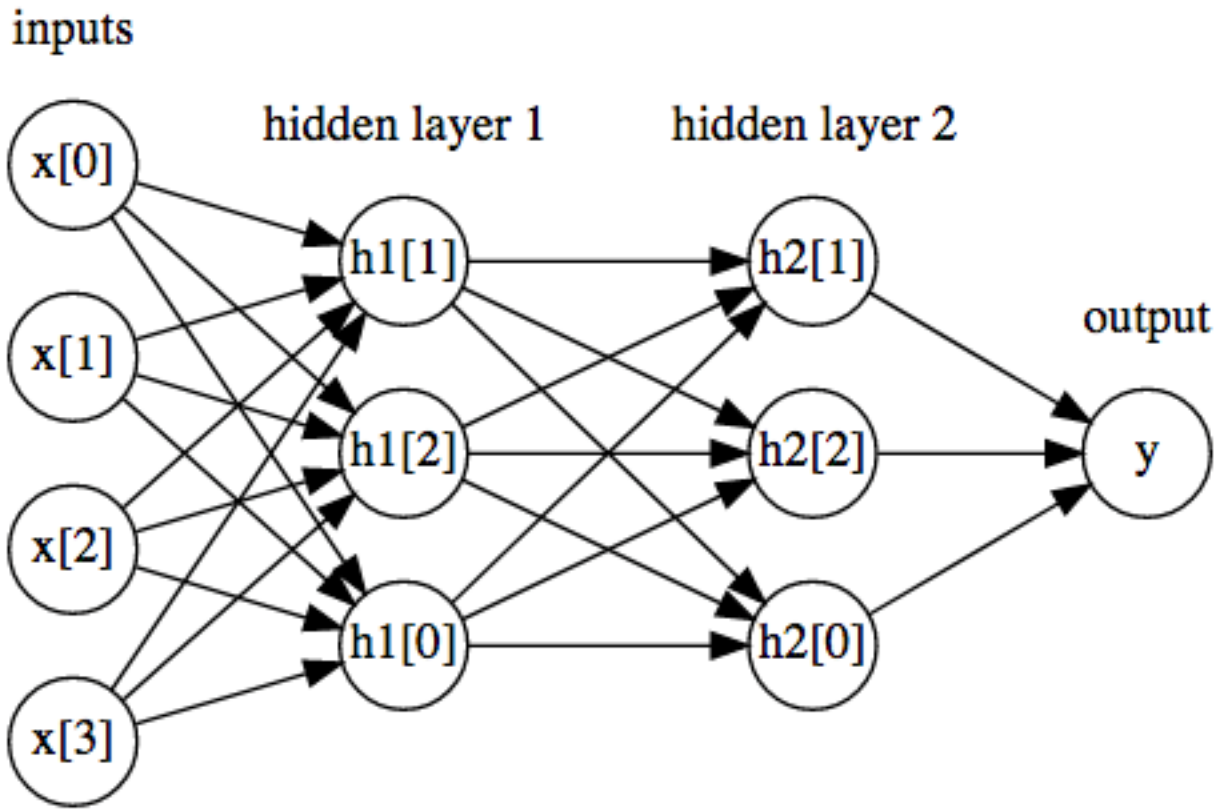
SINGLE HIDDEN LAYER



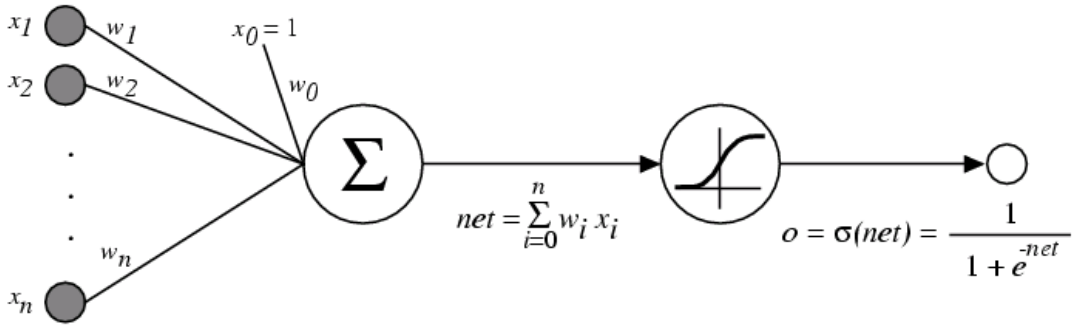
$$h[i] = \boldsymbol{w}_i^T \boldsymbol{x}$$

