WEEK 1 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

PREDICTIVE ANALYTICS

A P MOORE

A bit about me

MY BACKGROUND

CURRENT | PREVIOUS

UCL SCHOOL OF MANAGEMENT

Alastair Moore

Head of Analytics and Machine Learning

Mishcon de Reya

Senior Teaching Fellow
Predictive Analytics
MSc Business Analytics
UCL School of Management

UCL SCHOOL OF MANAGEMENT

MBA Programme,
Innovation and Entrepreneurship in Europe
Emerging Business Technologies
BiMBA Peking University

BIMBA 北京大学国家发展研究院

Co-Founder
Satalia.com

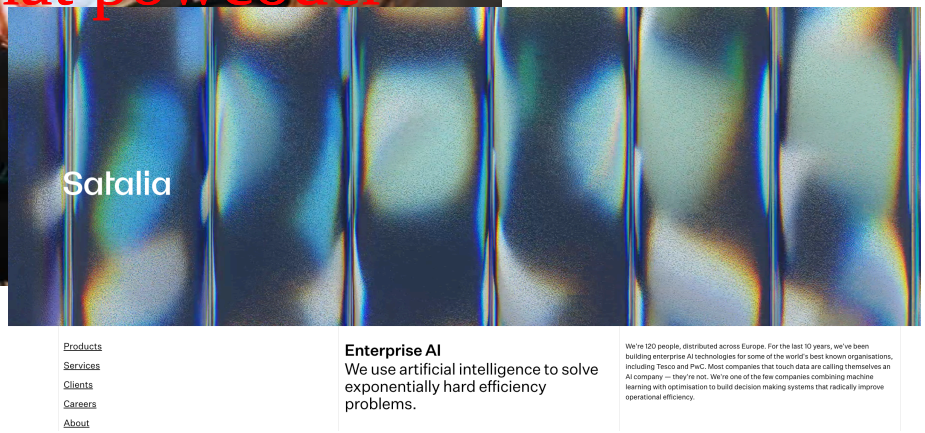
Co-Founder
WeArePopUp.com

WE ARE POP UP

