LECTURE 1 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

PREDICTIVE ANALYTICS

A P MOORE

Why do they call it intelligence?

Data + model → prediction

Assume there is enough data to find statistical associations to solve specific tasks

Data + model → prediction

Define how well the model solves the task and adapt the parameters to maximize performance

LEARNING A FUNCTION

UCL SCHOOL OF MANAGEMENT

$$x \rightarrow y$$

$$x \rightarrow f(x) \rightarrow y$$

LEARNING A FUNCTION

$$x \to y$$

$$x \to f(x) \to y$$

Measured data

Features

Inferred/Predicted/Estimated value

$$x' \to \hat{x} \to \boxed{f(x)} = \hat{y} \to y'$$

True initial value
(world state)

Learned/Fitted function
From n observations

True target value
(world state)

LEARNING A FUNCTION

$$x \rightarrow y$$

$$x \rightarrow f(x) \rightarrow y$$

Measured data

Features

Inferred/Predicted/Estimated value

$$x \rightarrow \hat{x} \rightarrow f(x) = \hat{y} \rightarrow y$$

True initial value
(world state)

True target value
(world state)

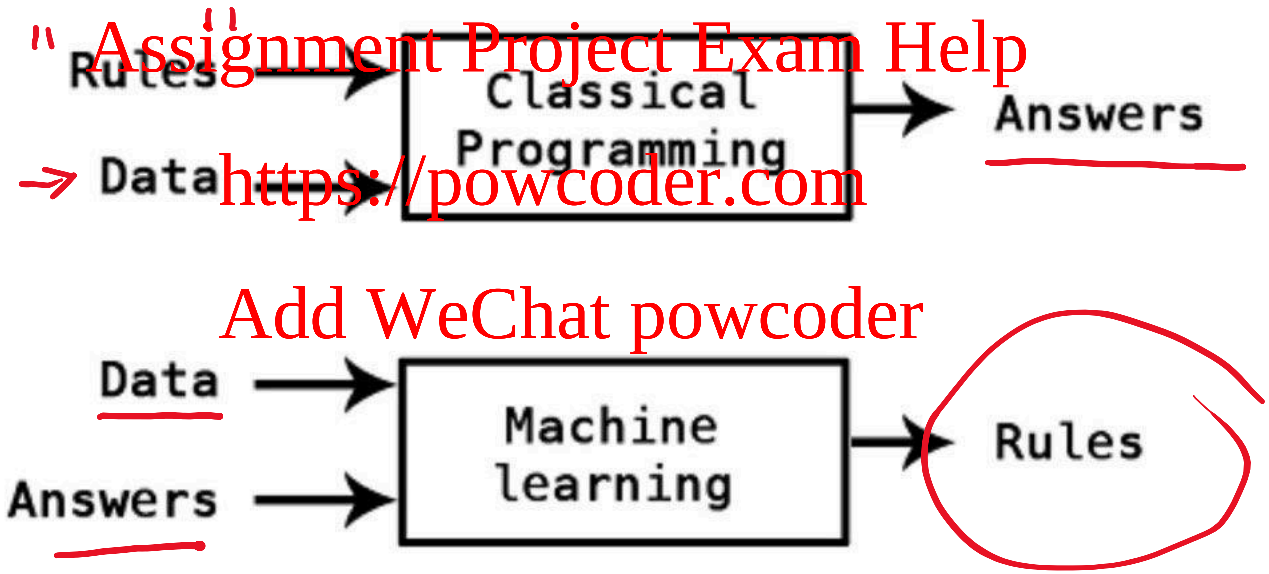
Learned/Fitted function
From n observations

input $x \rightarrow \boxed{f(x)} \rightarrow y$ output

data factor

MACHINE LEARNING

DATA DRIVEN AI



$x \rightarrow$

(x, y)

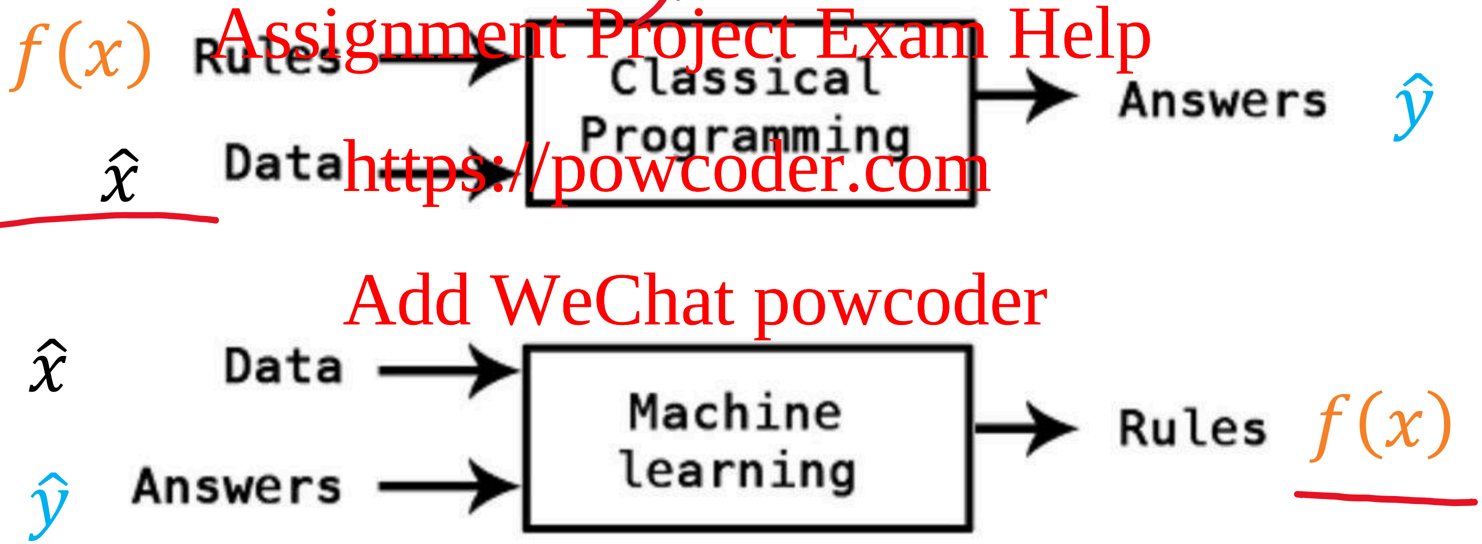
Source: https://twitter.com/Kpaxs/status/1163058544402411520

MACHINE LEARNING

DATA DRIVEN AI

UCL
SCHOOL OF
MANAGEMENT

$$x \to \hat{x} \to f(x) = \hat{y} \to y$$

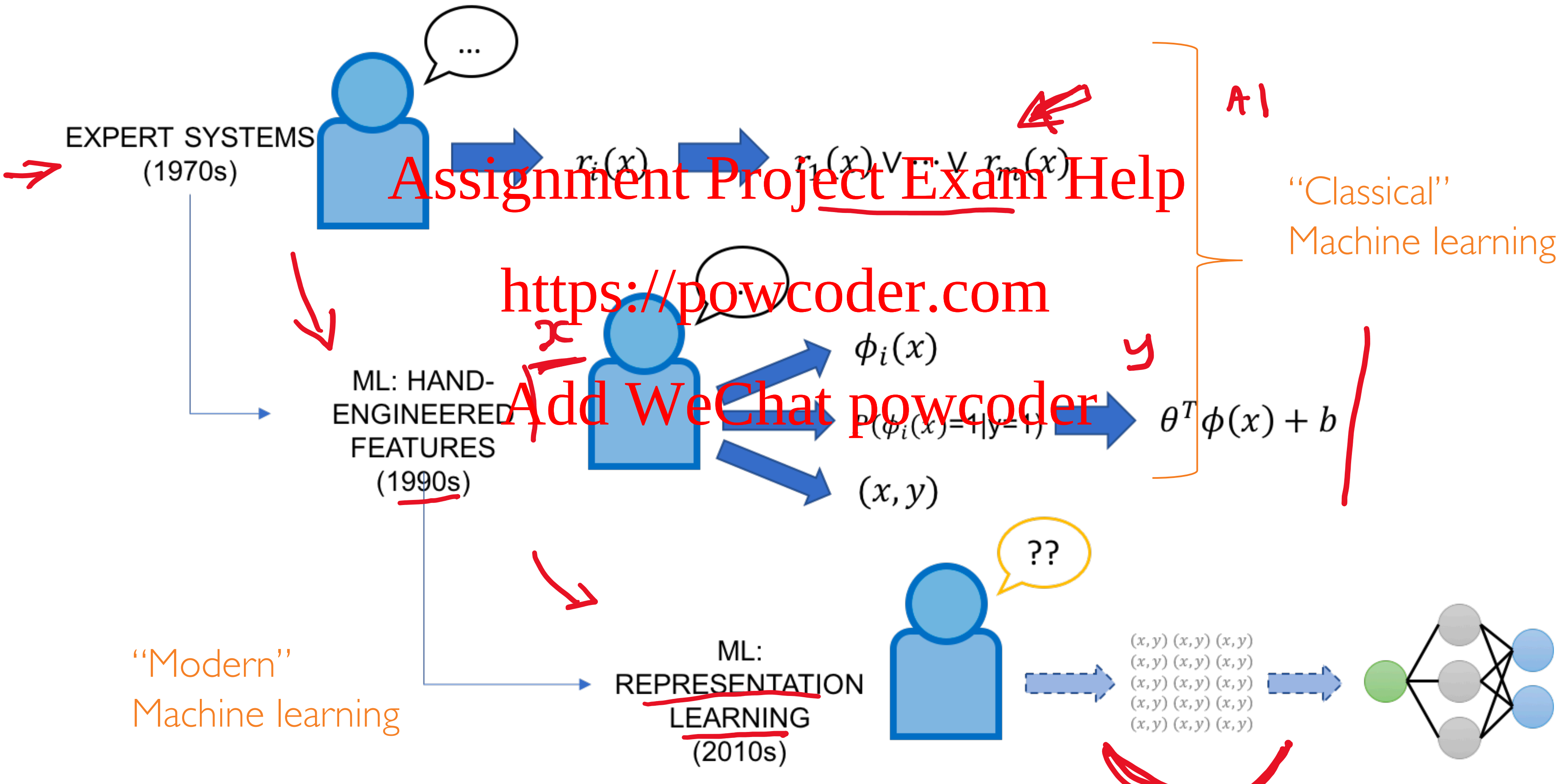


$\{x_i, y_i\}$ Labelled training data

Source: https://twitter.com/Kpaxs/status/1163058544402411520

Learning the rules

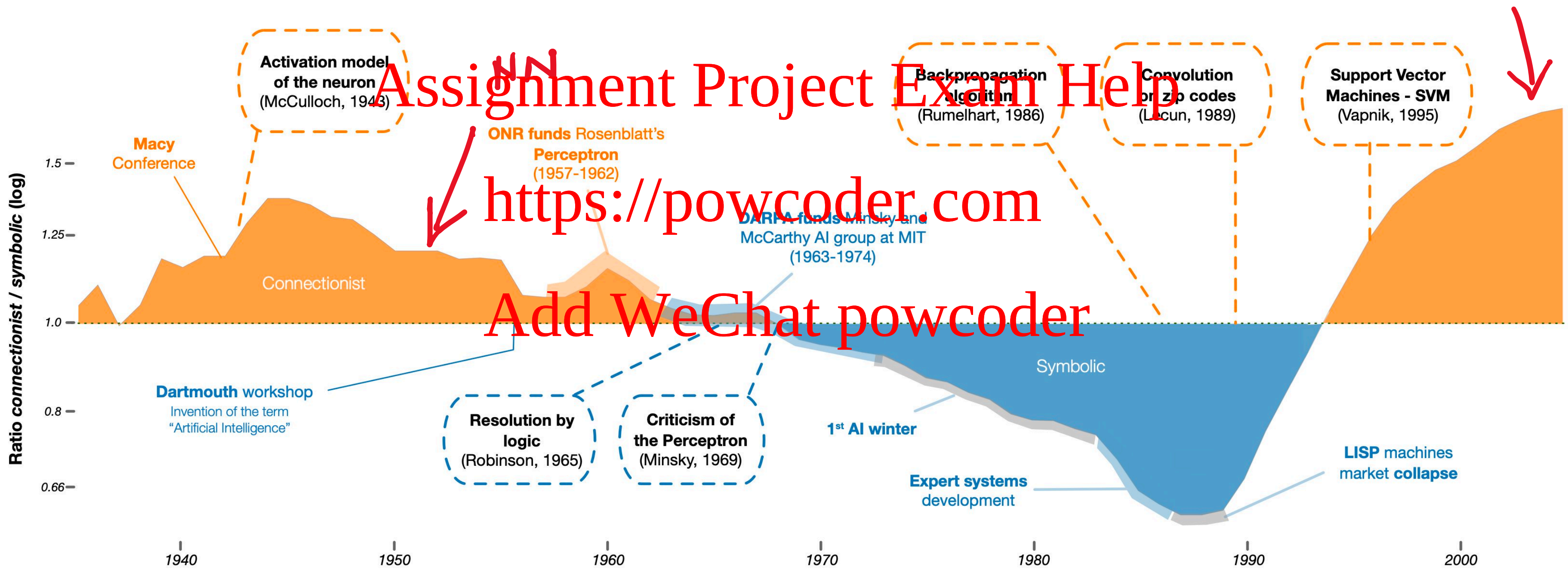
MATURITY OF APPROACHES ML



Source: hazyresearch.github.io/snorkel/blog/snorkel_programming_training_data.html

PARADIGMS IN ML

UCL SCHOOL OF MANAGEMENT



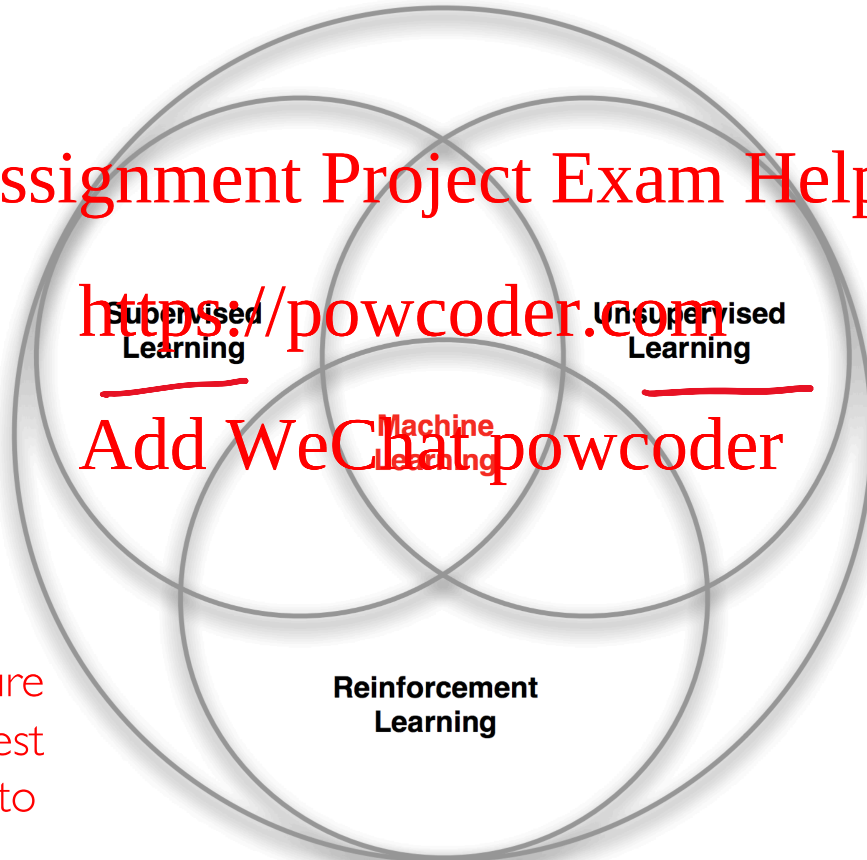
Source: https://twitter.com/IntuitMachine/status/1200796873495318528/photo/1

MACHINE LEARNING

BRANCHES

We know what the right answer is

$\{x, y\}$



We don't know what the right answer is – but we can recognize a good answer if we find it

$\{x, ?\}$

We have a way to measure how good our current best answer is, and a method to improve it

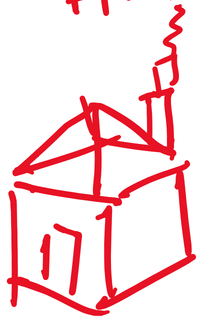
Source: Introduction to Reinforcement Learning, David Silver

BUILDING BLOCKS OF ML

UCL SCHOOL OF MANAGEMENT

Chapt 2

HML



A - B - C- D

A TAXONOMY OF PROBLEMS

A. ClAssification

B. Regression

B

week 2
week 3

C. Clustering

week 5

D. Decomposition

week 4

A - B - C- D

ALGORITHMIC APPROACHES

UCL SCHOOL OF MANAGEMENT

A. ClAssification

- Support vector machines
- Neural networks
- Random Forests
- Maximum entropy classifiers
- …

CART

B. Regression

- Logistic regression
- Support vector regression
- SGD regressor
- …

tree

100s

C. Clustering

- K-means
- KD Trees
- Spectral clustering
- Density estimation
- …

D. Decomposition

- PCA
- LDA
- t-SNE
- Umap
- VAE
- …

A - B - C- D

ALGORITHMIC APPROACHES

UCL SCHOOL OF MANAGEMENT

A. ClAssification

- Support vector machines
- Neural networks
- Random Forests
- Maximum entropy classifiers
- …

B. Regression

- Logistic regression
- Support vector regression
- SGD regressor
- …

Supervised

C. Clustering

- K-means
- KD Trees
- Spectral clustering
- Density estimation
- …

D. Decomposition

- PCA
- LDA
- t-SNE
- Umap
- VAE
- …

Unsupervised

A - B - C- D

ALGORITHMIC APPROACHES

A. ClAssification

Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

category

B. Regression

Model requirements

Assignment Project Exam Help

Regression
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

https://powcoder.com

quatity

Add WeChat powcoder

Supervised

C. Clustering

Model requirements

Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

D. Decomposition

Model requirements

Decomposition
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Unsupervised

A - B - C- D

ALGORITHMIC APPROACHES

A. ClAssification

Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

B. Regression

Model requirements

Regression
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

We know what the right answer is

Supervised

$x \rightarrow y$

C. Clustering

Model requirements

Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

D. Decomposition

Model requirements

Decomposition
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Unsupervised

A. ClAssification

Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

B. Regression

Model requirements

Regression
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Supervised

C. Clustering

Model requirements

Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

D. Decomposition

Model requirements

Decomposition
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Unsupervised

We don't know what the right answer is – but we can recognize a good answer if we find it