$$\hat{y} = \boldsymbol{v}^T \boldsymbol{h}$$

MULTIPLE HIDDEN LAYERS: DEEP LEARNING



SIGMOID UNIT



- $\sigma(x)$ is the sigmoid unit

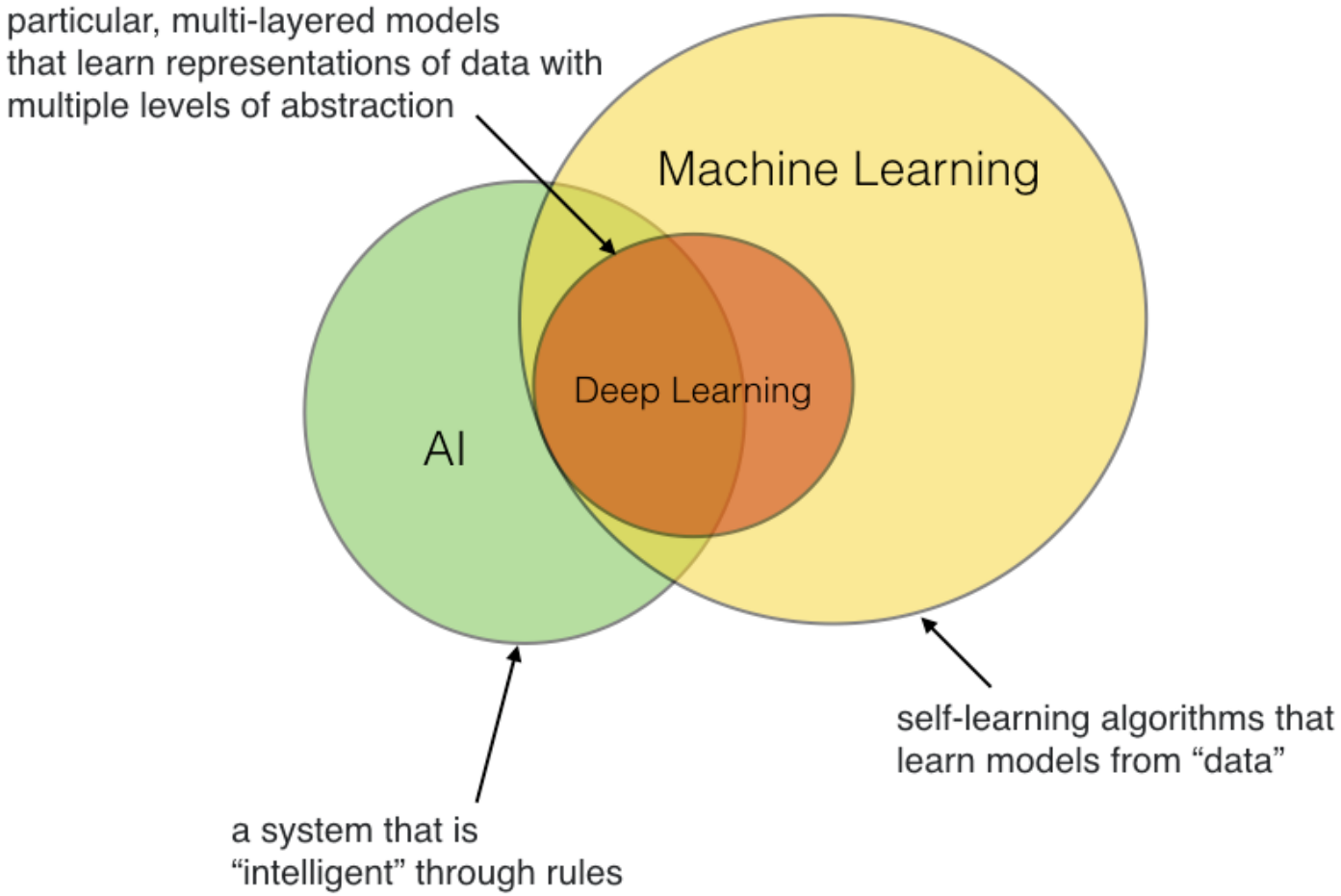
$$\frac{1}{1 + e^{-x}}$$

- $\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$

DEEP LEARNING EVOLUTION

- Deep learning has created a unique opportunity to deal with more complex problems.
- The core architecture established based on **Artificial Neural Network (ANN).**
- The processing elements and the architecture converted it into a powerful tool.
- Emergence of deep learning methods has been encouraged by a tremendous increase in **computational power and data availability**.