Partner
AMAAMS LLP

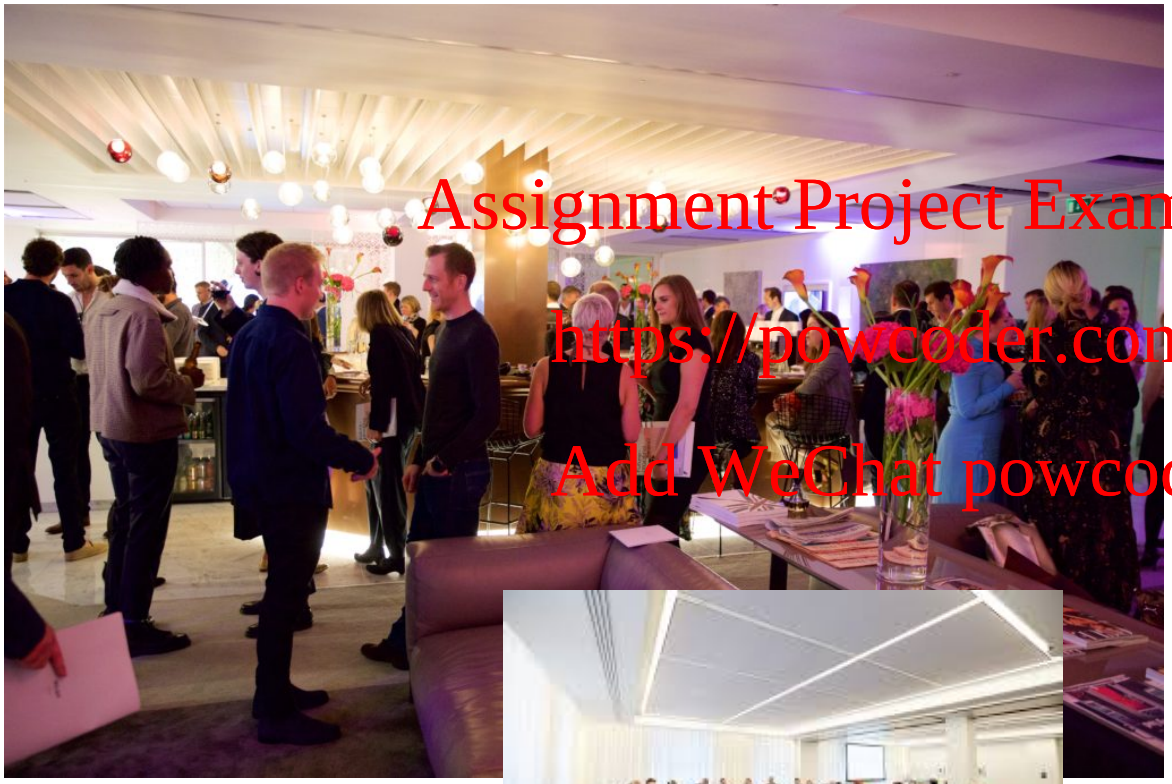
SATALIA.COM

UCL
SCHOOL OF
MANAGEMENT



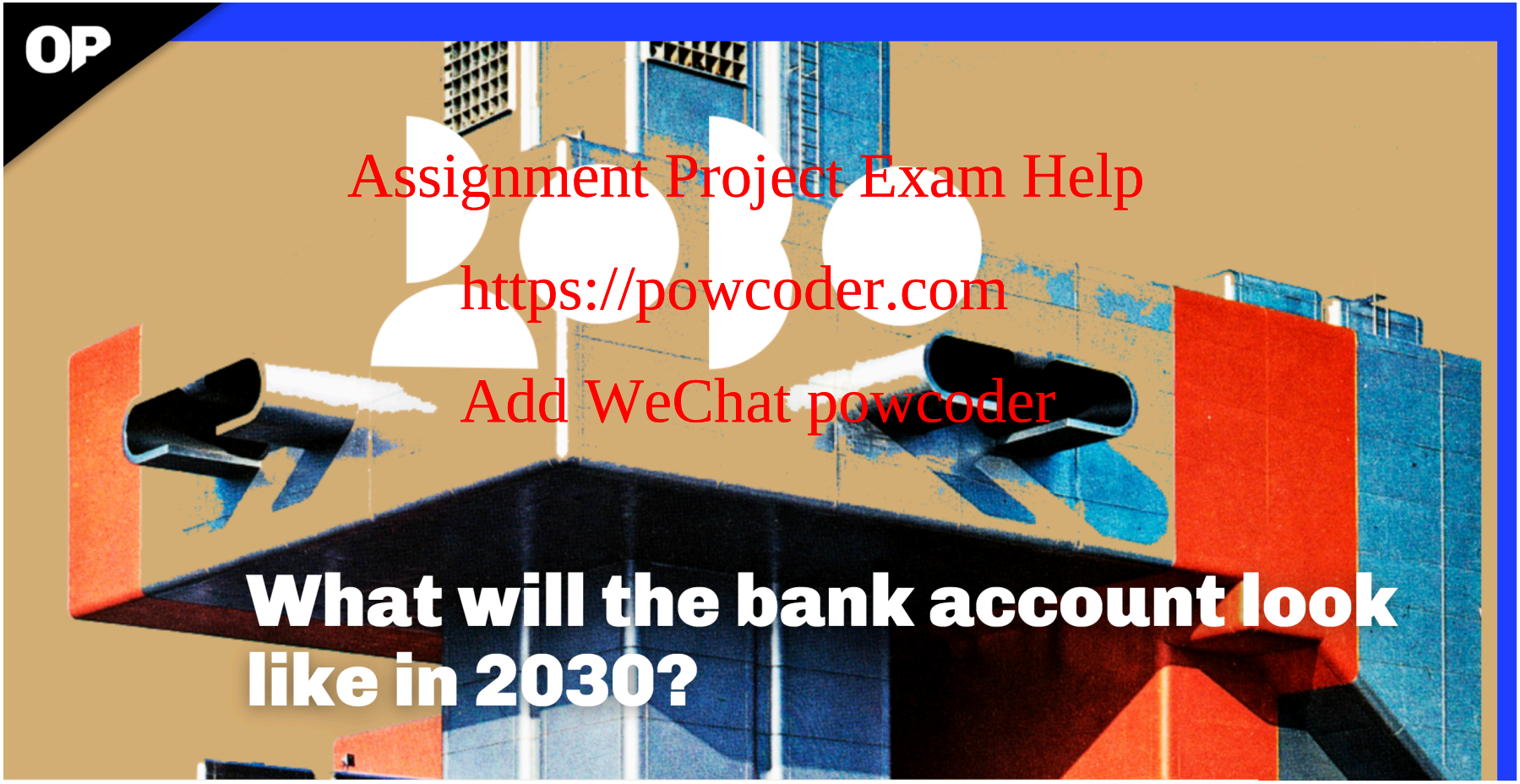
MISHCON DE REYA

UCL
SCHOOL OF
MANAGEMENT



OPEN BANKING

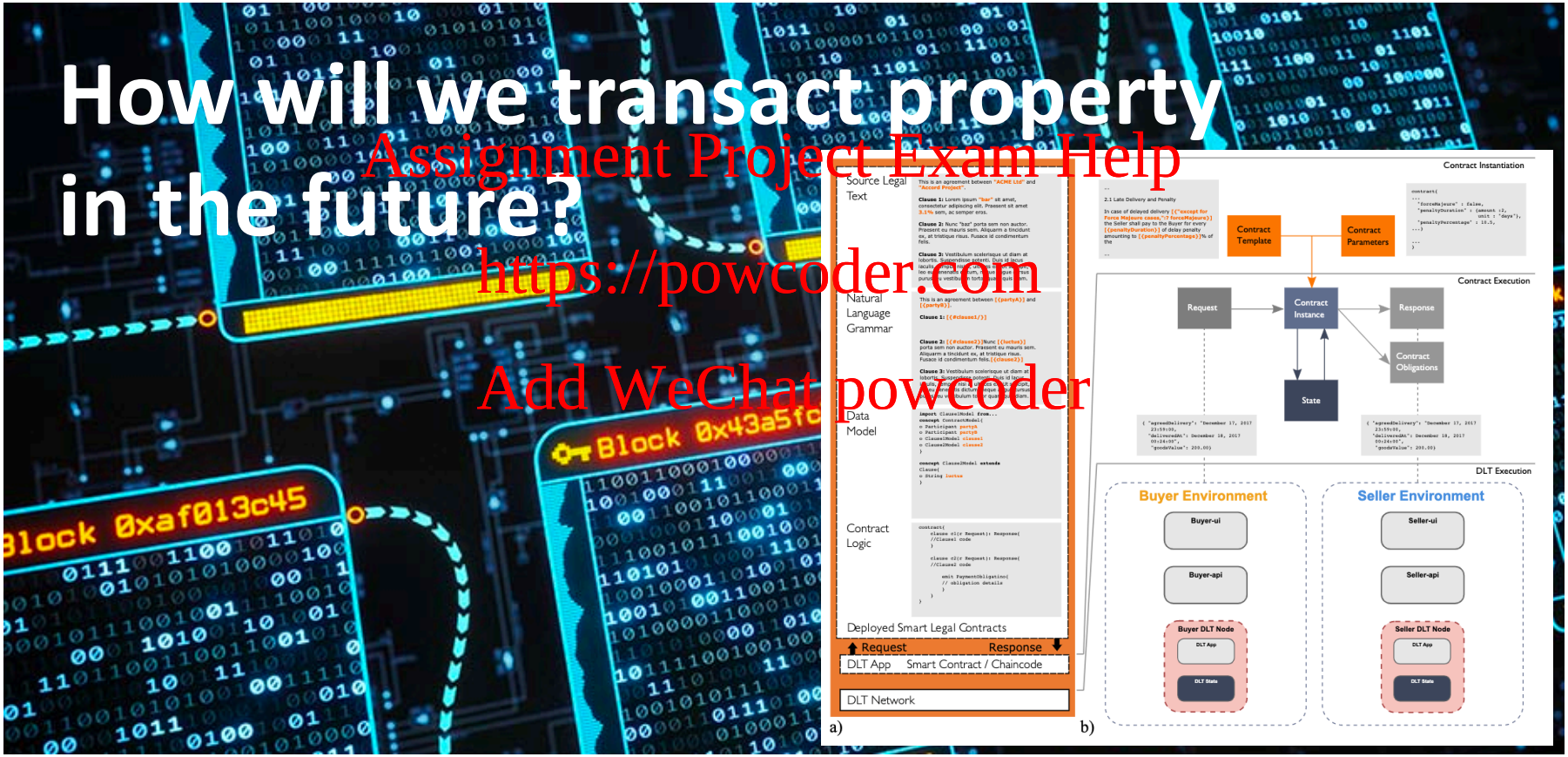
WORKING WITH NESTA ON CHALLENGE LEAD APPROACHES