A - B - C- D
ALGORITHMIC APPROACHES

UCL SCHOOL OF MANAGEMENT

A. ClAssification

Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

B. Regression

Model requirements

Regression
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

C. Clustering

Model requirements

Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

D. Decomposition

Model requirements

Decomposition
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

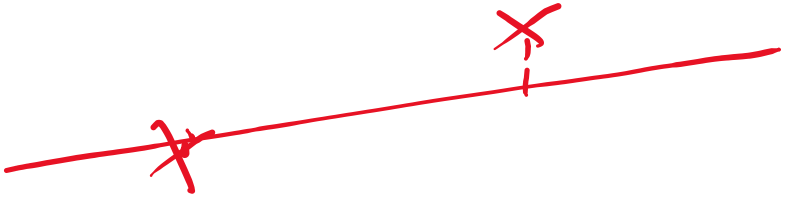
Supervised

We have a way to measure how good our current best answer is

Reinforcement Learning

Unsupervised

B. Regression



B. REGRESSION

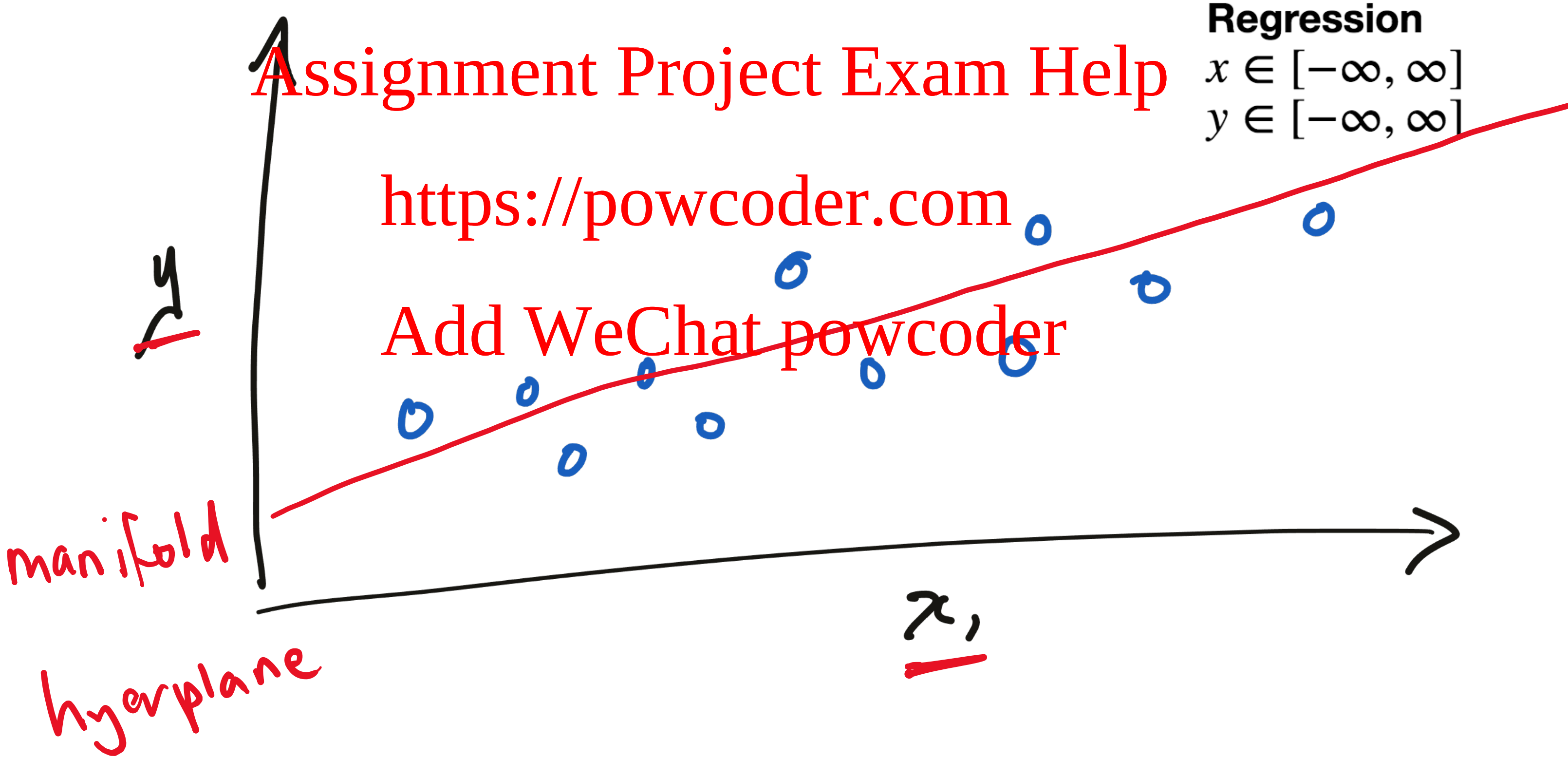
REAL VALUED VARIABLE

Model requirements

Regression

$x \in [-\infty, \infty]$

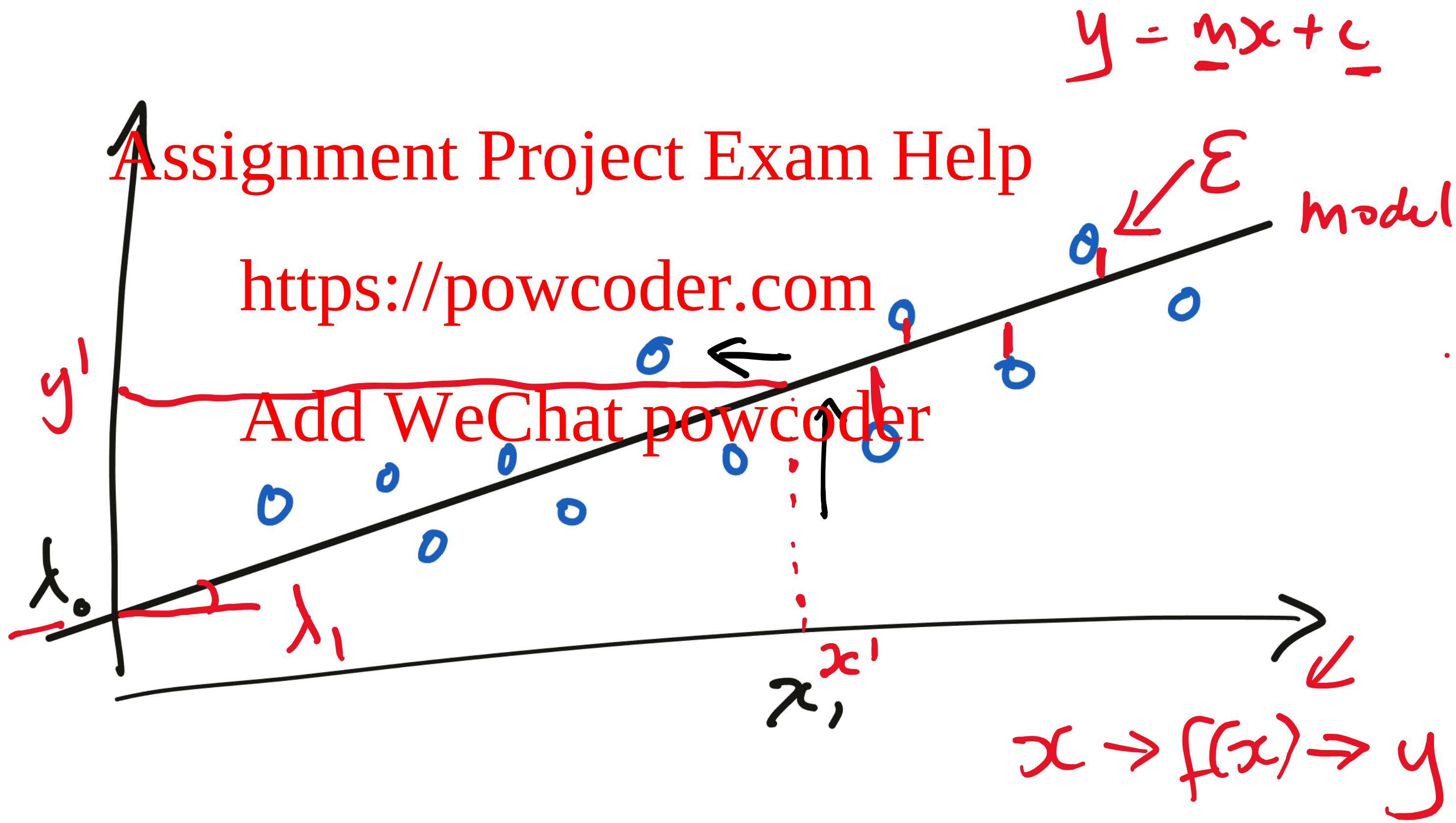
$y \in [-\infty, \infty]$



B. REGRESSION

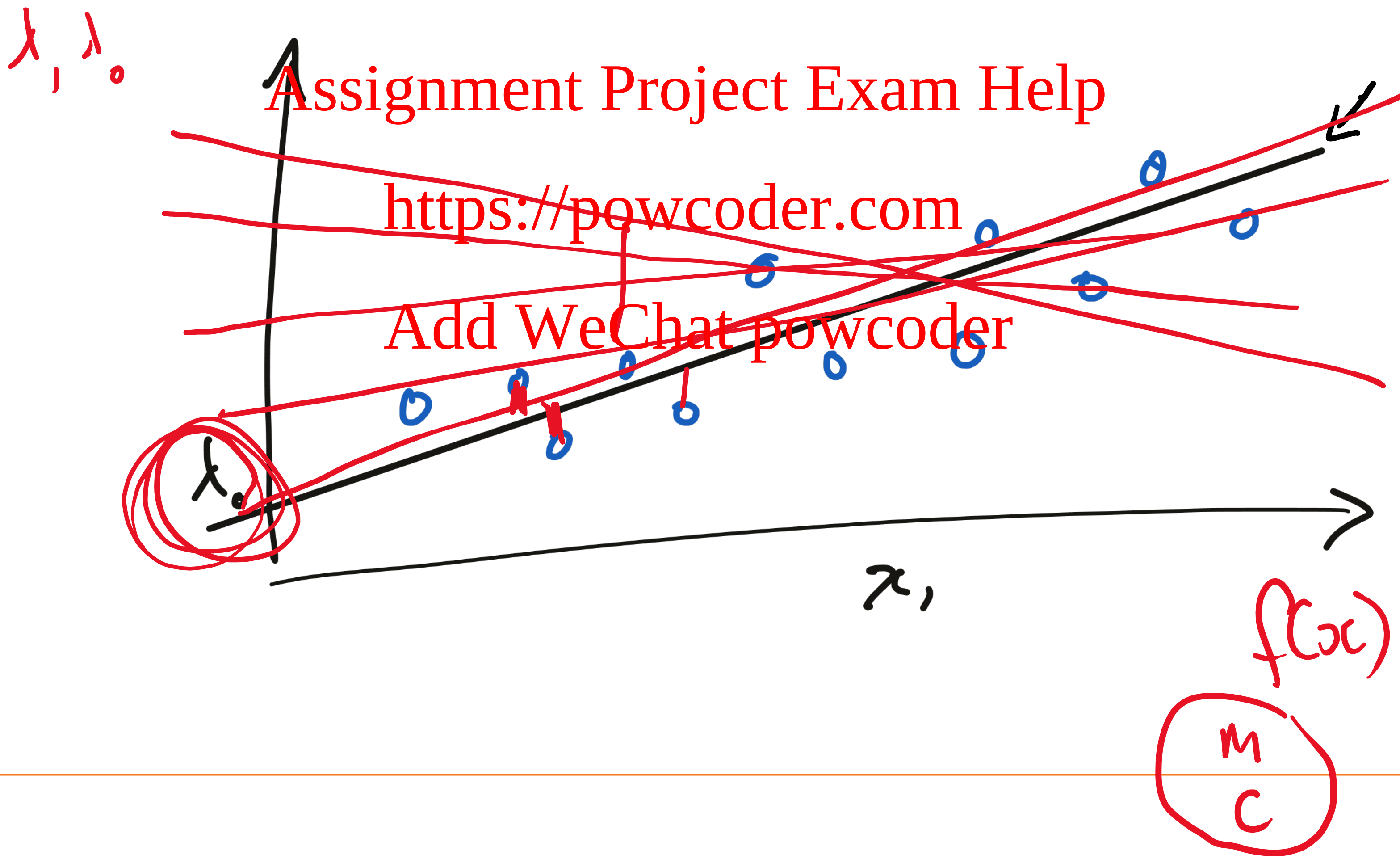
REAL VALUED VARIABLE

UCL SCHOOL OF MANAGEMENT



B. REGRESSION

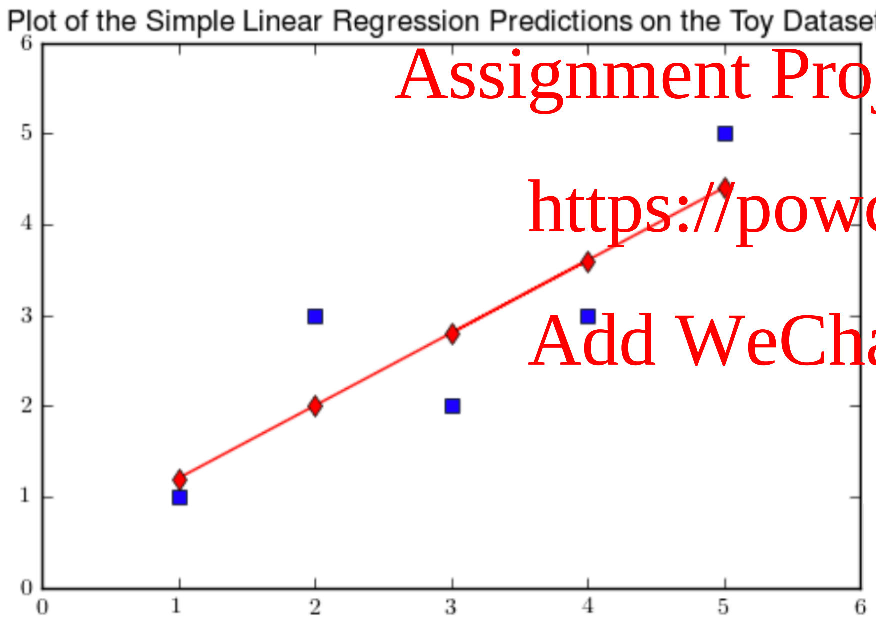
REAL VALUED VARIABLE



LINEAR REGRESSION

Algorithm Taxonomy

Discriminative
Parametric
Linear
Deterministic



In simple linear regression we can use statistics on the training data to estimate the coefficients required by the model to make predictions on new data. The line for a simple linear regression model can be written as:

$$y = \beta_0 + \beta_1 x$$

Where β_0 and β_1 are the coefficients we must estimate from the training data.

The variables are commonly referred to the following:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Figure. 2 Graph to show the trend line on predicted data for the Toy Dataset

REGRESSION BY MODELING PROBABILITIES

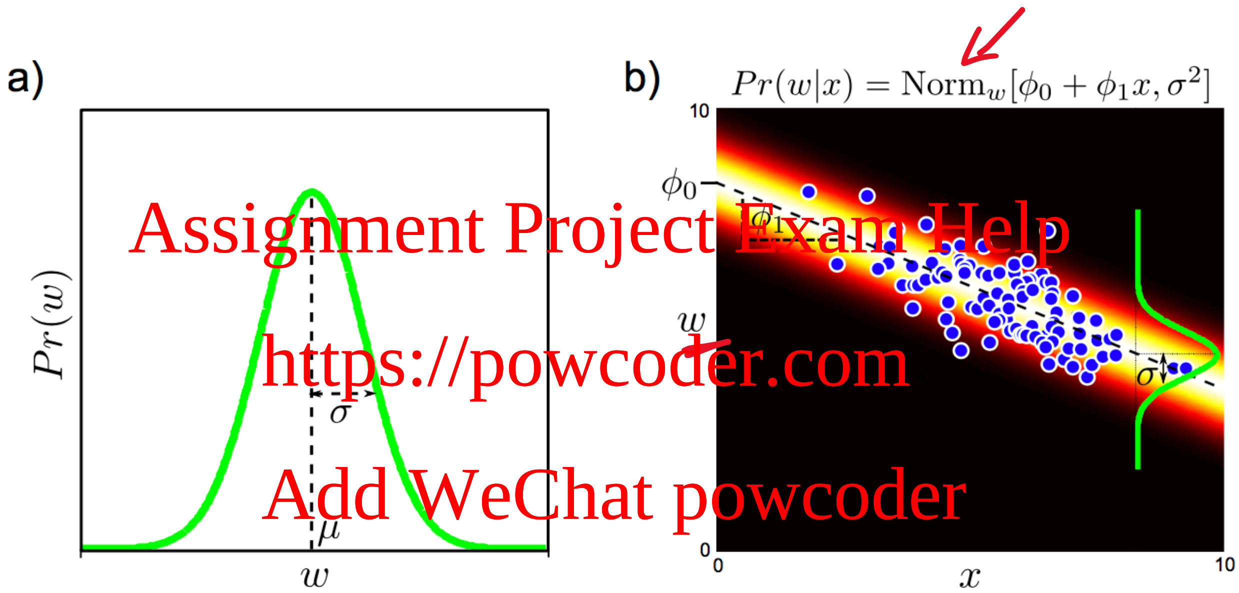
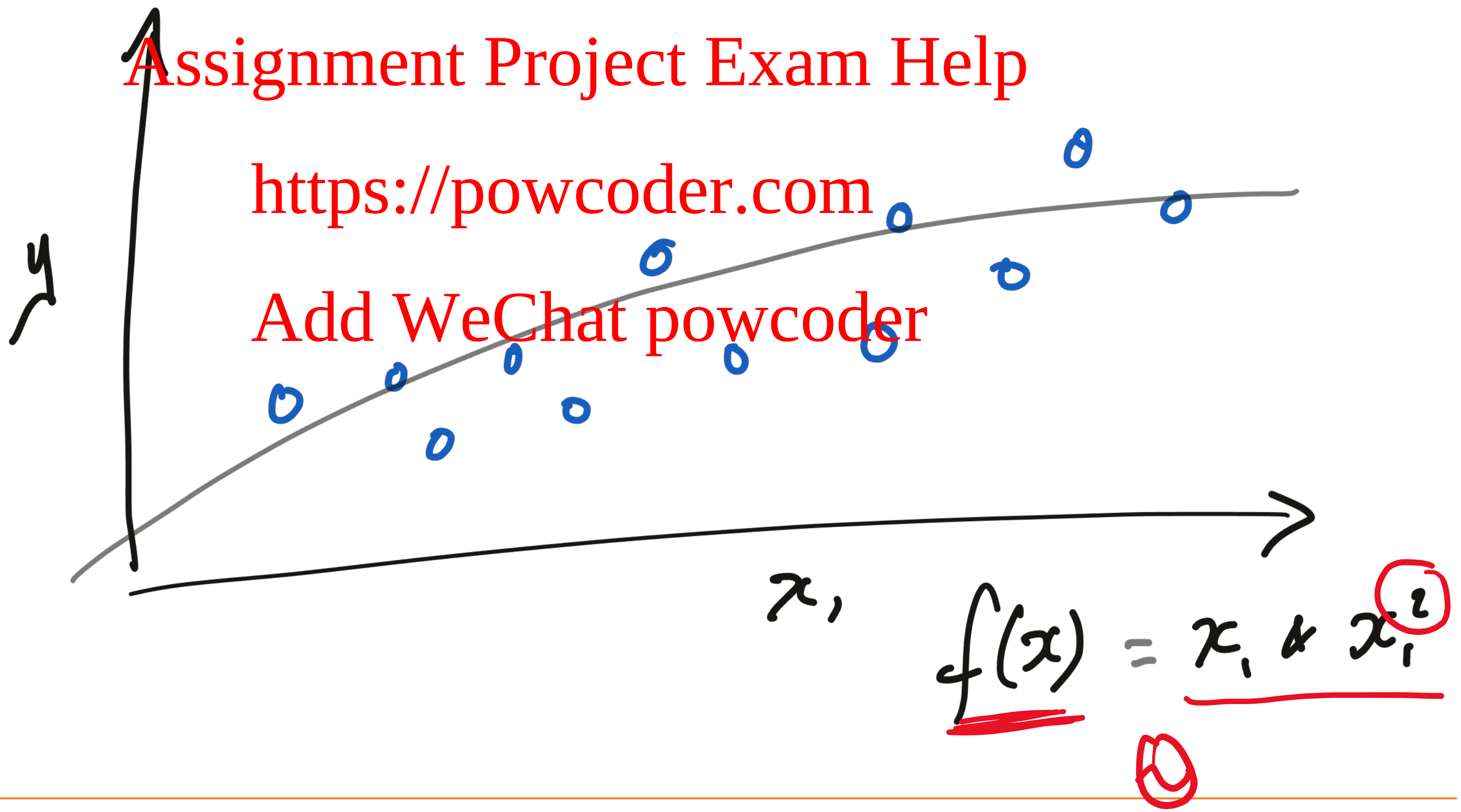


Figure 6.1 Regression by modeling the posterior $Pr(w|x)$ (discriminative). a) We model the world state w as normally distributed. b) We make the normal parameters a function of the observations x: the mean is a linear function $\mu = \phi_0 + \phi_1 x$ of the observations, and the variance σ^2 is fixed. The learning algorithm fits the parameters $\boldsymbol{\theta} = \{\phi_0, \phi_1, \sigma^2\}$ to example training pairs $\{x_i, w_i\}_{i=1}^{I}$ (blue dots). In inference we take a new observation x and compute the posterior distribution $Pr(w|x)$ over the state.