DEEP LEARNING



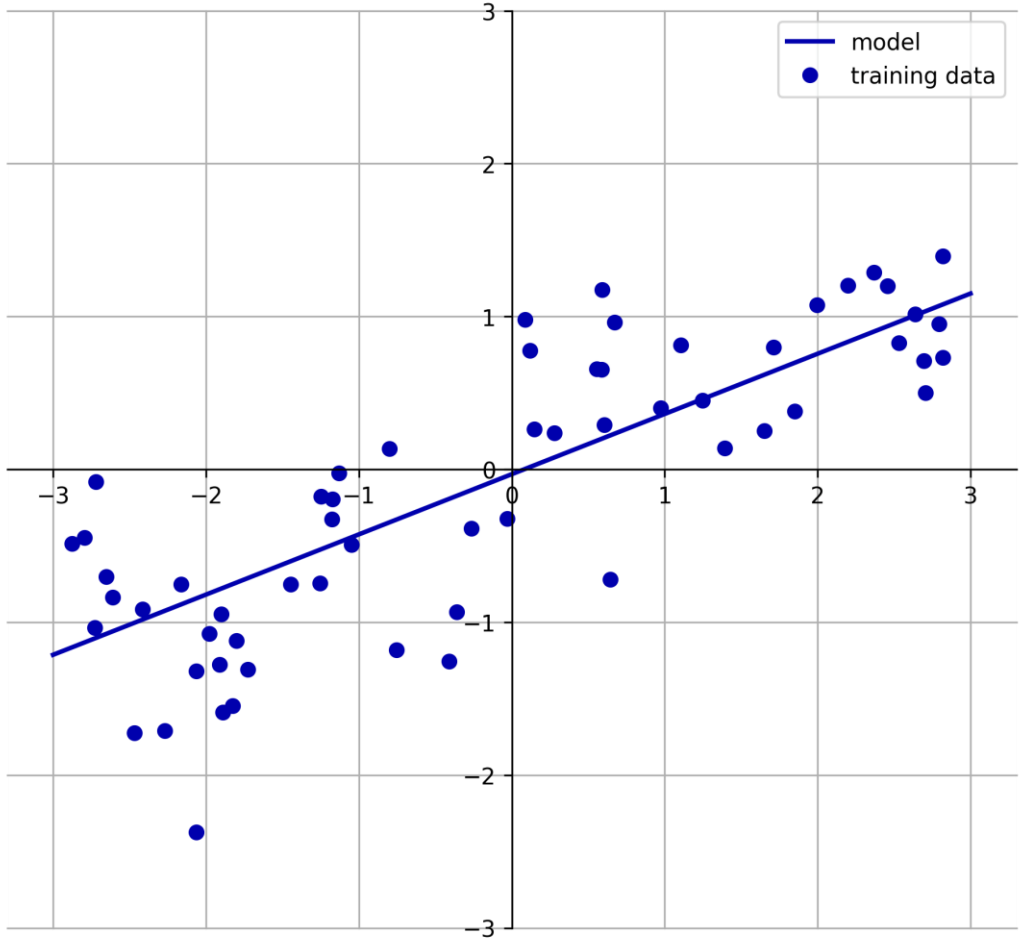
DEEP LEARNING

- More Efficient in Learning high dimensional (graph, image etc.) data representation