Source: www.openupchallenge.io

DLT SYSTEMS

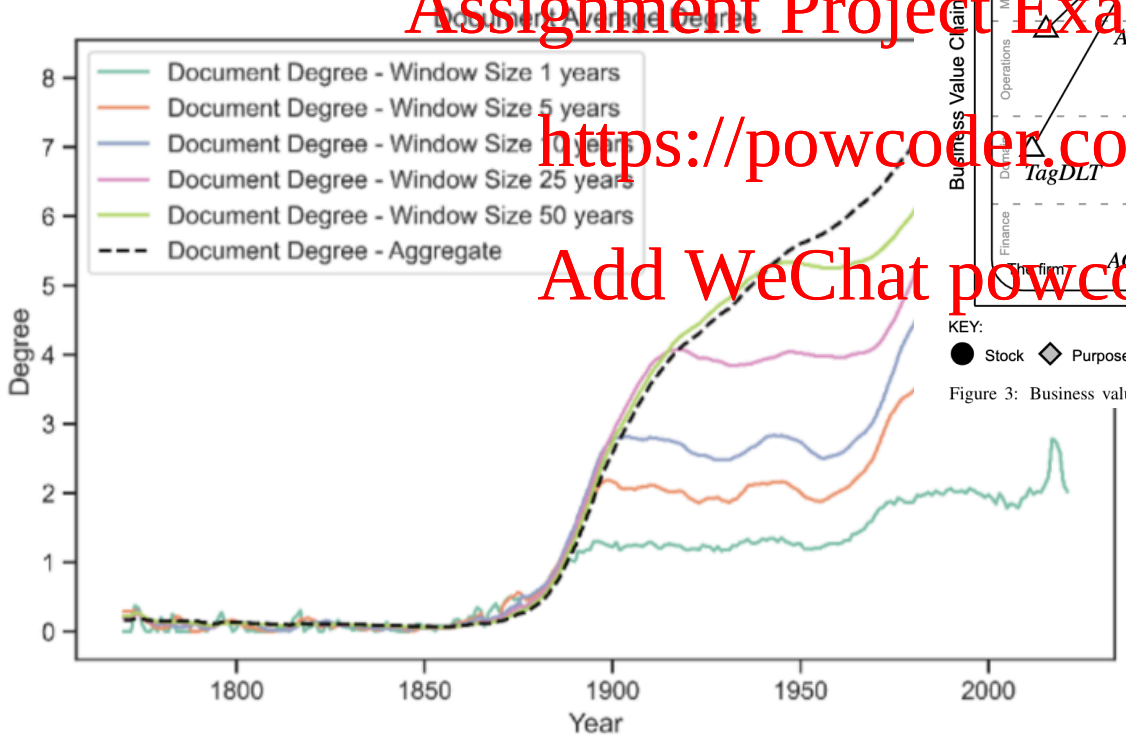
USE OF DISTRIBUTED LEDGERS IN REAL ESTATE



Source: www.gov.uk/government/news/hm-land-registry-to-explore-the-benefits-of-blockchain