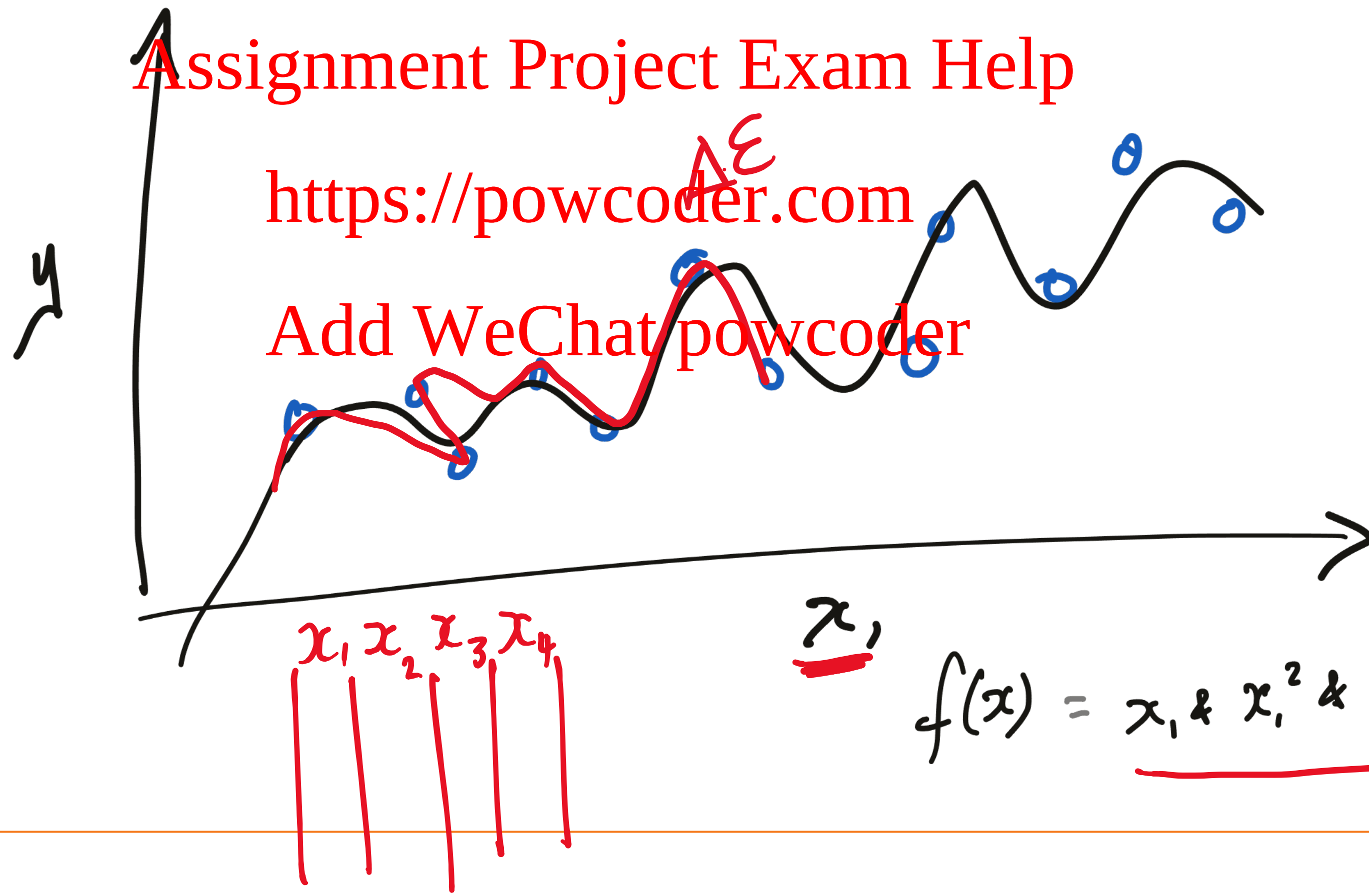
B. REGRESSION

REAL VALUED VARIABLE



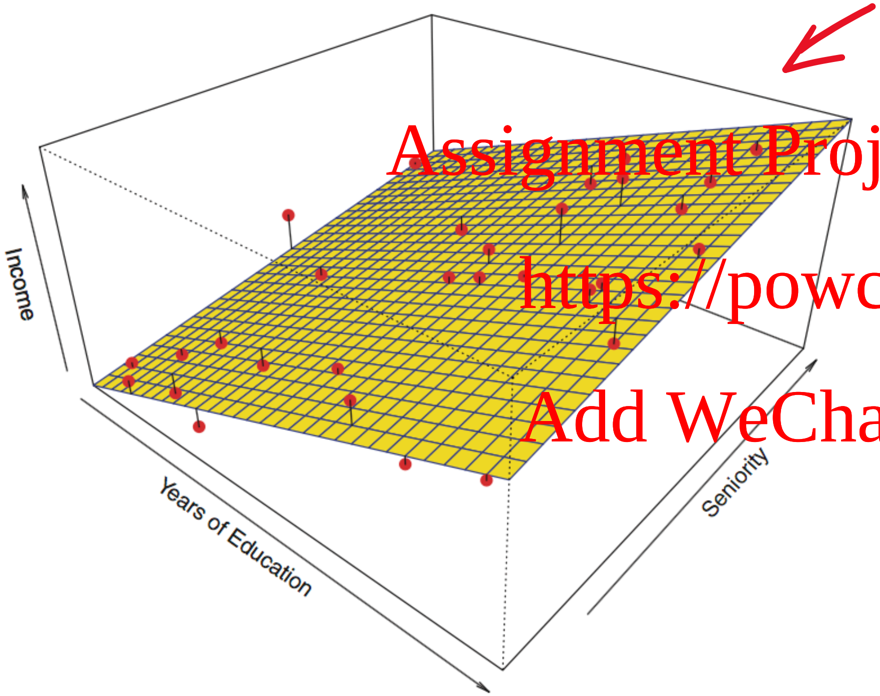
B. REGRESSION

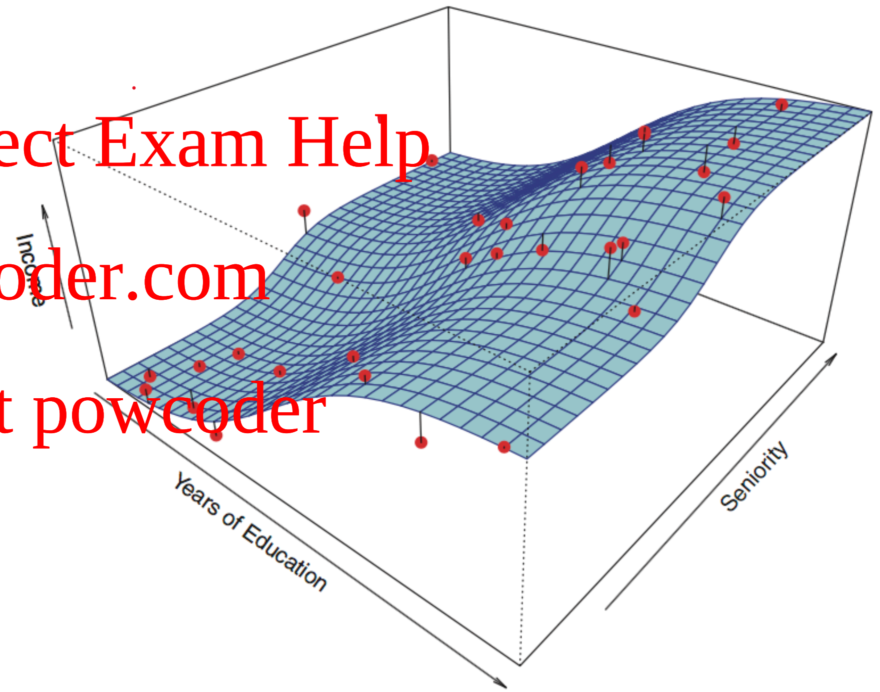
REAL VALUED VARIABLE



$$f(x) = \underline{x_1 \,\&\, x_1^2 \,\&\, x_1^3 \dots}$$

MULTIPLE DIMENSIONS



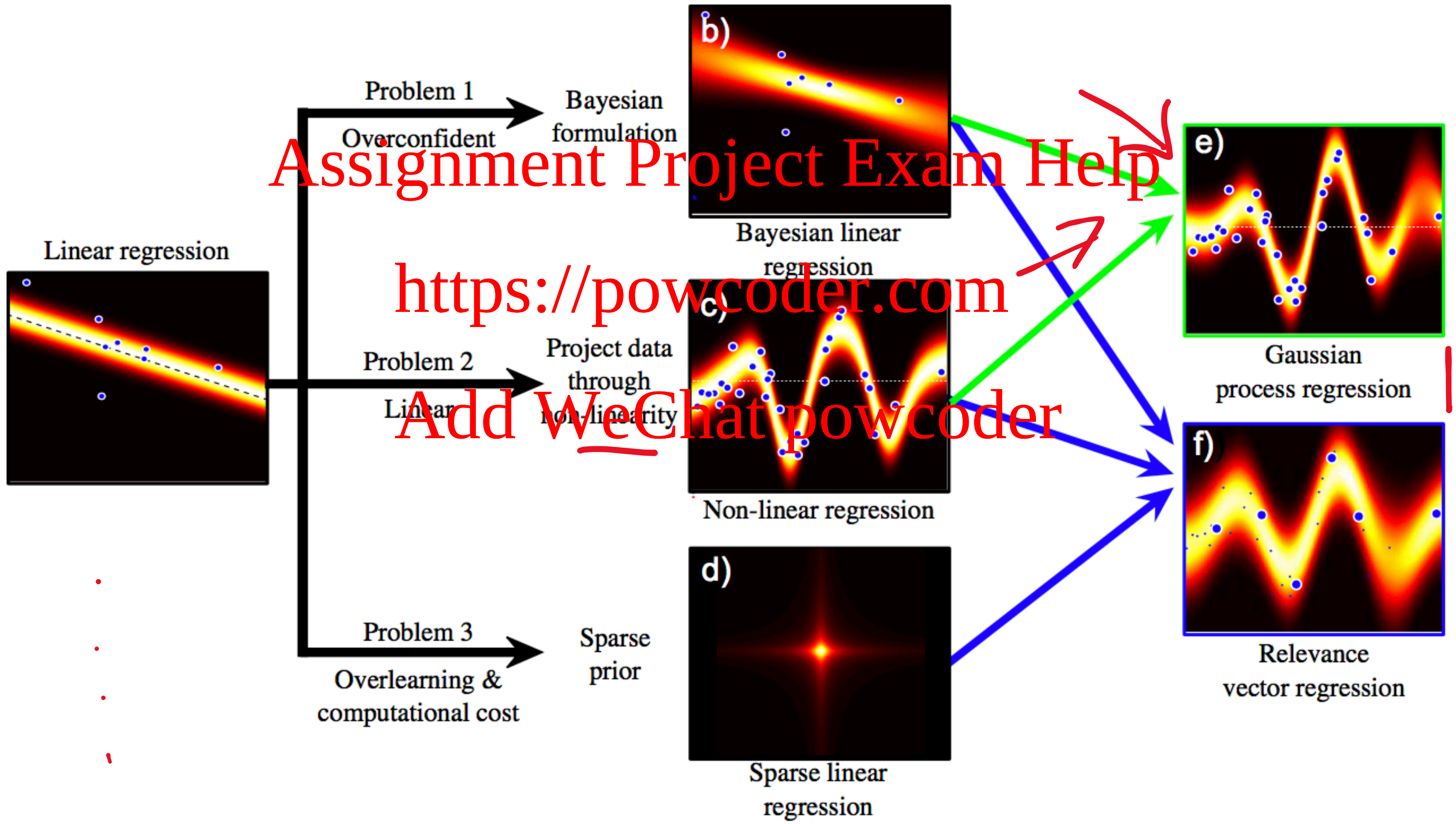


1000's dim

hyper cube

DEVELOPING MORE COMPLEX ALGORITHMS

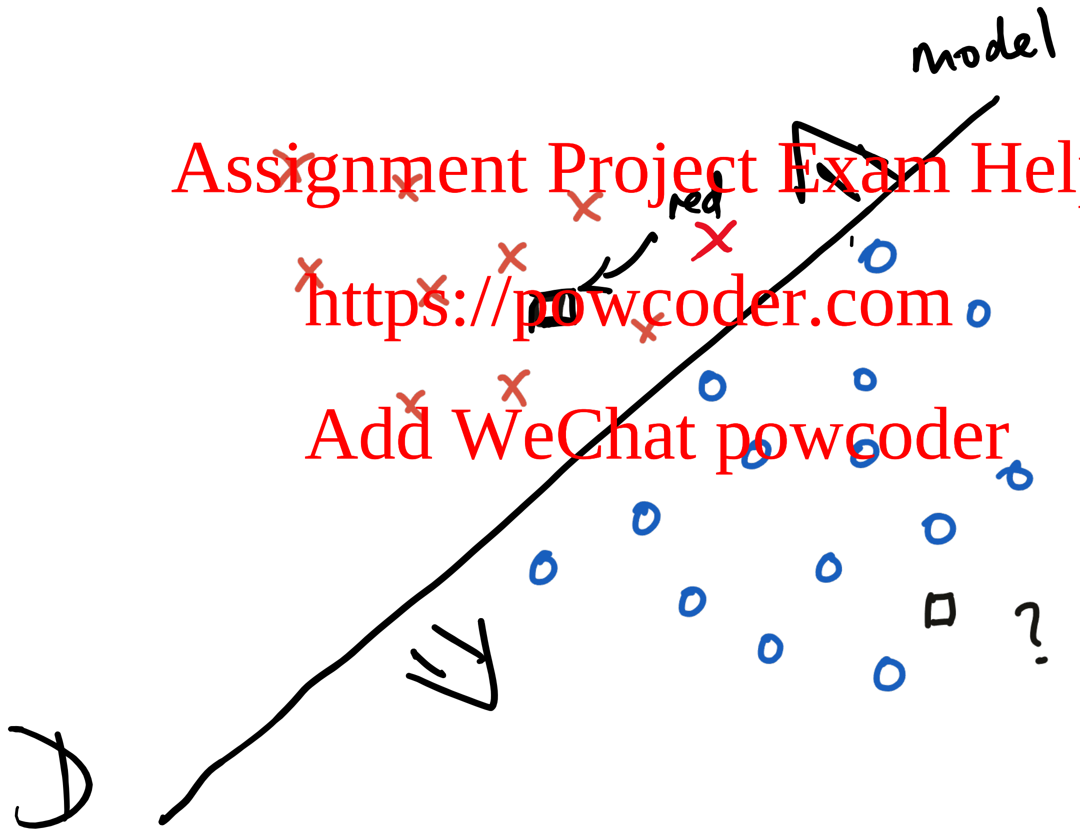
B



A. Classification

A. CLASSIFICATION

CATEGORICAL VARIABLE



Model requirements

Classifiation

$x \in [-\infty, \infty]$

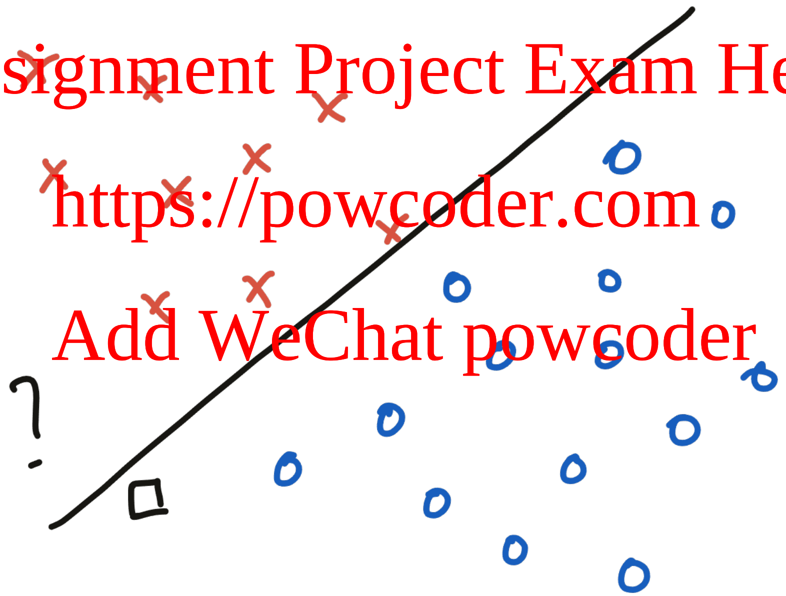
$y \in \{0, \underline{N}\}$

2

decision surface

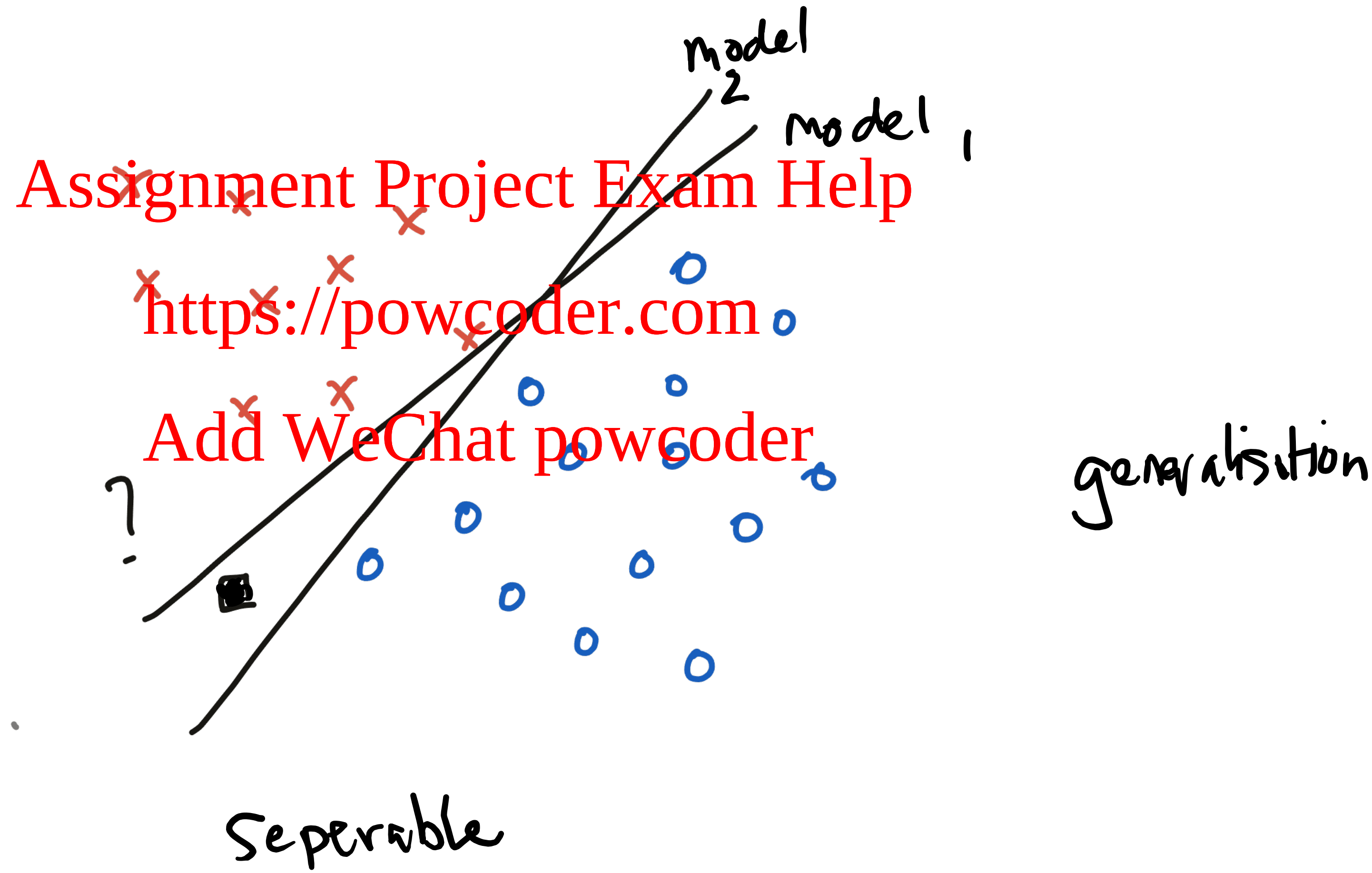
A. CLASSIFICATION

CATEGORICAL VARIABLE



A. CLASSIFICATION

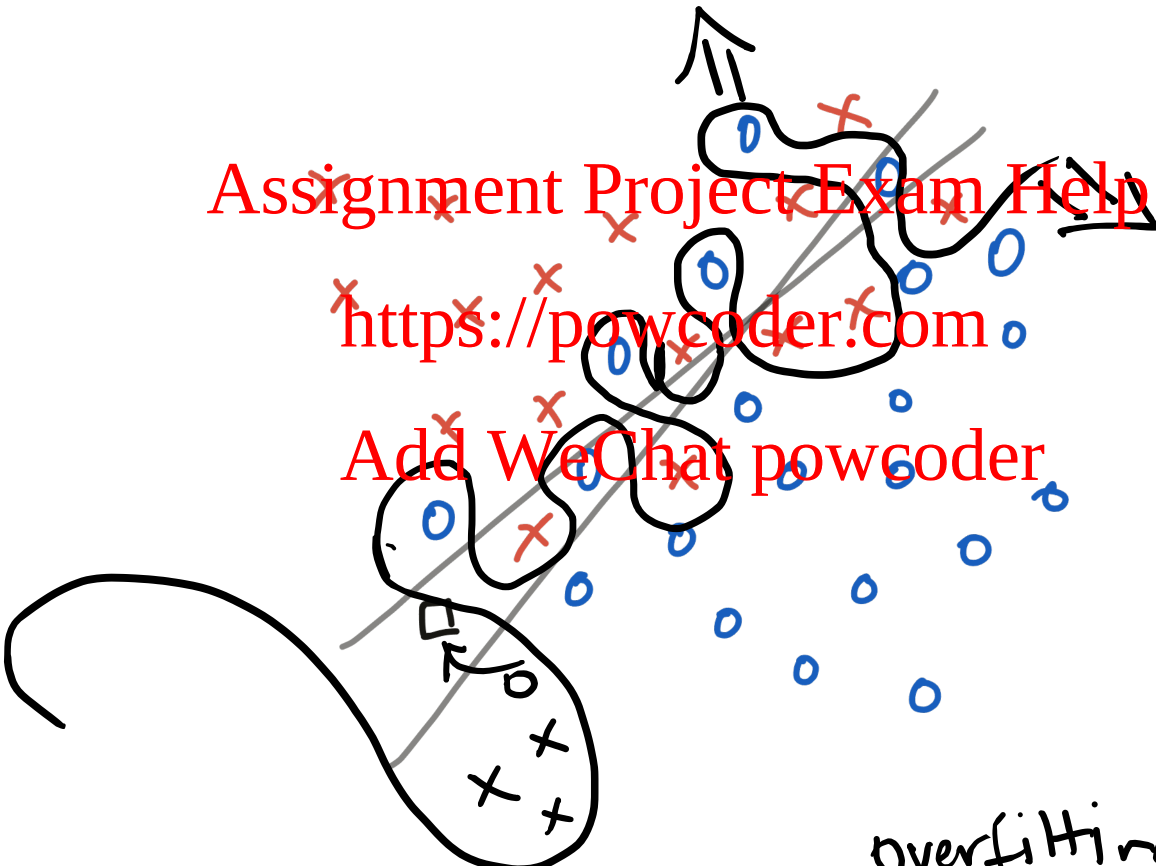
CATEGORICAL VARIABLE



A. CLASSIFICATION

CATEGORICAL VARIABLE

turtuous



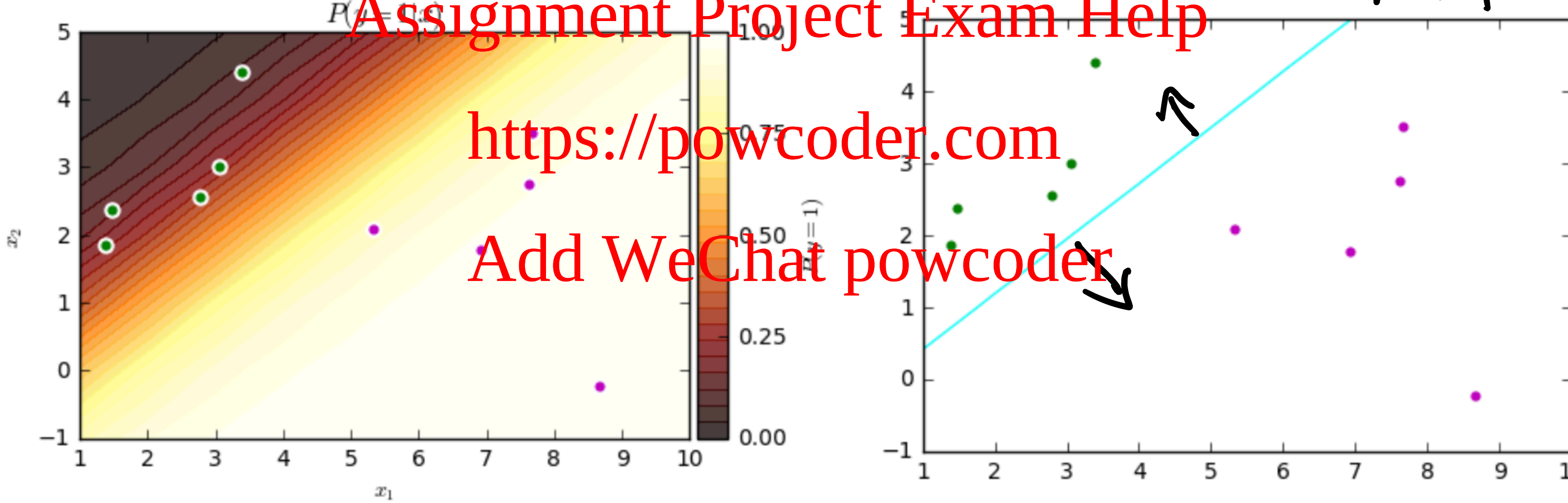
overfitting

train
dev
test

LOGISTIC REGRESSION

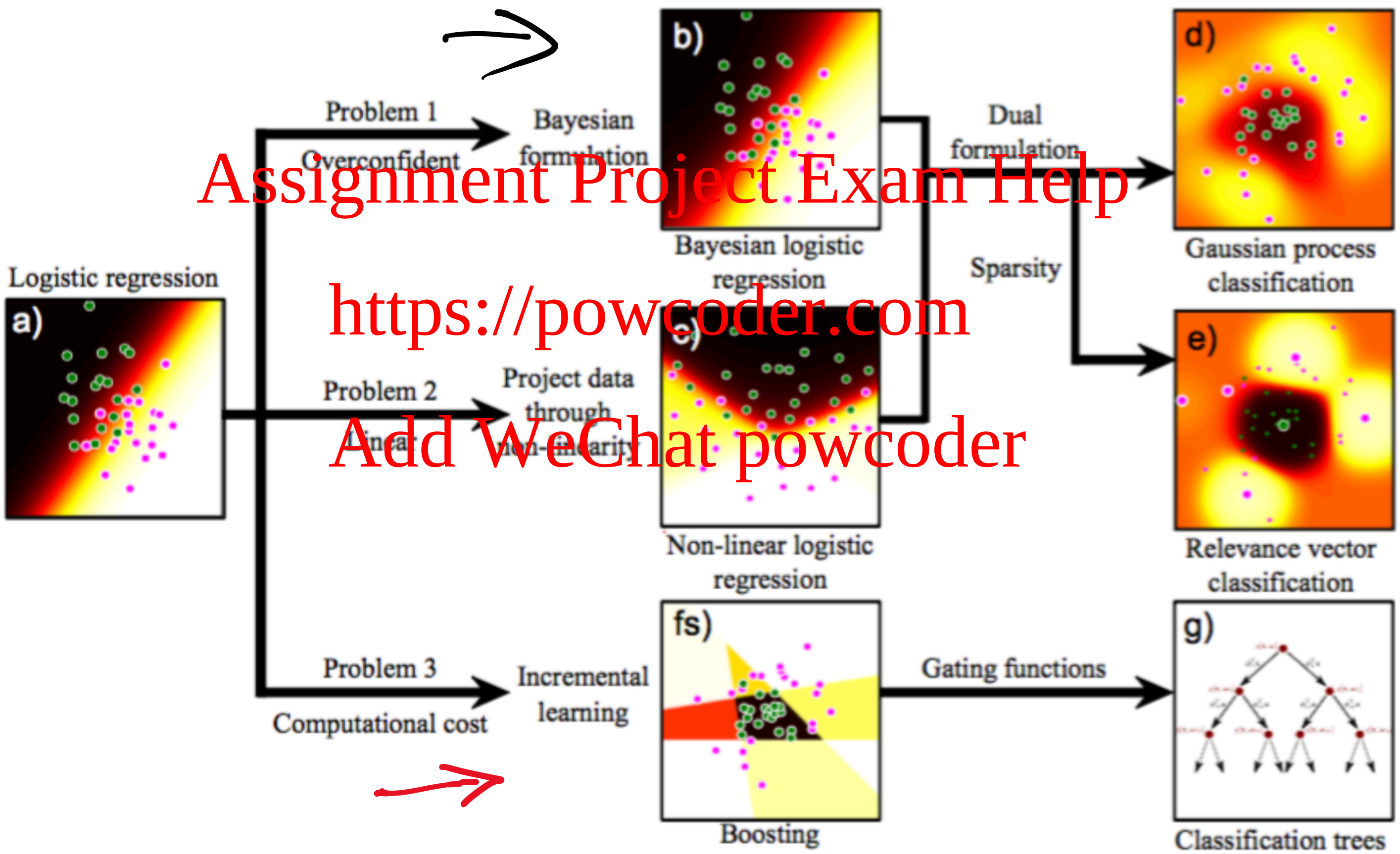
Algorithm Taxonomy
Discriminative
Parametric
Linear
Deterministic