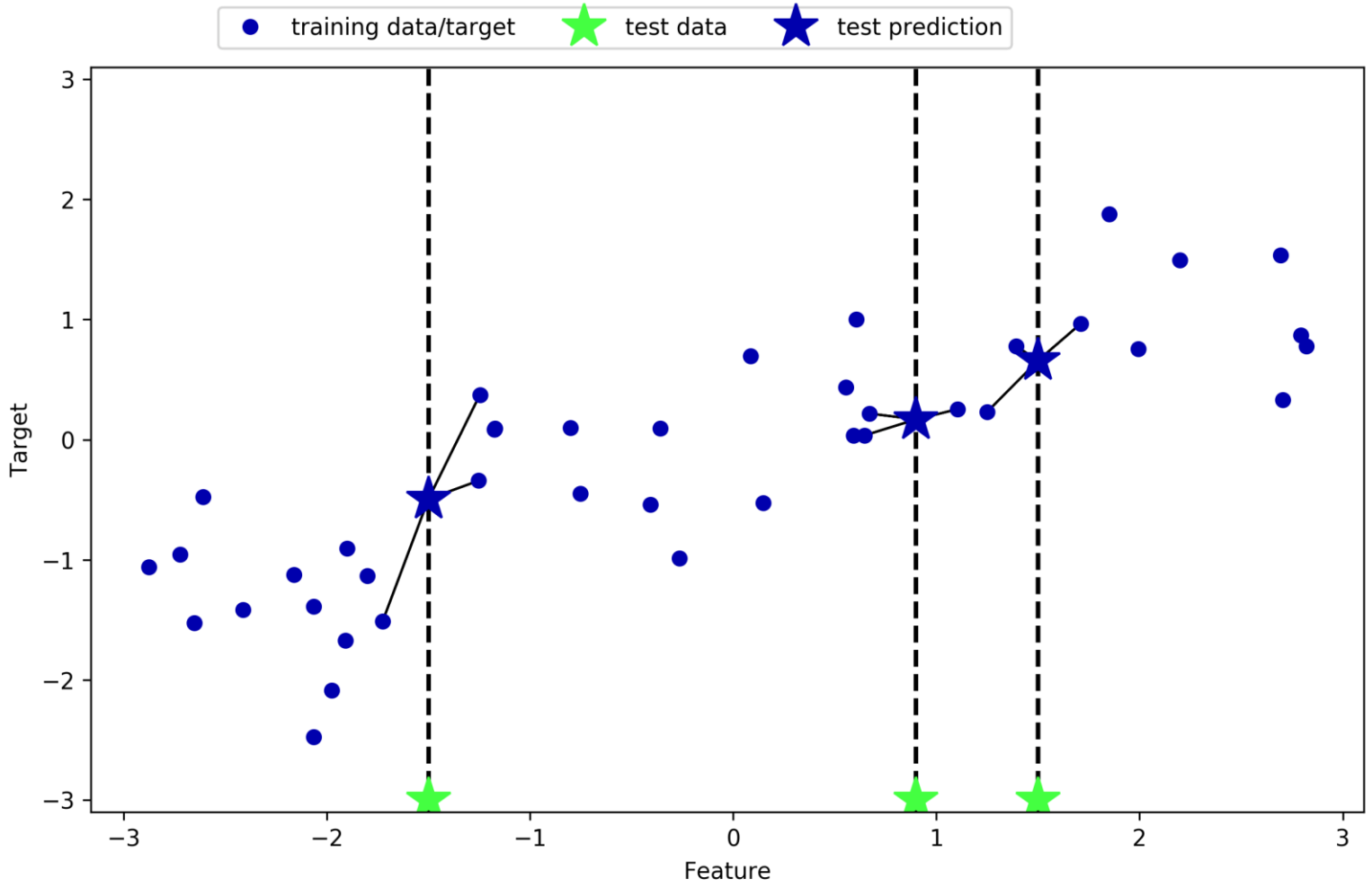
REGRESSION

- Some regression models
  - Linear regression
  - *k*-nearest neighbor regression

LINEAR REGRESSION



K-NEAREST NEIGHBOR REGRESSION



SUPERVISED ALGORITHMS BASIC PROBLEM

- Supervised algorithms
  - assume that labeled training data are available
- Labeling may be expensive, error prone, or sometimes impossible
- Examples
  - assign a topic to each tweet based on its contents
  - assign a pattern for each day's data of activity-travel diaries
  - identify and find the cluster of destinations of a group of individuals (e.g., commuters, tourists, evacuees)

UNSUPERVISED ALGORITHMS: CLUSTERING

- No need to have a labelled dataset
- Typically solved by using clustering algorithms
- Divided the data some sort of clusters

K-MEANS ALGORITHM

- *K*-means is a prototypical clustering algorithm
- Given data $X = \{x_1, x_2, \ldots, x_m\}$
- *K*-means partitions X into k clusters such that
  - each point in a cluster is similar to points from its own cluster than with points from some other cluster.