RESEARCH INTERESTS

UCL SCHOOL OF MANAGEMENT



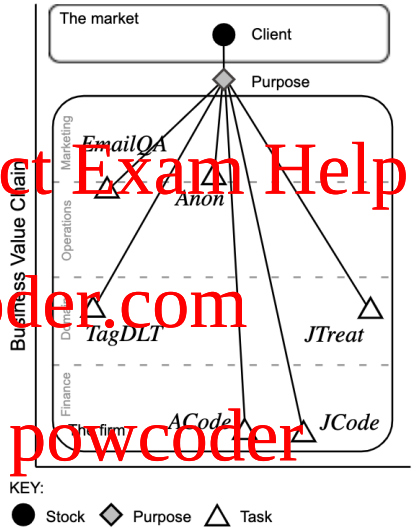


Figure 3: Business value chains. Distribution of se-

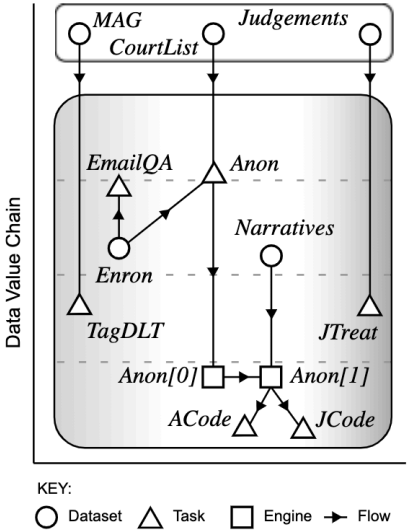


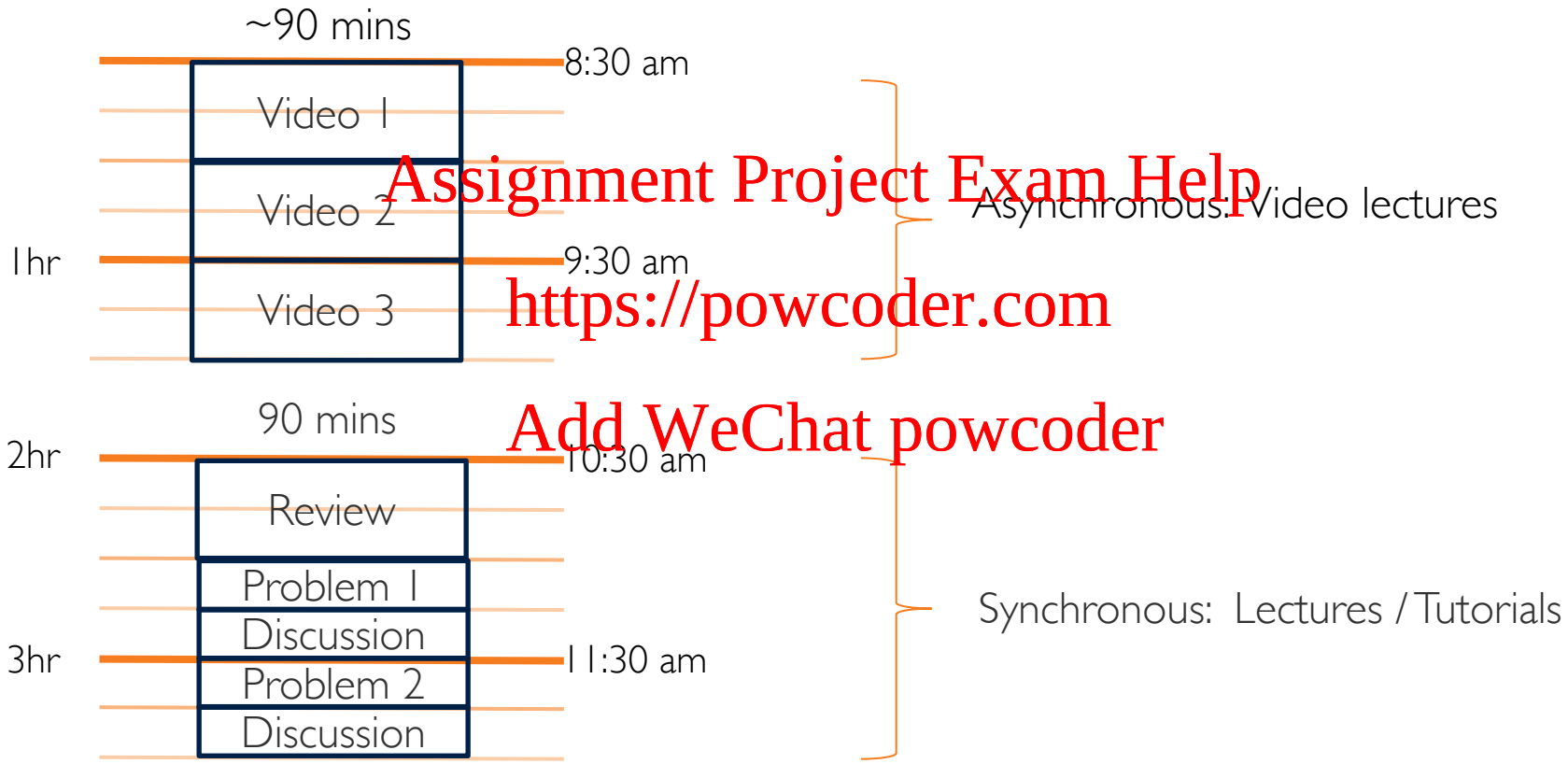
Figure 4: Data value chains. Showing the dependencies