manifold



DEVELOPING MORE COMPLEX ALGORITHMS

UCL SCHOOL OF MANAGEMENT



week 3

CONFUSION MATRIX

BINARY FORCED CHOICE

UCL SCHOOL OF MANAGEMENT

predicted

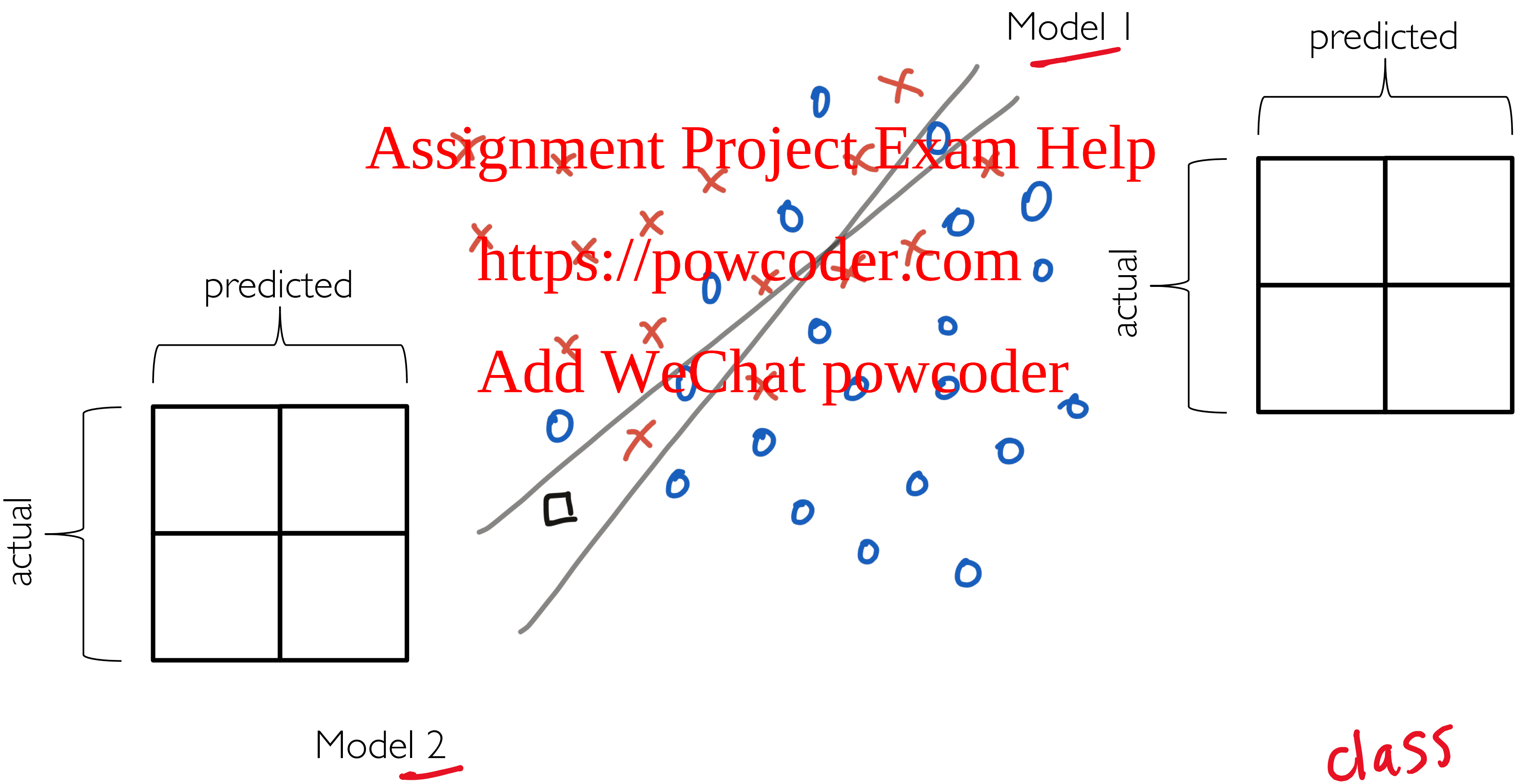
actual

TP	FN	P
FP	TN	N

	1	0
1	TP	
0		TN

A. CLASSIFICATION

CATEGORICAL VARIABLE



CLASSIFICATION
MNIST DATASET

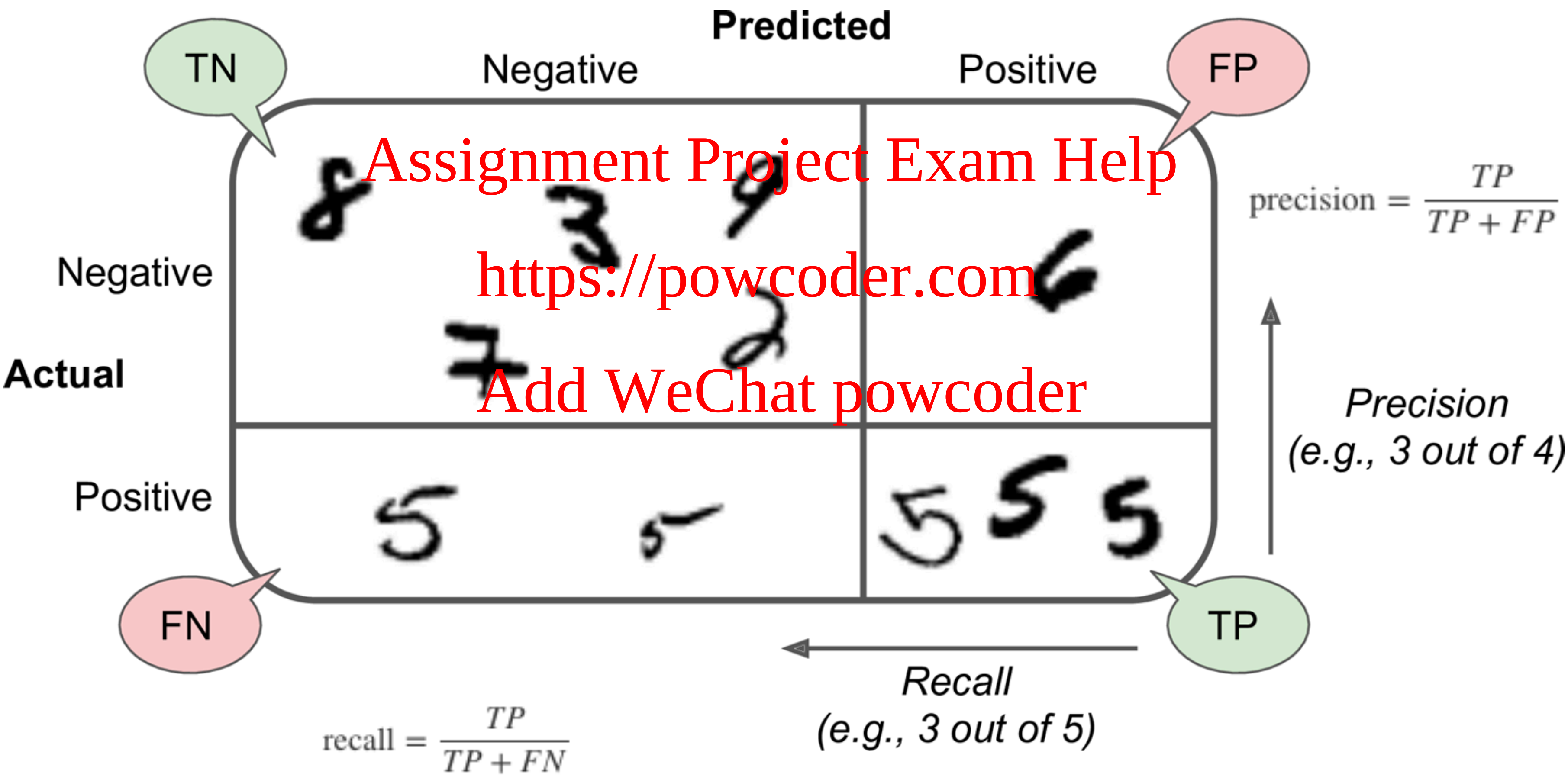
UCL
SCHOOL OF
MANAGEMENT

class

CONFUSION MATRIX

UCL SCHOOL OF MANAGEMENT

class

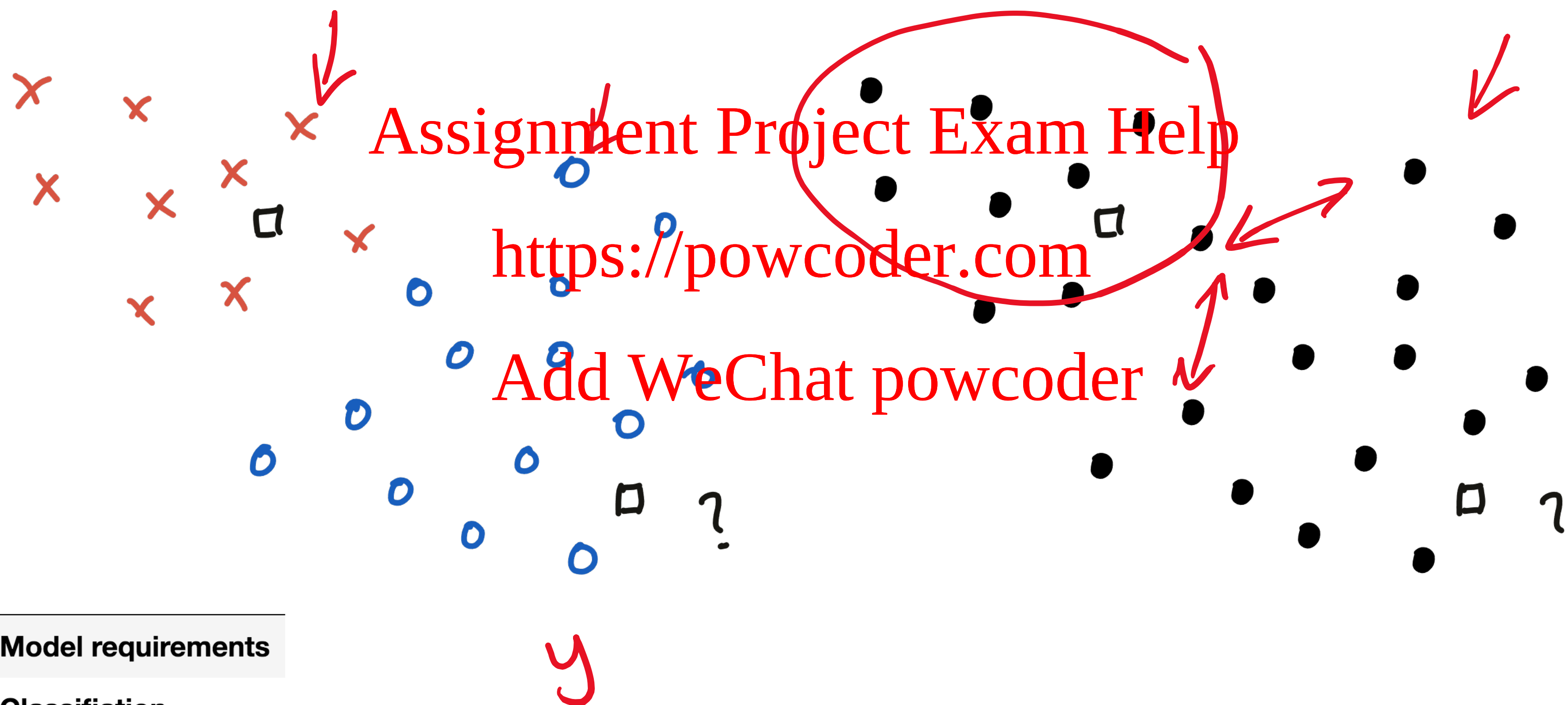


C. Clustering

CLASSIFICATION VS CLUSTERING

CATEGORICAL VARIABLE

UCL SCHOOL OF MANAGEMENT

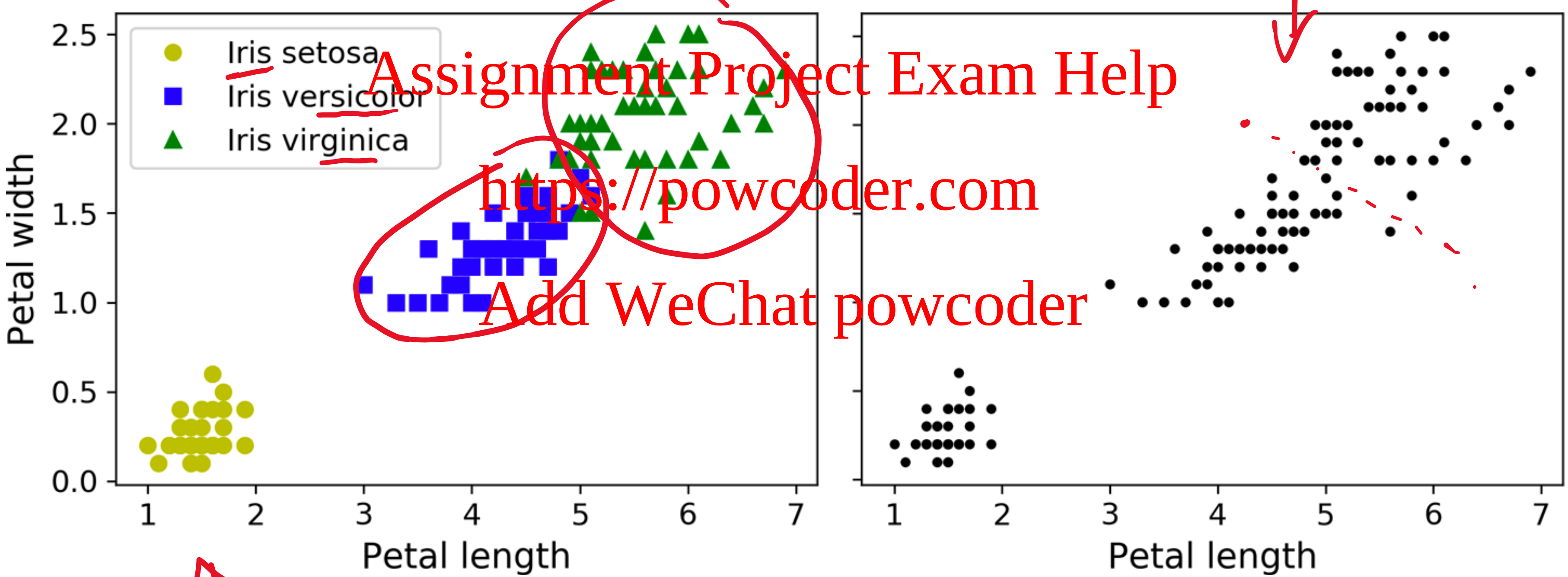


Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

CLASSIFICATION VS CLUSTERING

UCL SCHOOL OF MANAGEMENT



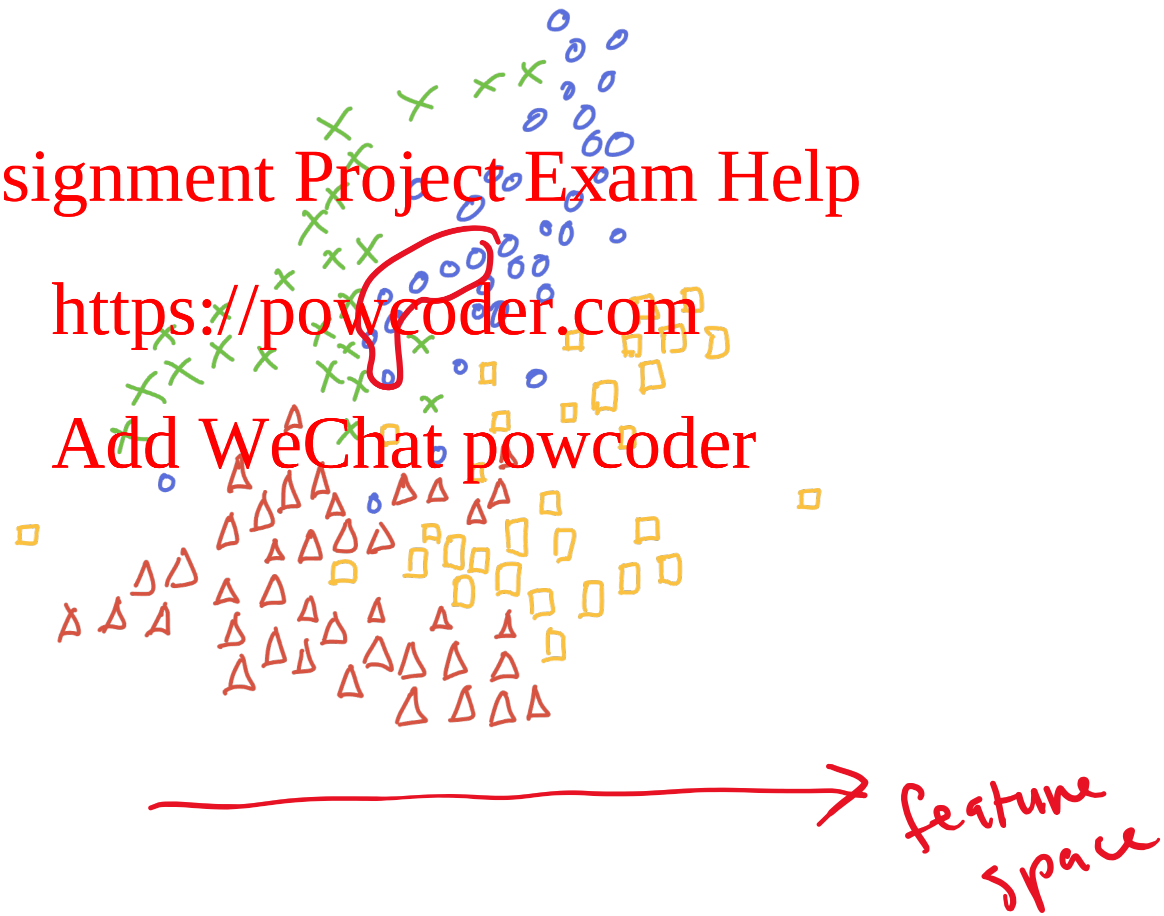
C. CLUSTERING

Model requirements

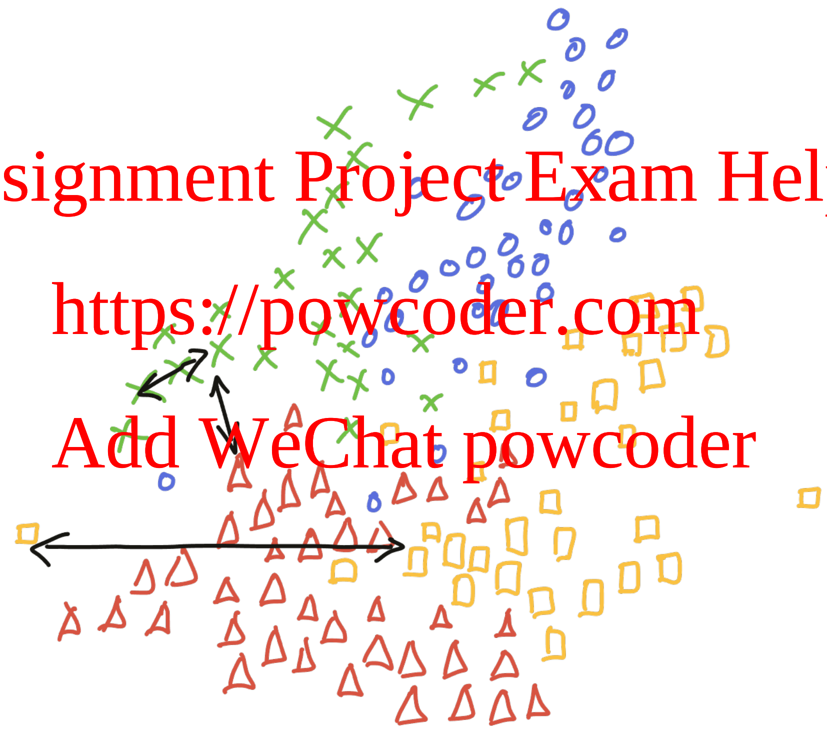
Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