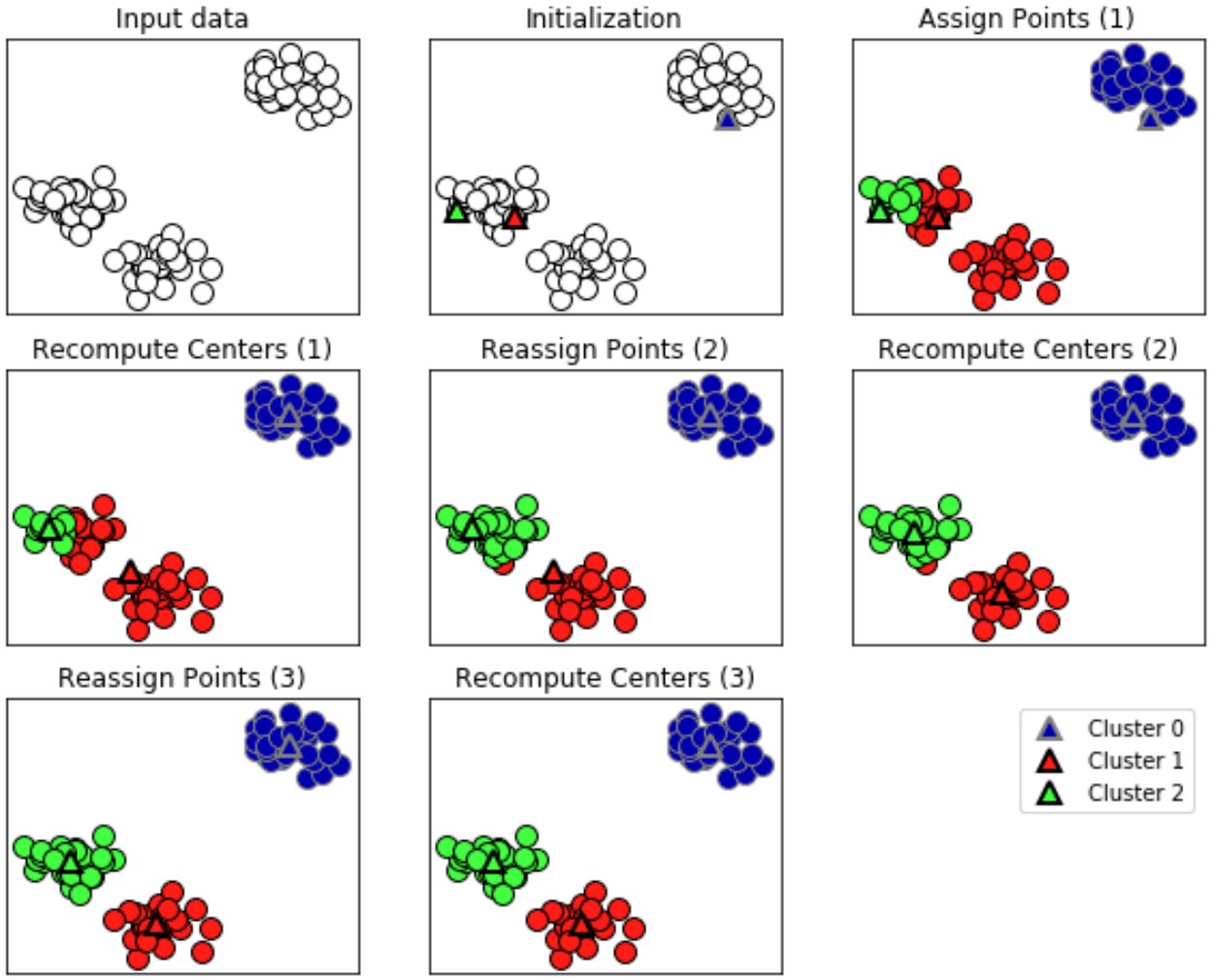
K-MEANS ALGORITHM: PROBLEM DEFINITION

- Define
  - prototype vectors $\mu_1, \ldots, \mu_k$ and
  - an indicator vector r_{ij} which is 1 if, and only if, x_i is assigned to cluster j
- Minimize the distortion measure