Course overview

TODAY

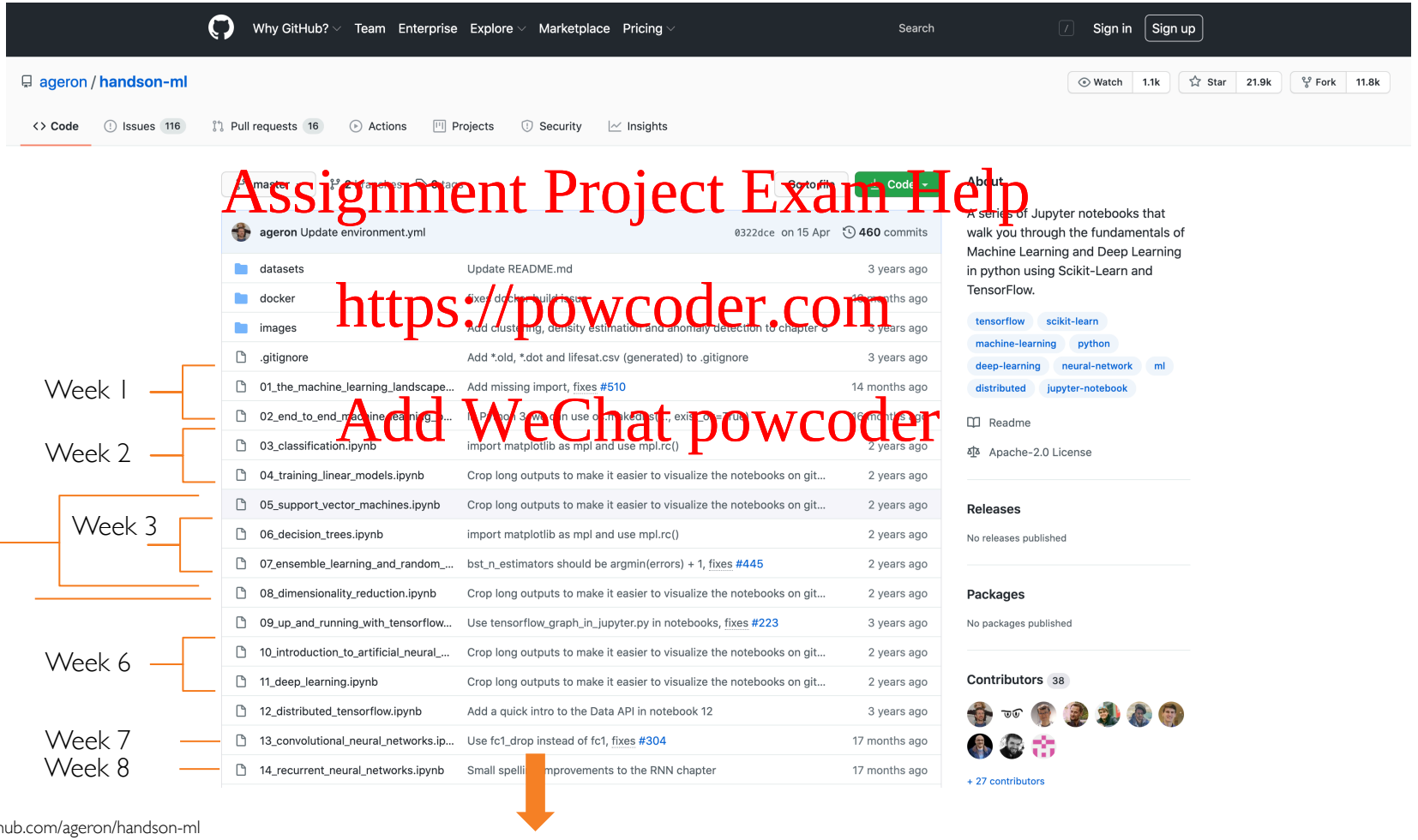
LECTURE / WORKSHOP SCHEDULE

UCL
SCHOOL OF
MANAGEMENT



HANDS ON MACHINE LEARNING GIT

UCL SCHOOL OF MANAGEMENT



Source: https://github.com/ageron/handson-ml

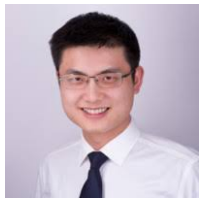
TEACHING SUPPORT

UCL SCHOOL OF MANAGEMENT

Kamil Tylinski

Teaching Assistant

kamil.tylinski.16@ucl.ac.uk



Jiangbo Shangguan

Teaching Assistant

j.shangguan.17@ucl.ac.uk

Bartos Kultys

Teaching Assistant

bartosz.kultys.18@ucl.ac.uk

Dr Viviana Culmone

Teaching Assistant

v.culmone@ucl.ac.uk

Editha Nemsic

Teaching Assistant

editha.nemsic.19@ucl.ac.uk

Walter Hernandez

Teaching Assistant

walter.hernandez.18@ucl.ac.uk

ASSESSMENTS

UCL SCHOOL OF MANAGEMENT

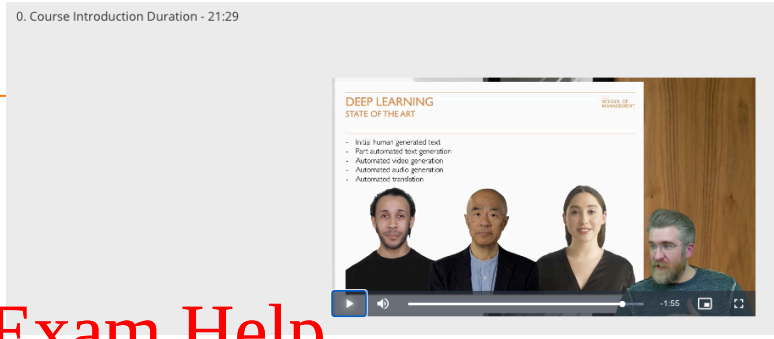
- Individual Coursework - 60%
  - 2000 words
  - Due Date: Friday 26th February 2021
- Group Coursework - 40%
  - 4-5 per group
  - Deadline for Group formation: **Friday 29th January 2021 (week 3)**
  - 2000 words
  - Due Date: Thursday 18th March 2021