C. CLUSTERING

UCL SCHOOL OF MANAGEMENT



C. CLUSTERING



C. CLUSTERING

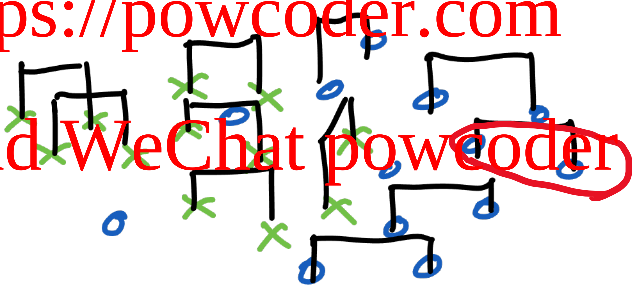
I. AGGLOMERATIVE

UCL SCHOOL OF MANAGEMENT

C. CLUSTERING

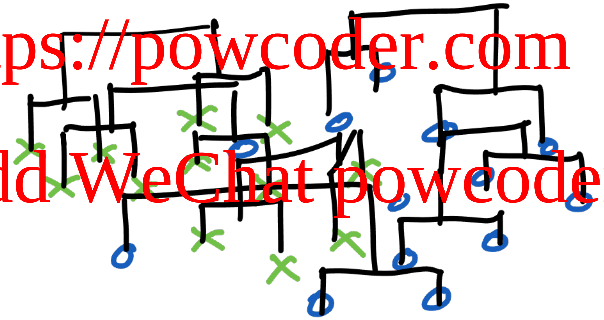
1. AGGLOMERATIVE

UCL SCHOOL OF MANAGEMENT



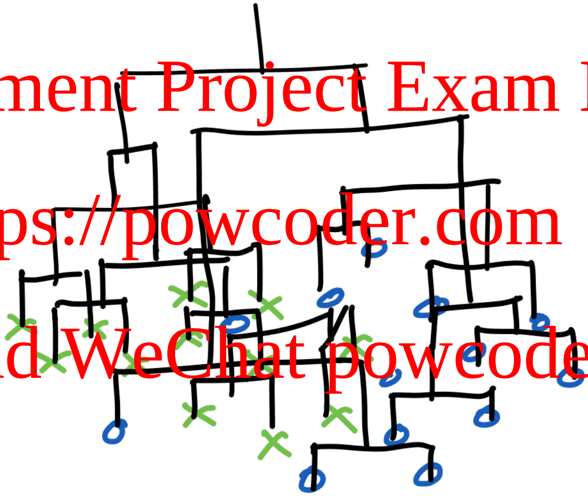
C. CLUSTERING

1. AGGLOMERATIVE



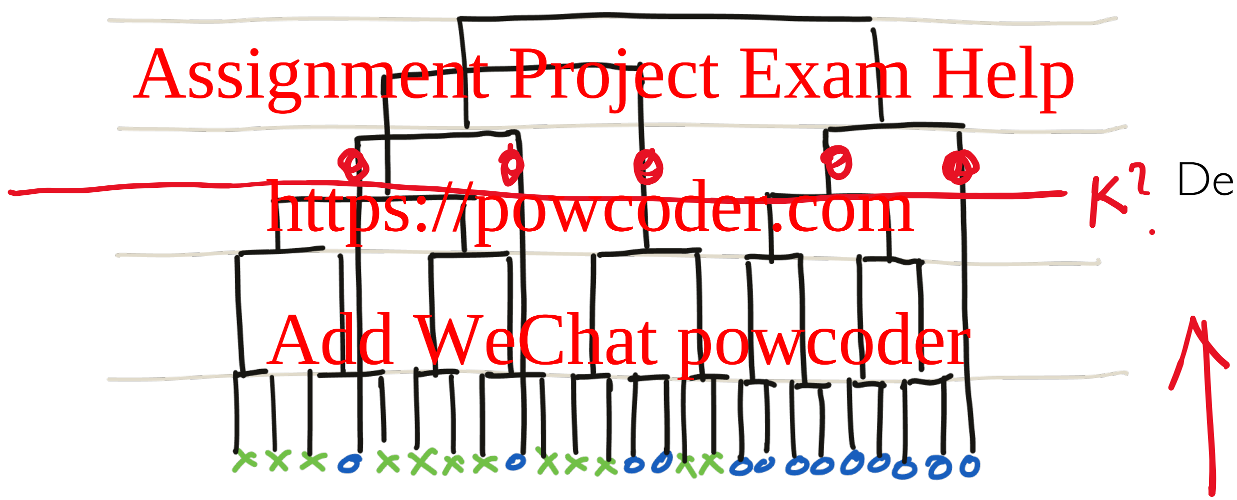
C. CLUSTERING

1. AGGLOMERATIVE



C. CLUSTERING

1. AGGLOMERATIVE



Dendrogram

C. CLUSTERING

2. DIVISIVE

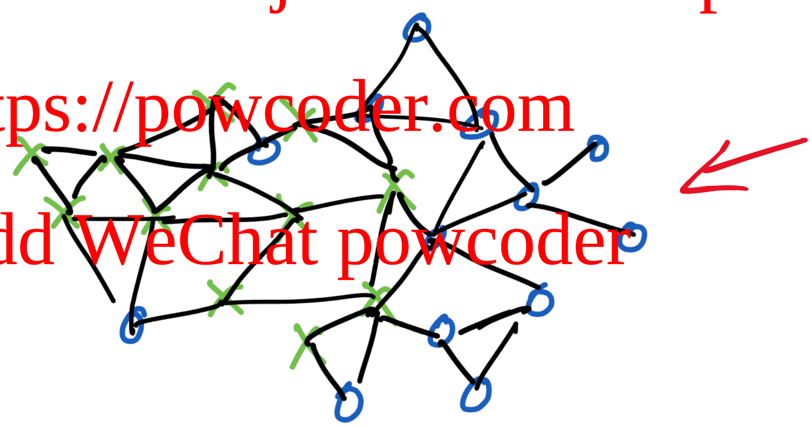
UCL
SCHOOL OF
MANAGEMENT

C. CLUSTERING

2. DIVISIVE

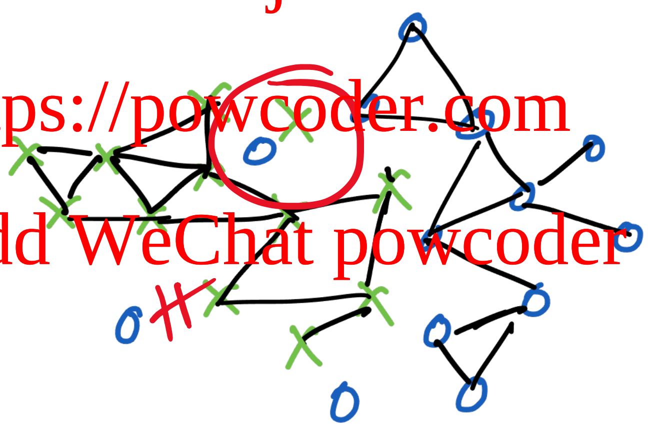
C. CLUSTERING

2. DIVISIVE



C. CLUSTERING

2. DIVISIVE



C. CLUSTERING

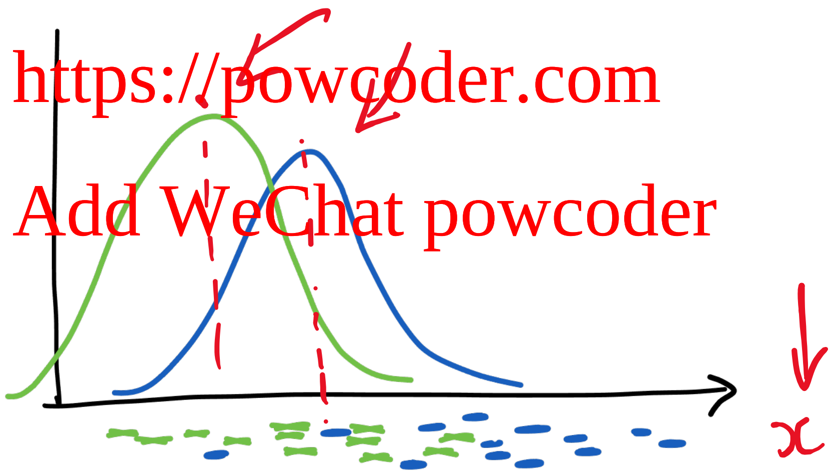
3. PARTITIONAL

UCL SCHOOL OF MANAGEMENT

$k = 2$

C. CLUSTERING

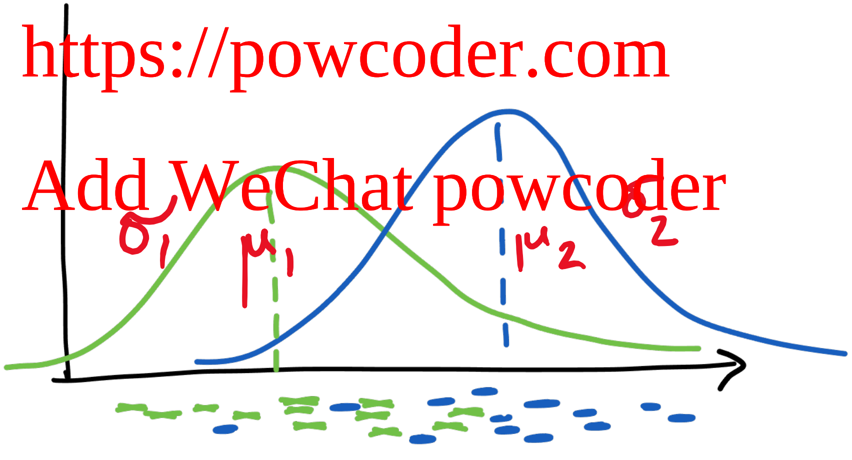
3. PARTITIONAL



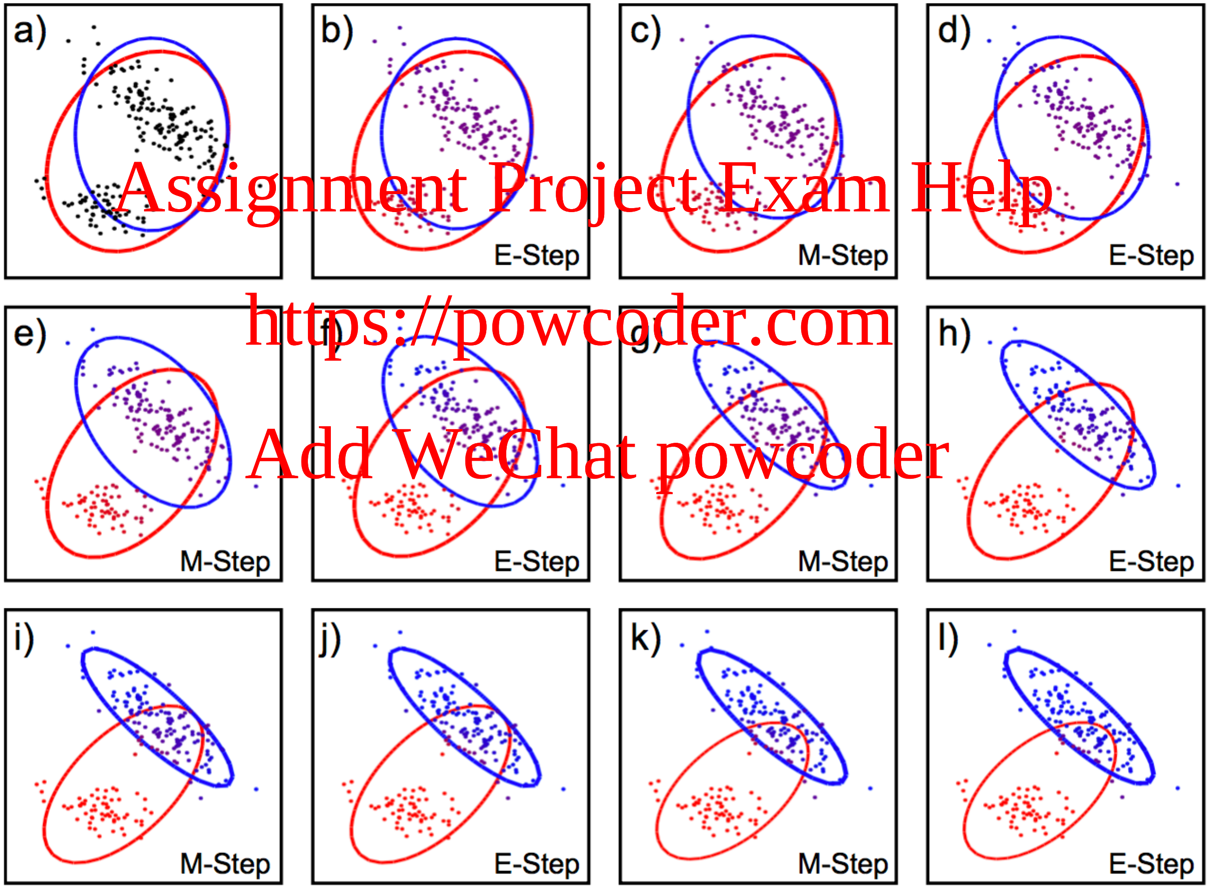
week 5

C. CLUSTERING

3. PARTITIONAL



EXPECTATION MAXIMISATION



D. Decomposition

D. DECOMPOSITION

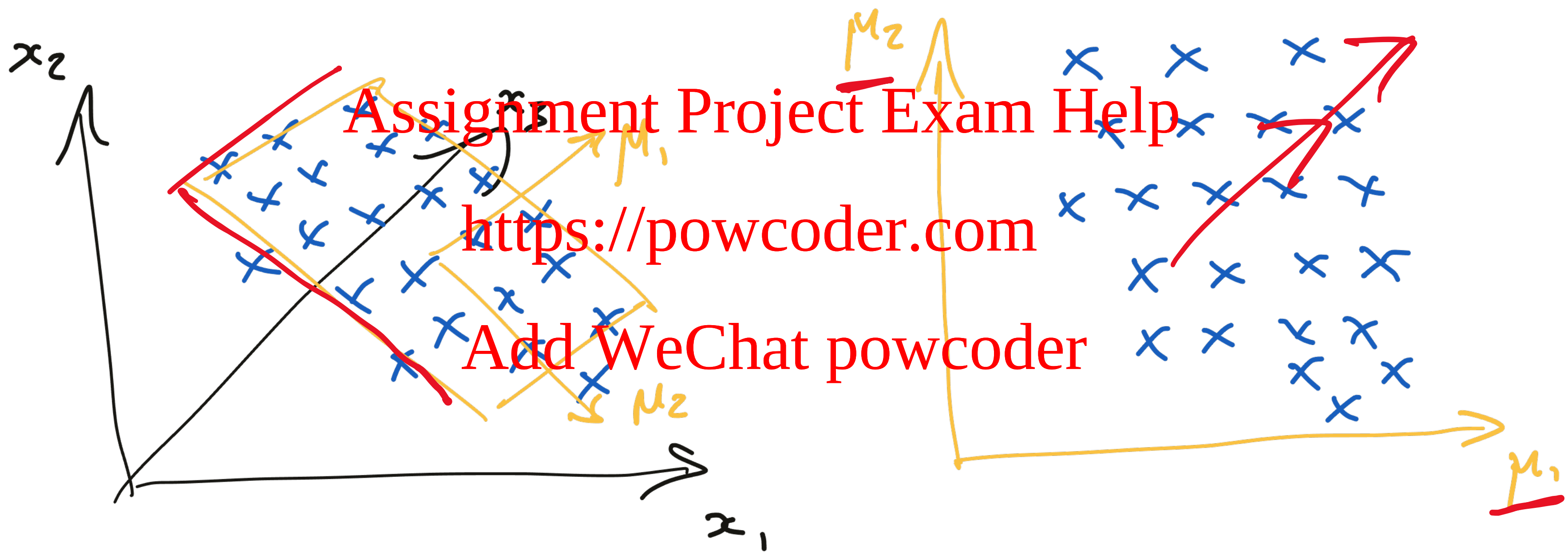
1. PROJECTION METHODS

Model requirements

Decomposition

$x \in [-\infty, \infty]$

$y \in [-\infty, \infty]$

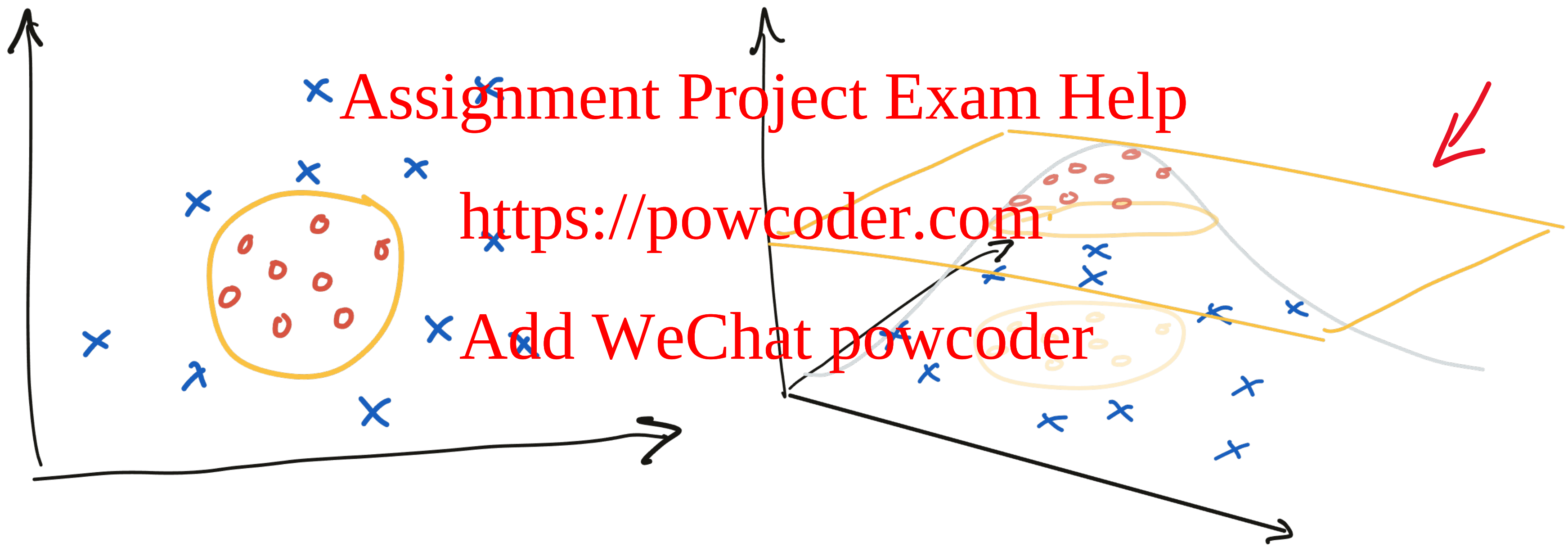


Dimensionality reduction

high dim space $\Rightarrow$ low dim space

D. DECOMPOSITION

2. KERNEL METHODS



low dim

high dim

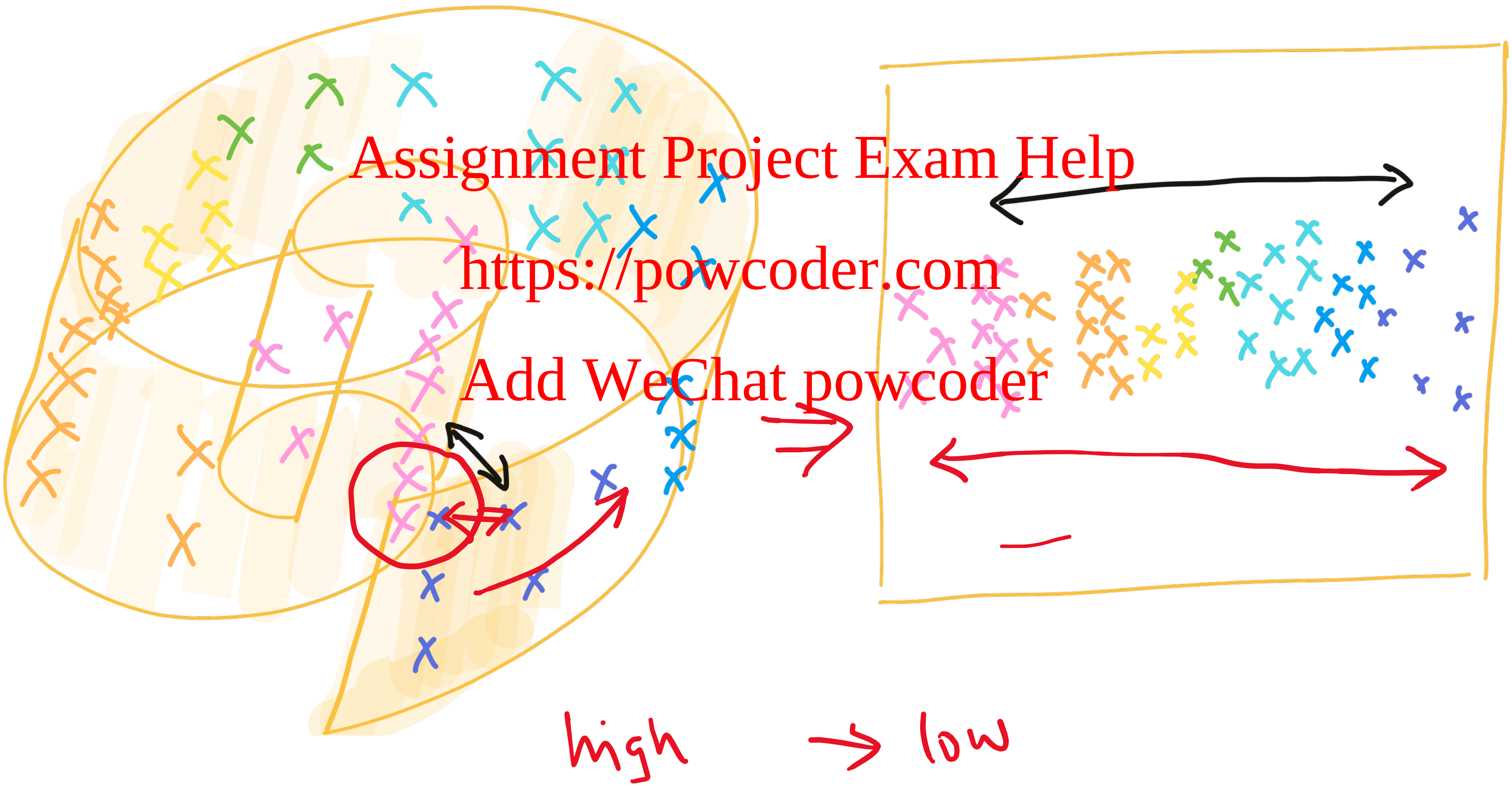
2D → 3D

Kernel

D. DECOMPOSITION

3. MANIFOLD LEARNING

UCL SCHOOL OF MANAGEMENT



A - B - C- D

ALGORITHMIC APPROACHES

UCL SCHOOL OF MANAGEMENT

A. ClAssification

Model requirements

Classifiation