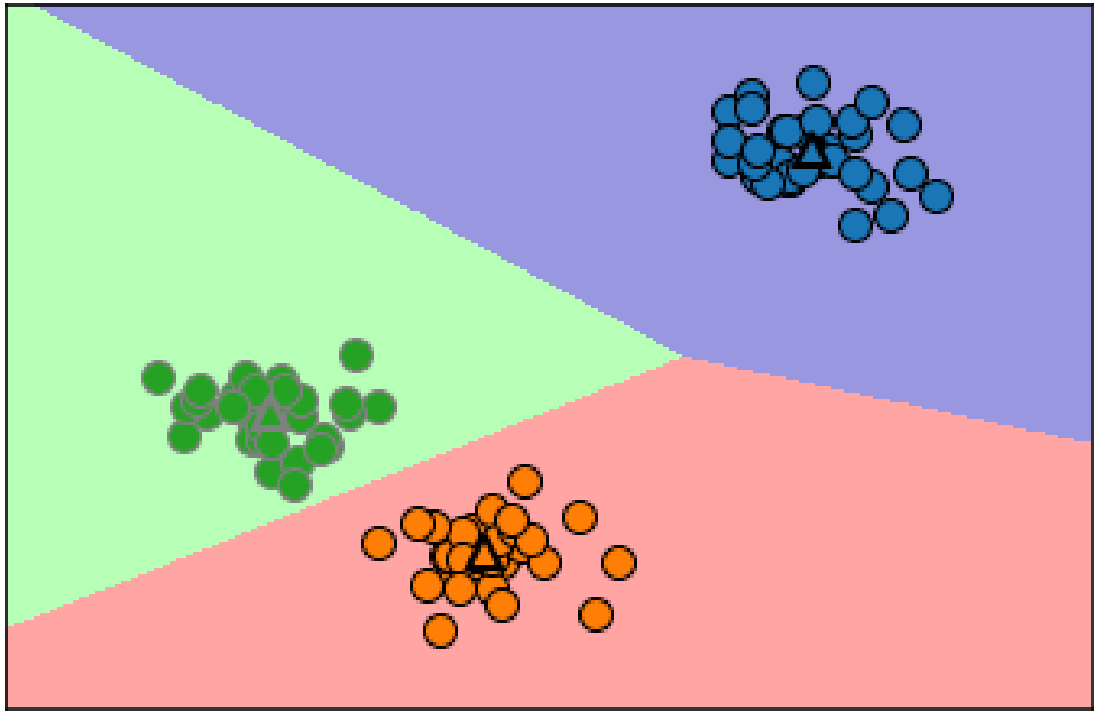
$$J(r,\mu) := \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{k} r_{ij}\left\|x_i - \mu_j\right\|^2$$

distance of each point from the prototype vector

EXAMPLE


Input data
Initialization
Assign Points (1)
Recompute Centers (1)
Reassign Points (2)
Recompute Centers (2)
Reassign Points (3)
Recompute Centers (3)
Cluster 0
Cluster 1
Cluster 2

K-MEANS DECISION BOUNDARIES



ISSUES WITH K-MEANS

- Sensitive to the choice of initial cluster center
- Makes a hard assignment of every point to a cluster center

REFERENCES

- Ben-Akiva, M. E. and S. R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, The MIT Press.
- Ortuzar, J. D. and L. G. Willumsen (2011). Modelling transport, Wiley. 4th ed.
- Koppelman, F. S. and C. Bhat (2006). "A self instructing course in mode choice modeling: multinomial and nested logit models"
- Train, K. (2003). Discrete choice methods with simulation, Cambridge University Press (Available for free online at http://www.econ.berkeley.edu/books/choice2.html)
- Bhat, C. R., N. Eluru, R. Copperman (2008). "Flexible model structures for discrete choice analysis." Handbook of transport modelling 1: 71–90.
- Bhat, C. R. (1998). "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling." Transportation Research Part A: Policy and Practice 32(7): 495-507.
- Cameron A. C. and P.K. Trivedi "Microeconometrics: Methods and Applications" Cambridge University Press 2005.
- Bhat, C.R. (2003), "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences", Transportation Research Part B, Vol. 37, No. 9, pp. 837-855
- Bhat, C.R. (2001), "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model", Transportation Research Part B, Vol. 35, No. 7, pp. 677-693.