Review

DEEP LEARNING

STATE OF THE ART

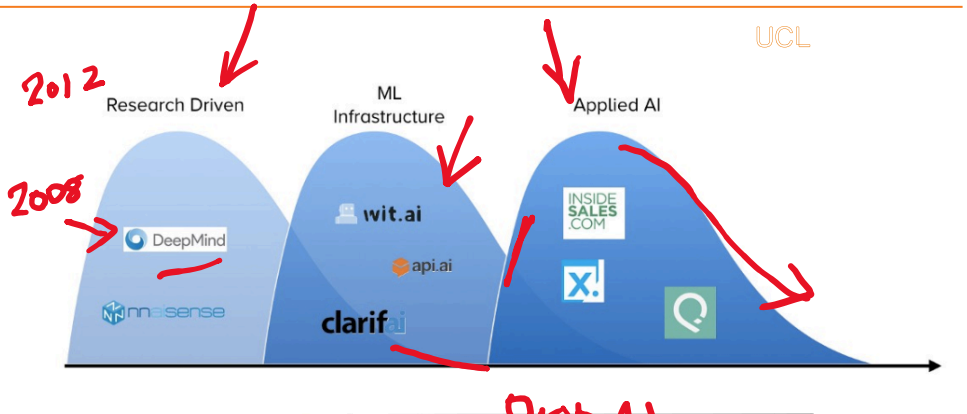
UCL SCHOOL OF MANAGEMENT

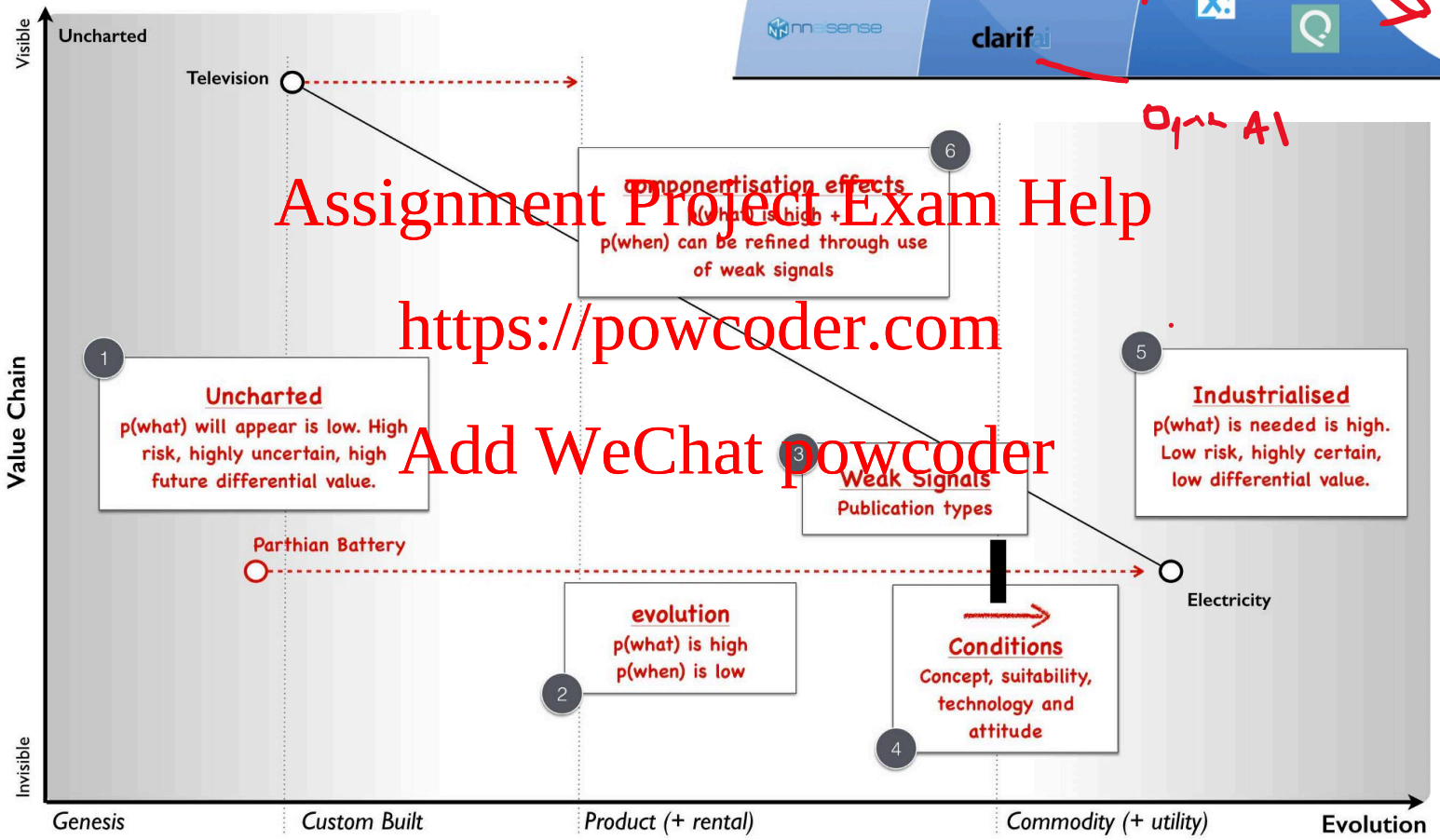


- Initial human generated text
- Part automated text generation
- Automated video generation
- Automated audio generation
- Automated translation