$x \in [-\infty, \infty]$

$y \in \{0, N\}$

B. Regression

Model requirements

Regression

$x \in [-\infty, \infty]$

$y \in [-\infty, \infty]$

C. Clustering

Model requirements

Clustering

$x \in [-\infty, \infty]$

$y \in \{0, N\}$

D. Decomposition

Model requirements

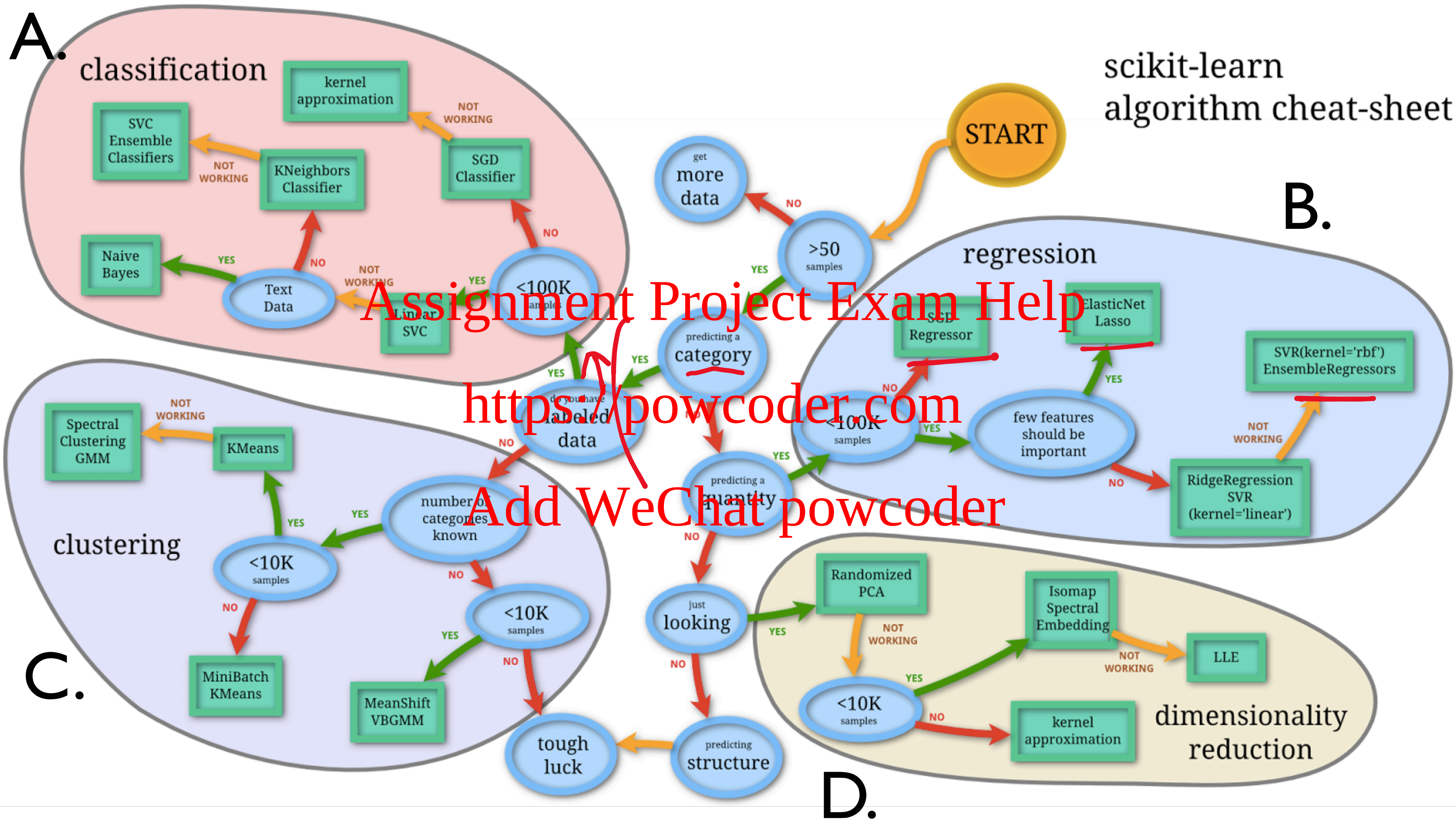
Decomposition

$x \in [-\infty, \infty]$

$y \in [-\infty, \infty]$

TAXONOMY

UCL SCHOOL OF MANAGEMENT



UCL SCHOOL OF MANAGEMENT

A. ClAssification

Model requirements
Classification $x \in [-\infty, \infty]$ $y \in \{0, N\}$

	Discriminative	Generative
Method / Model	$Pr(y\|x)$	$Pr(x\|y)$
Regression $x \in [-\infty, \infty]$ $y \in [-\infty, \infty]$	Linear Regression	Linear Regression
Classifiation $x \in [-\infty, \infty]$ $y \in \{0, 1\}$	Logistic Regression	Probability density function

B. Regression

Model requirements
Regression $x \in [-\infty, \infty]$ $y \in [-\infty, \infty]$

Source: Computer Vision: Learning, Models and Inference

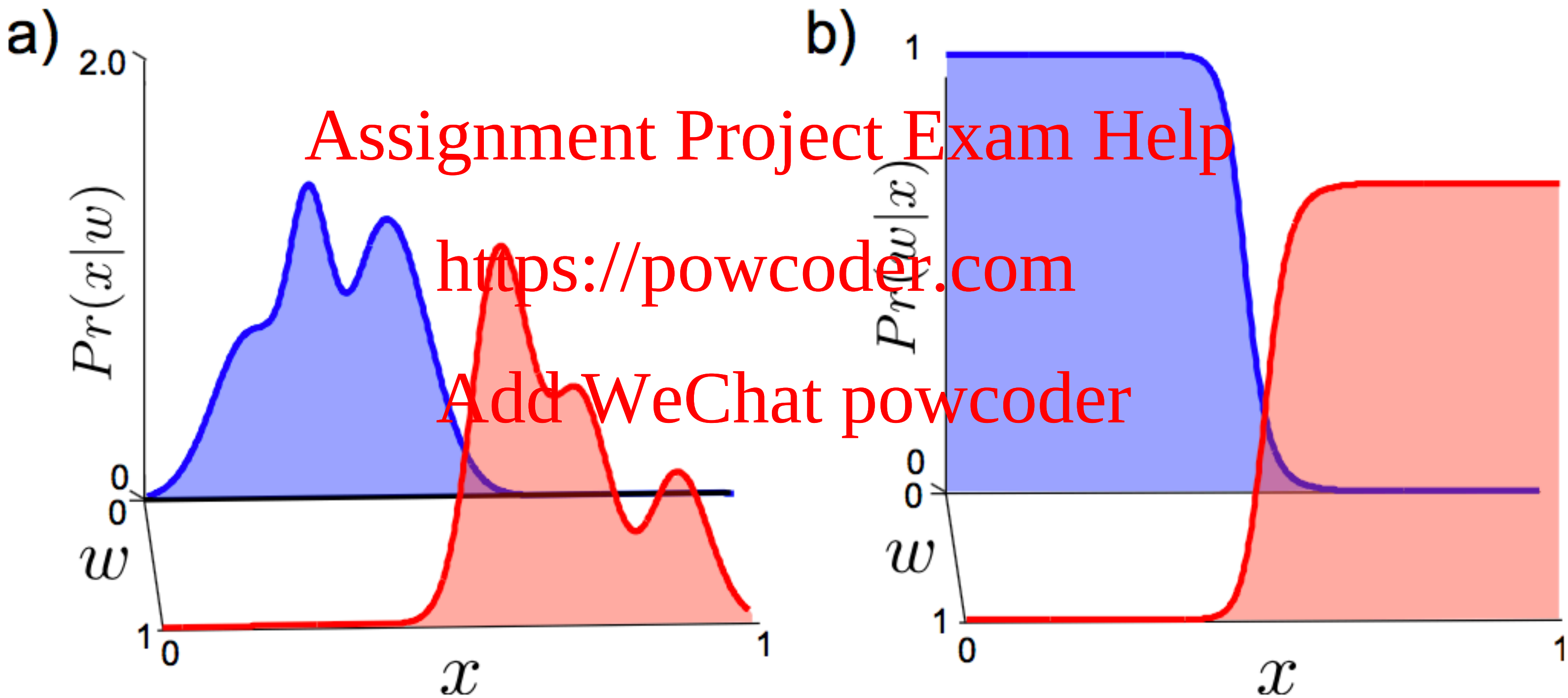
DISCRIMINATIVE VS GENERATIVE

A SIMPLE EXAMPLE

UCL SCHOOL OF MANAGEMENT

class

3 weeks



PARAMETRIC VS NON-PARAMETRIC

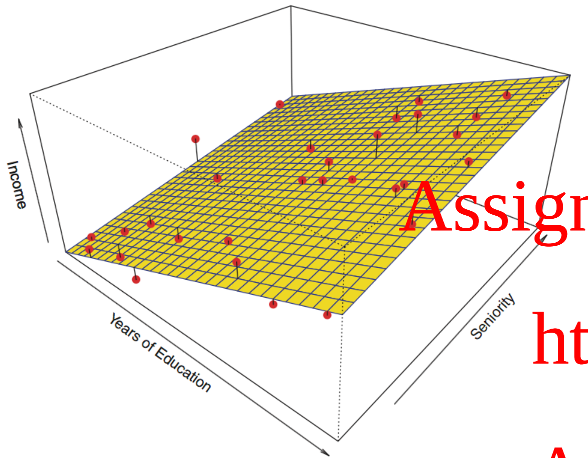
- With data gathered from uncontrolled observations on complex systems involving unknown [physical, chemical, biological, social, economic] mechanisms, the ***a priori*** assumption that nature would generate the data through a parametric model selected by the statistician can result in questionable conclusions that cannot be substantiated by appeal to **goodness-of-fit tests** and **residual analysis**.

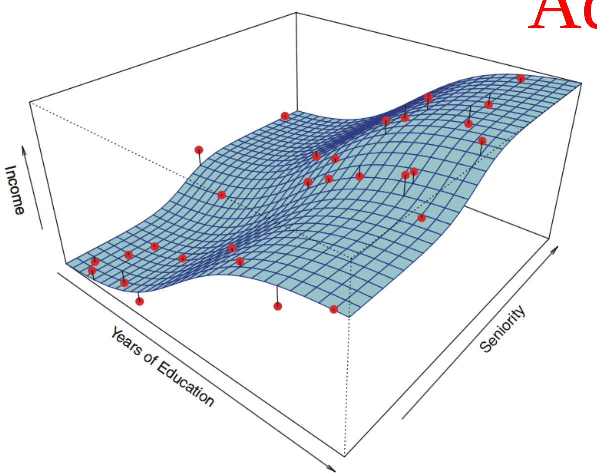
- Usually, simple parametric models imposed on data generated by complex systems, for example, medical data, financial data, result in a loss of accuracy and information as compared to algorithmic models

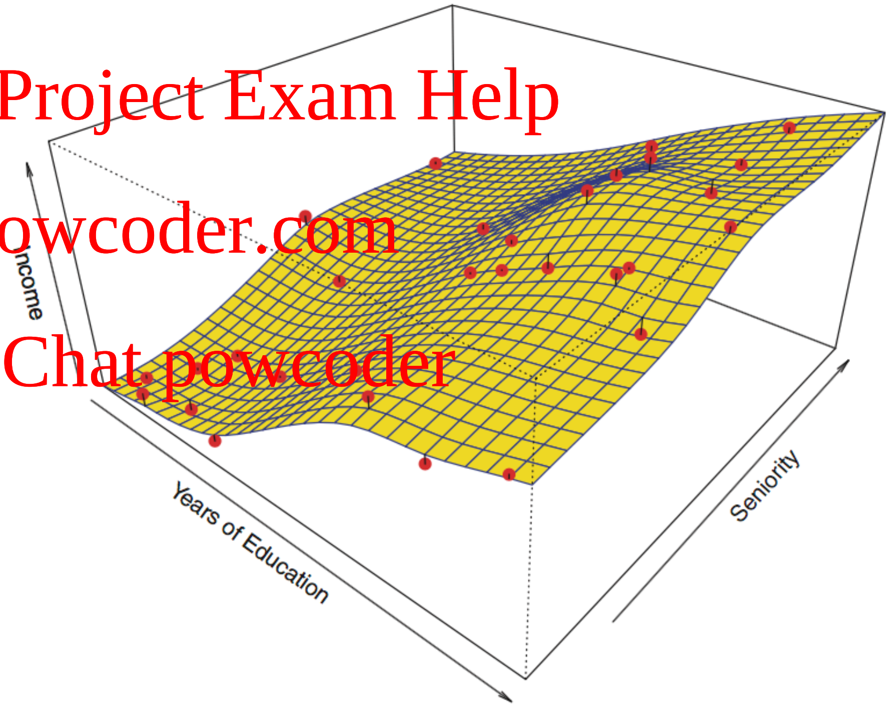
Source: Statistical Science 2001, Vol. 16, No. 3, 199–231 **Statistical Modeling: The Two Cultures Leo Breiman**

REGULARIZATION

IMPOSING ADDITIONAL CONSTRAINTS

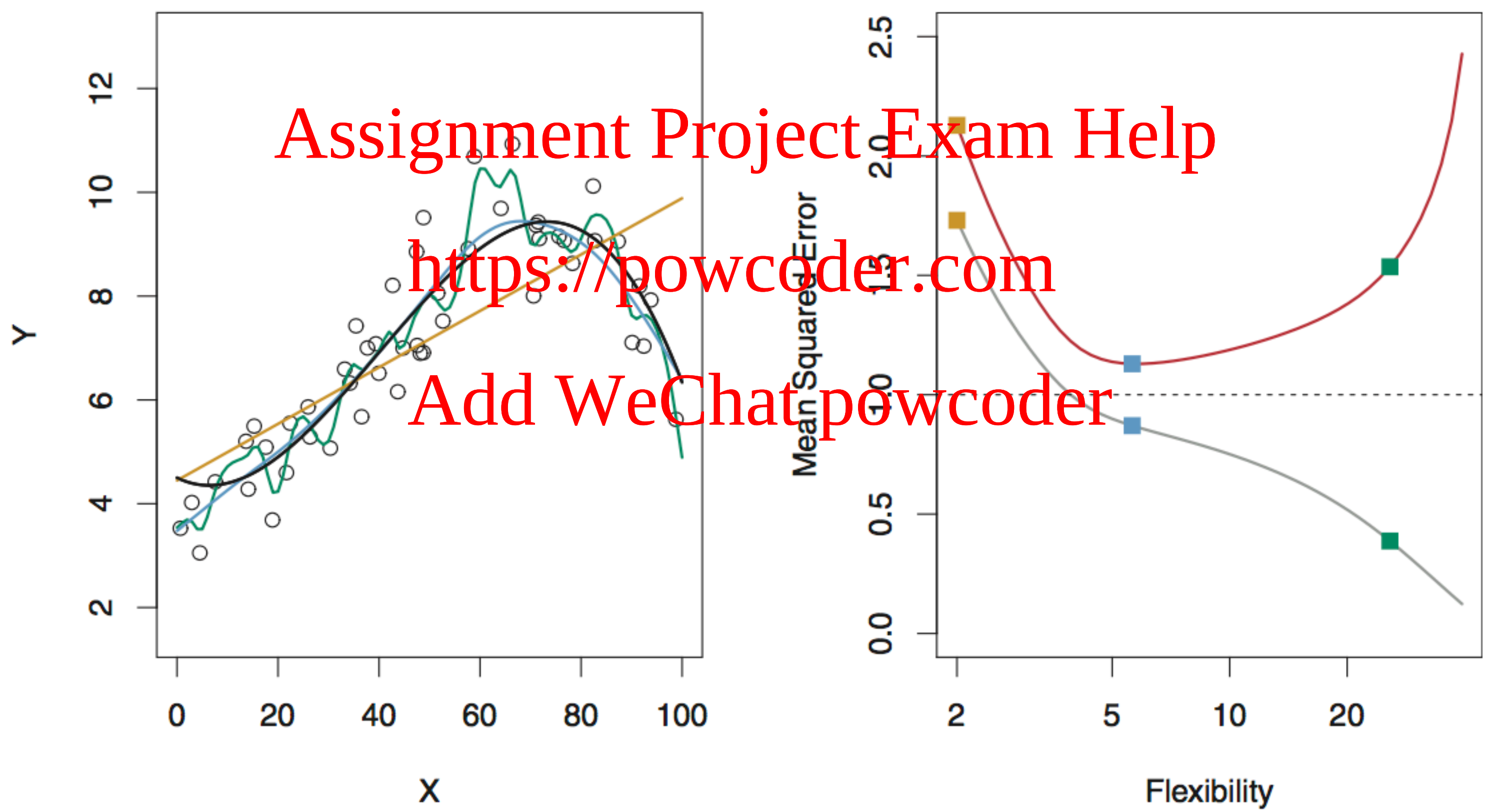




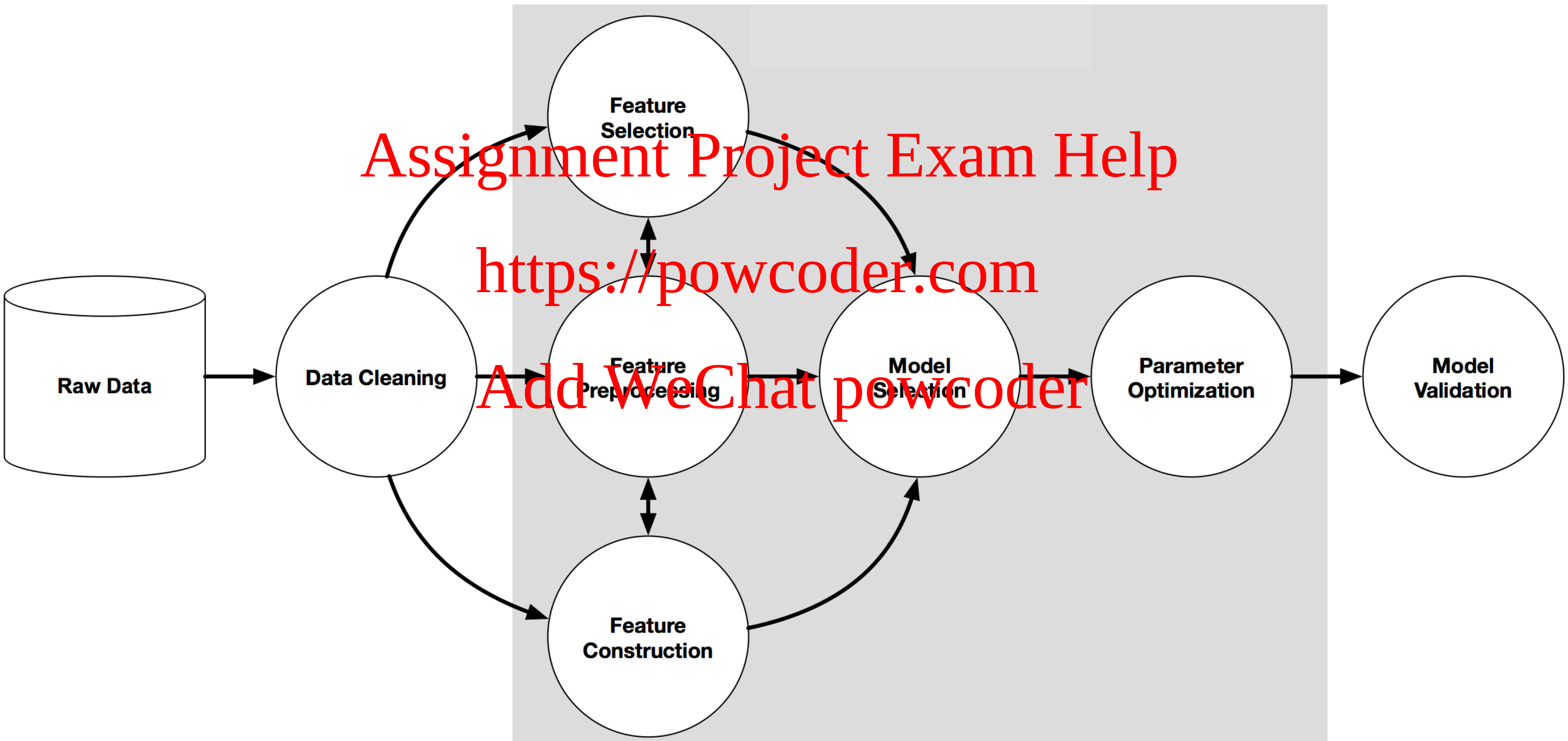


ASSESSING GOODNESS OF FIT

UCL SCHOOL OF MANAGEMENT



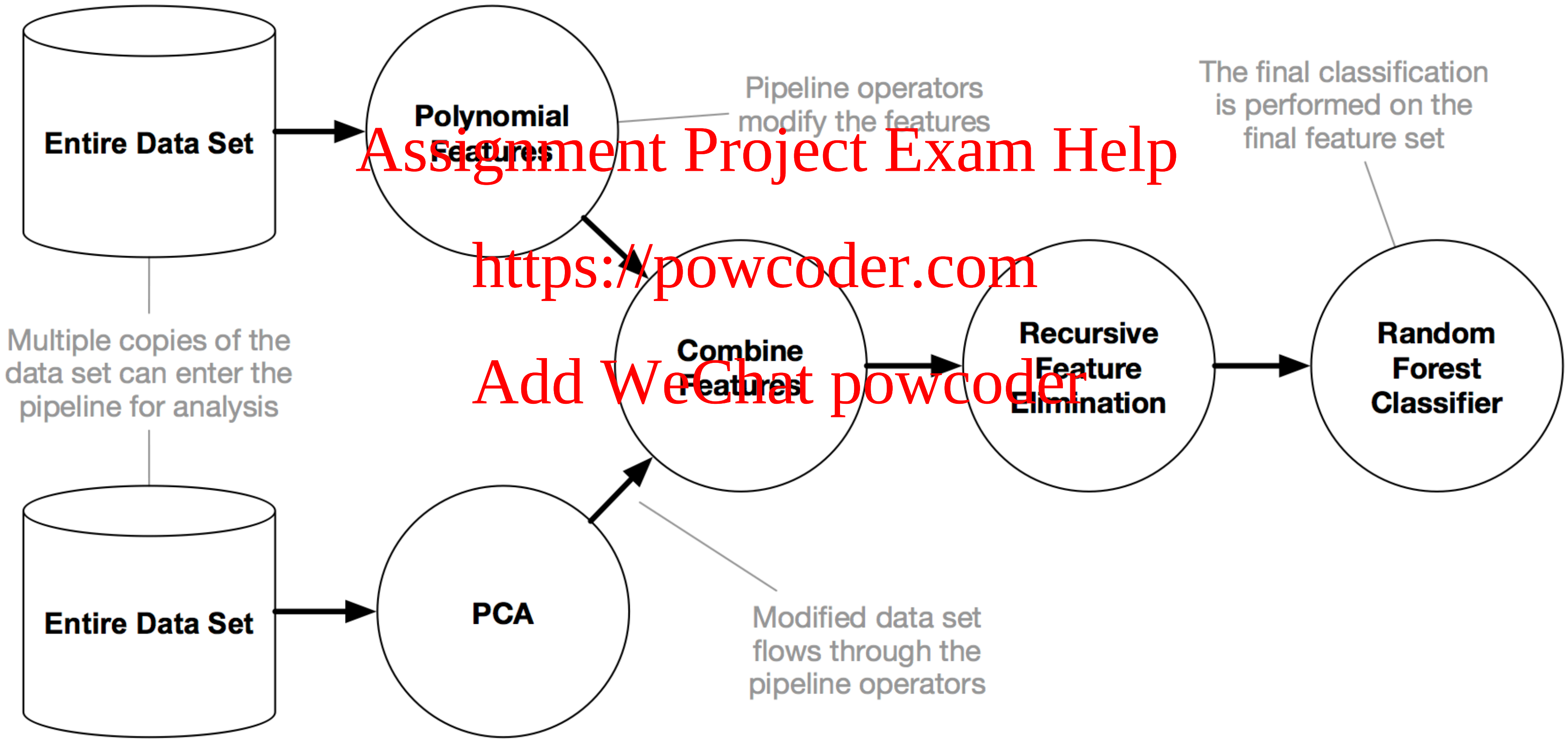
ML PIPELINES



Source: https://epistasislab.github.io/tpot/

ML PIPELINES

FEATURE SELECTION AND AUTOMATION



Source: https://epistasislab.github.io/tpot/

HOMEWORK

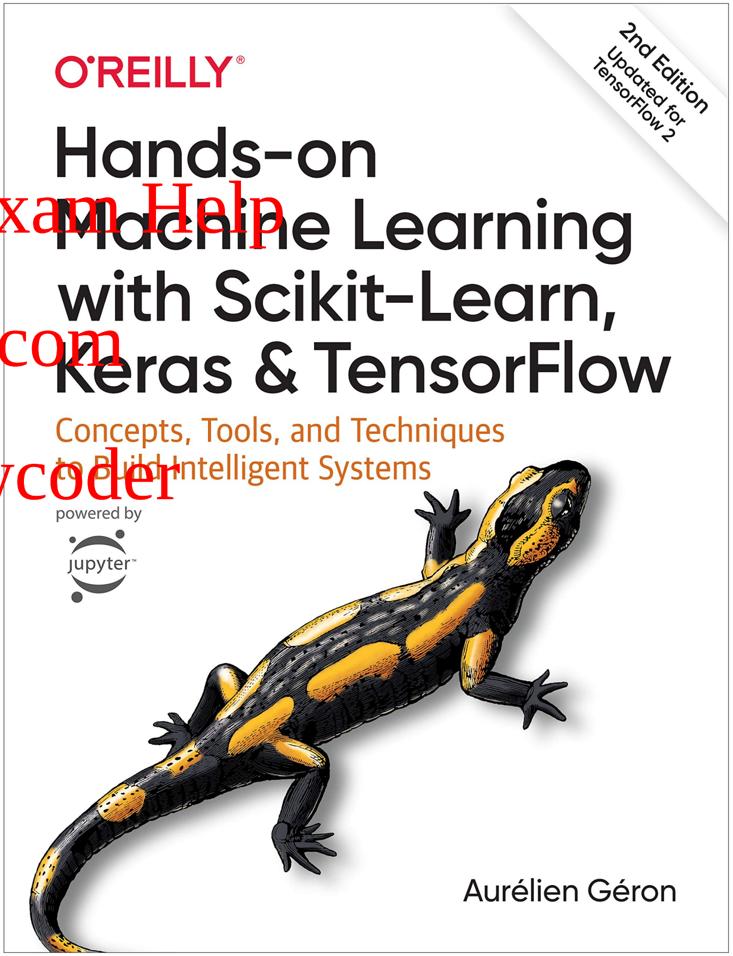
1, 2

Hands-on Machine Learning

Chapter 2: End-to-End Machine Learning Project

Try reading the Chapter from start to finish. We will work through the problem in class but please come prepared to discuss the case study.

It is easier to understand the different stages of a ML project if you follow one from start to finish.



END TO END

UCL
SCHOOL OF
MANAGEMENT

1. **Look at the big picture**
2. **Get some data**
3. **Discover, explore, vizualise to gain insight**
4. **Prepare data for Machine Learning**
5. **Select a model and train it**
6. (*Fine-tune a model*)
7. **Present a solution**
8. (*Launch, monitor and maintain the system*)

EDA
Summary
stat.

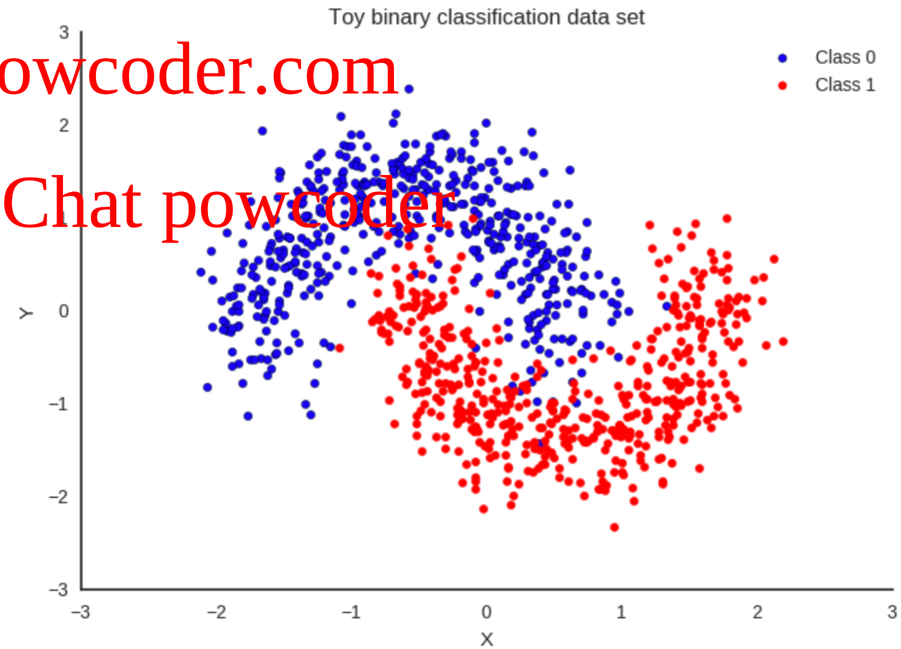
TESTING AND VALIDATION

- Generalization of data
- Generalization of feature representation
- Generalization of the ML model

TOY VS REAL DATA

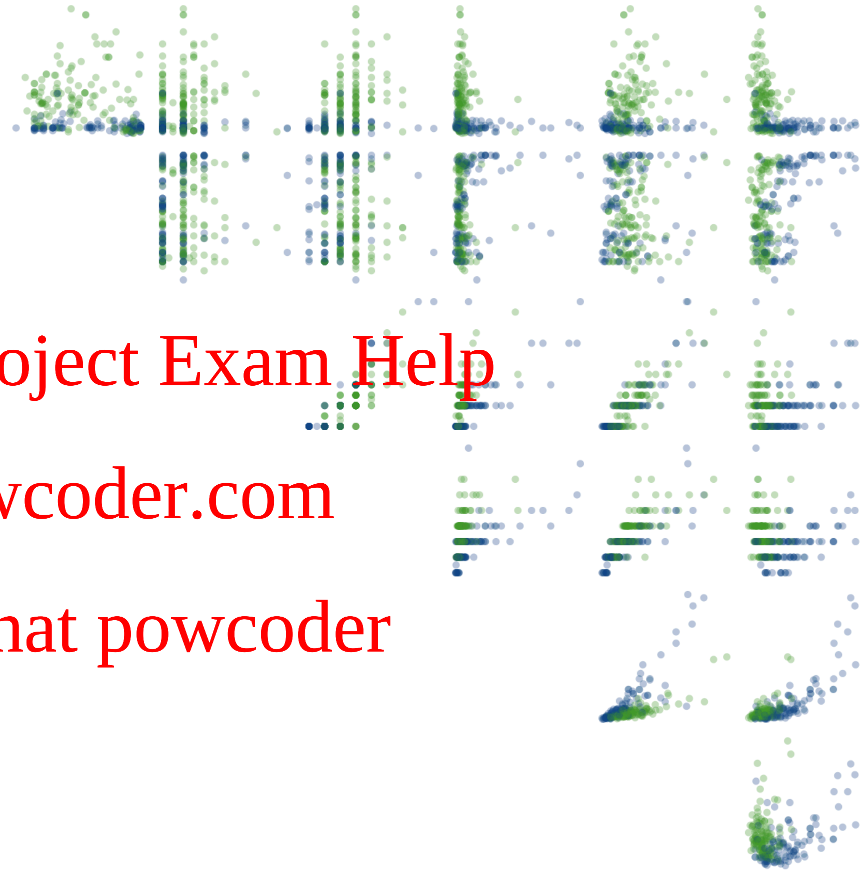
UCL SCHOOL OF MANAGEMENT

- Toy data is useful for exploring behaviour of algorithms
- Demonstrating the advantages and disadvantages of an algorithm
- However, best not to use *just* **Toy** datasets
- Use real datasets



R2
D3

A visual introduction to machine learning



English

In machine learning, computers apply **statistical learning** techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.

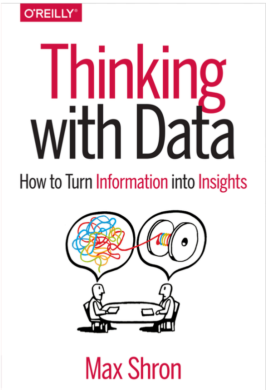
- overfitting
- generalisation

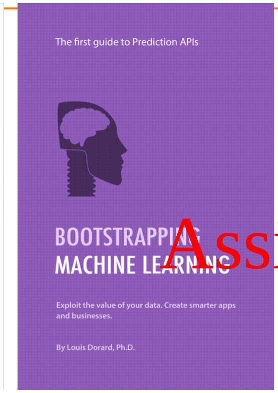
week 3

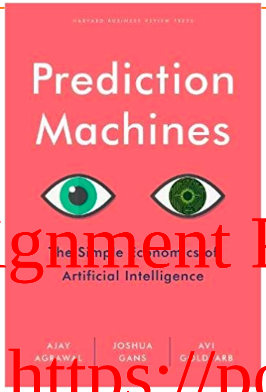
Source: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

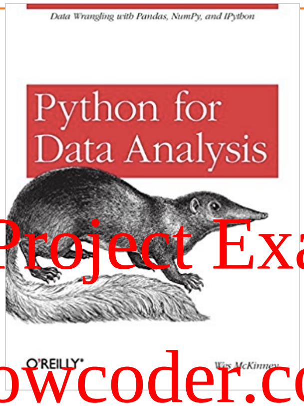
BOOKS

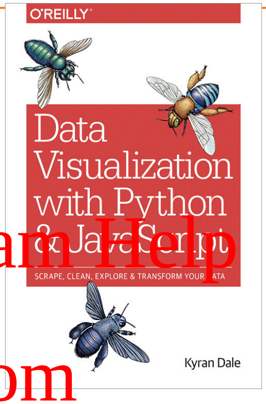
UCL SCHOOL OF MANAGEMENT

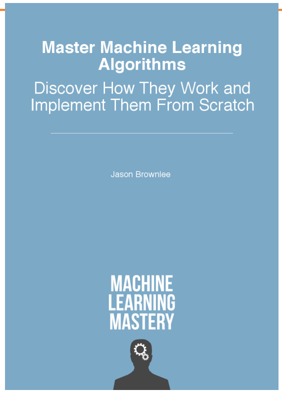


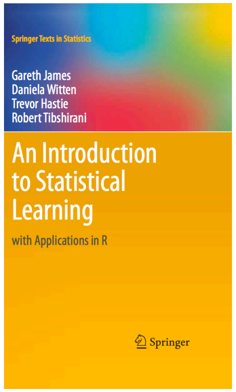


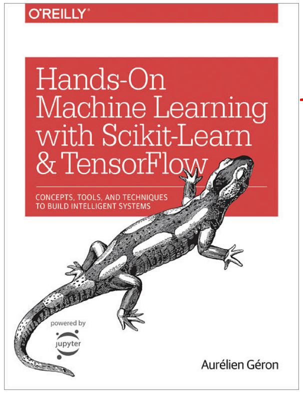


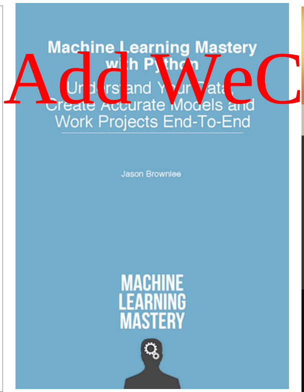


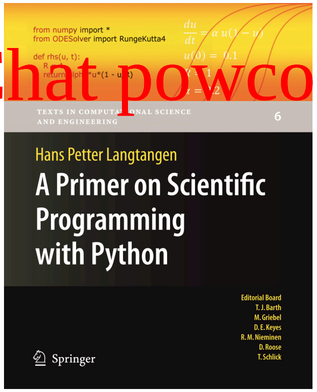


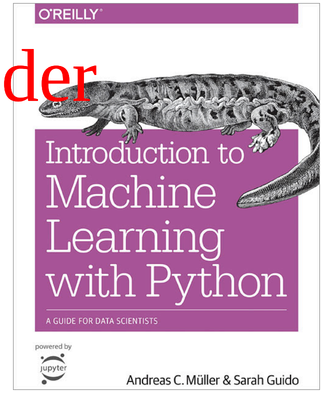












THINKING ABOUT BUSINESS

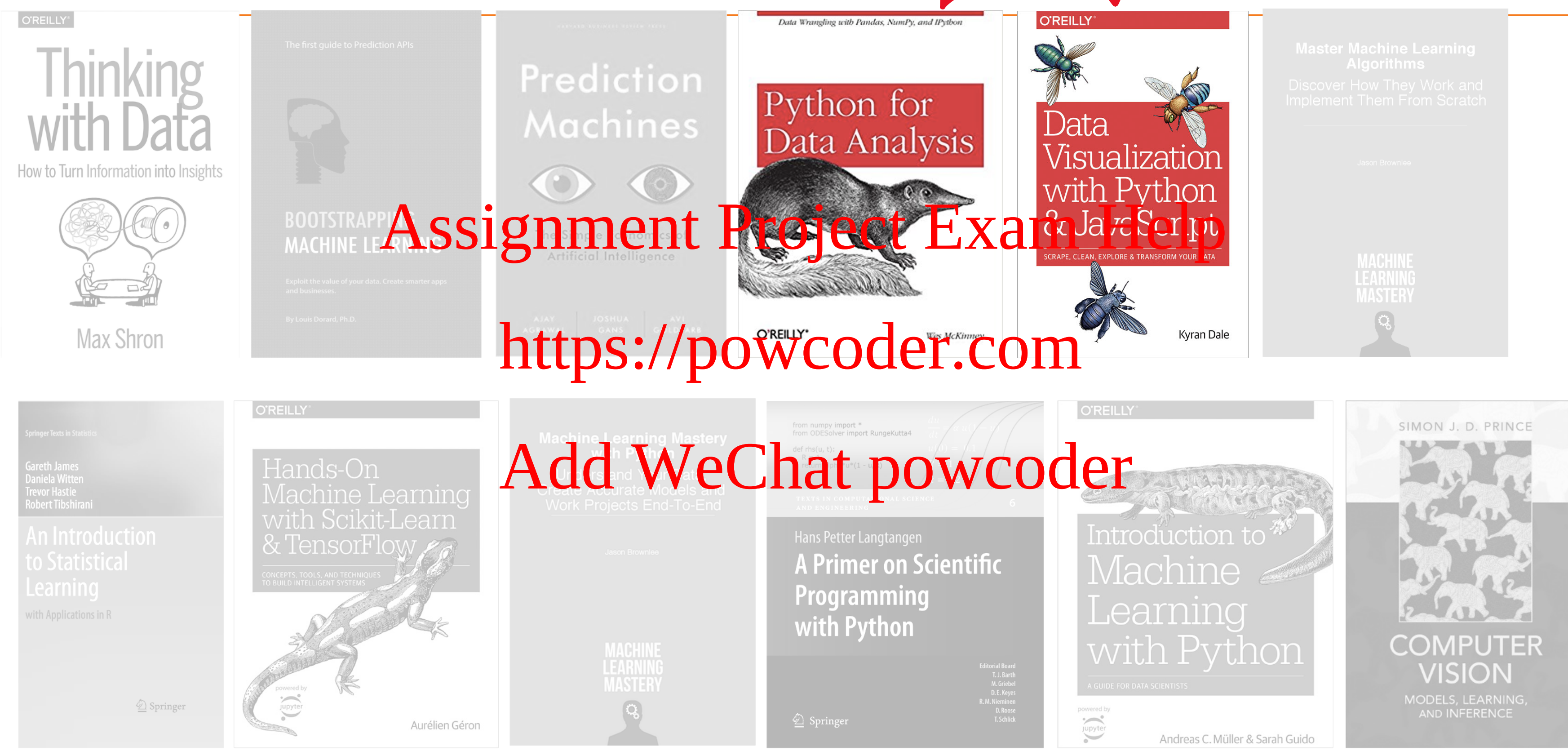
UCL
SCHOOL OF
MANAGEMENT

E-level

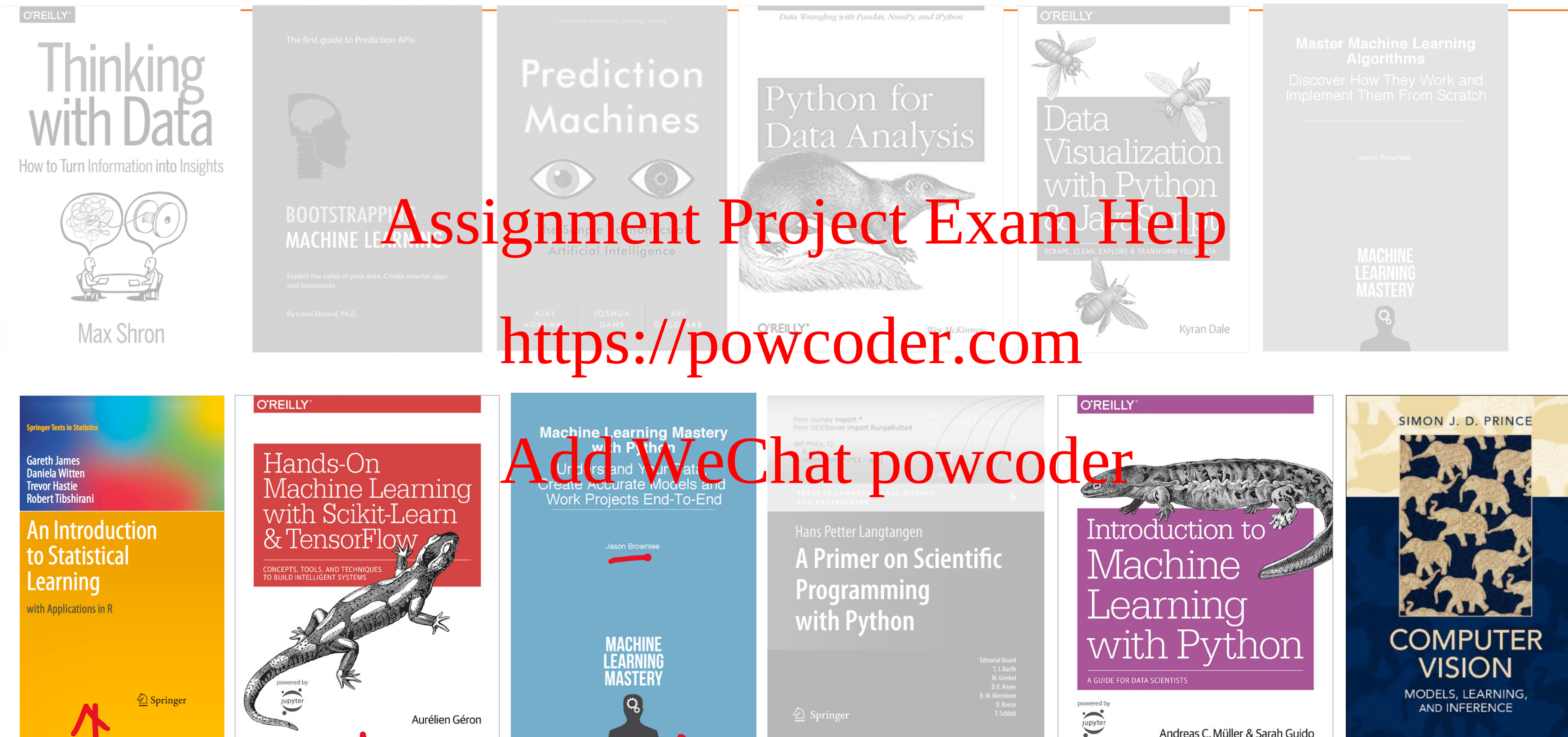
WORKING WITH DATA

UCL
SCHOOL OF
MANAGEMENT

D3

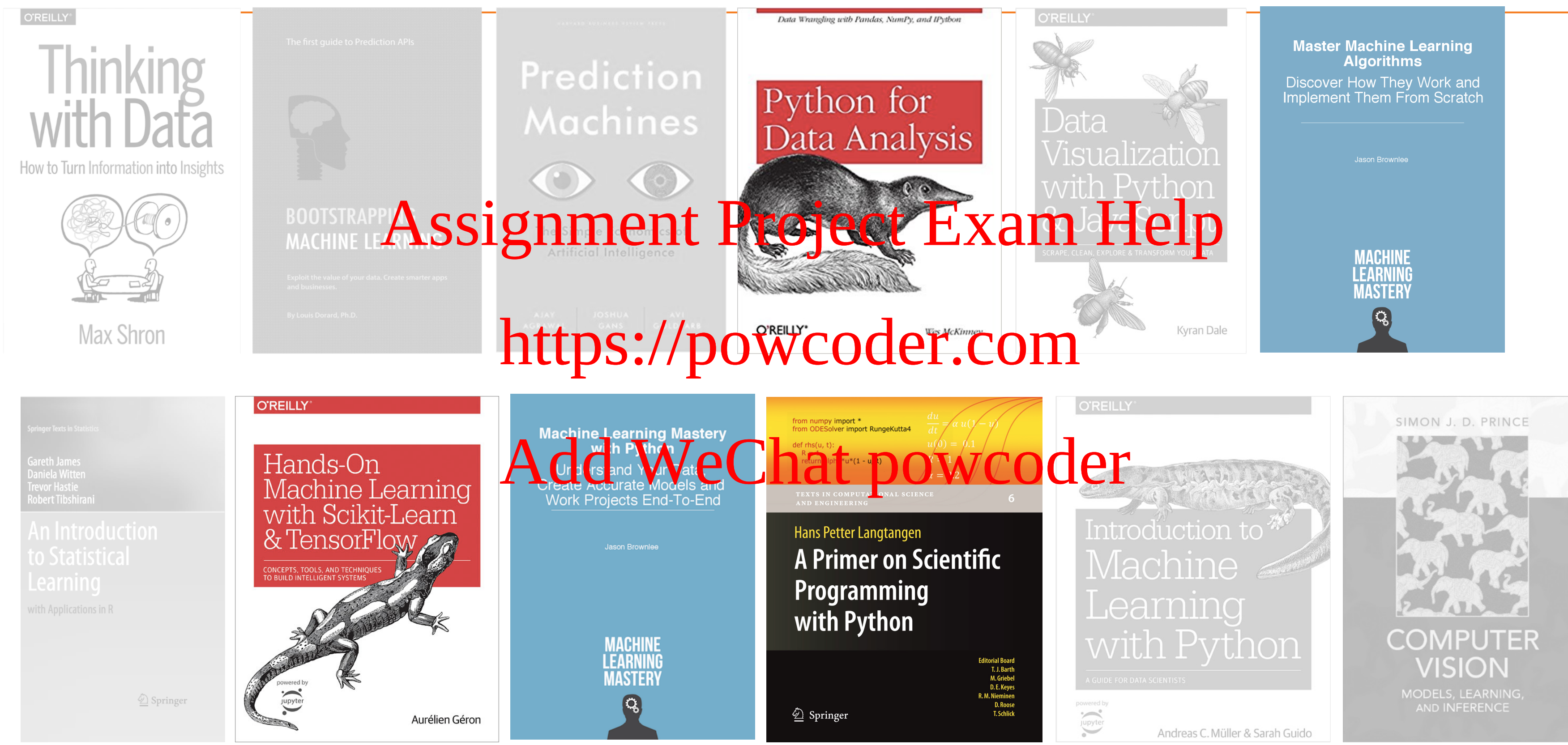


DESIGNING PREDICTIVE MODELS



PYTHON PROGRAMMING

UCL SCHOOL OF MANAGEMENT



A - B - C- D

A TAXONOMY OF PROBLEMS

UCL
SCHOOL OF
MANAGEMENT

A. ClAssification

Week 2 – Classification and Regression

Week 3 – Trees and Ensembles

B. Regression

- Chpt. 2.
- End-2-end

C. Clustering

Week 5 – Clustering

D. Decomposition

Week 4 – Kernel spaces and Decomposition

LECTURE 1 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

PREDICTIVE ANALYTICS

A P MOORE