WAVES OF INVESTMENT

THE ROLE OF INDUSTRIALIZATION

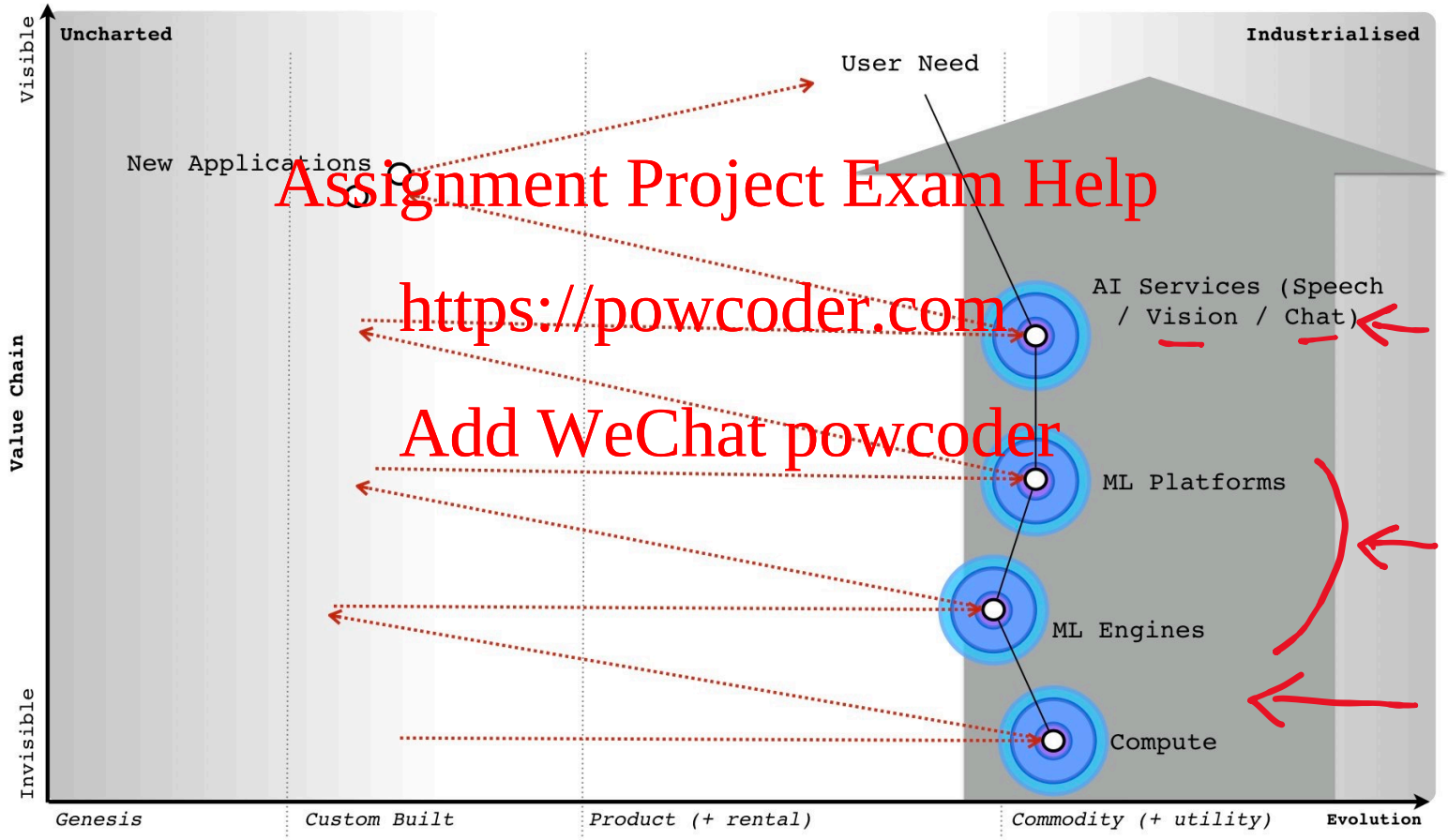




https://machinelearnings.co/winning-strategies-for-applied-ai-companies-f02cac0a6ad8

TOWER & MOAT

UCL SCHOOL OF MANAGEMENT



Source: https://blog.gardeviance.org/2014/07/tower-and-moat.html

MACHINE LEARNING JARGON

UCL SCHOOL OF MANAGEMENT

- Model
- Interpolating / Extrapolating
- Data Bias
- Noise / Outliers
- Learning algorithm
- Inference algorithm
- **Supervised learning**
- **Unsupervised learning**
- Classification
- Regression
- Clustering
- Decomposition

- Parameters
- Optimisation
- Training data
- Testing data
- Error metric
- Linear model
- Parametric model
- Model variance
- Model bias
- Model generalization
- Overfitting
- Goodness-of-fit
- Hyper-parameters

- Failure modes
- Confusion matrix
- True Positive
- False Negative
- Data density
- Partition
- Hidden parameter
- High dimensional space
- Low dimensional space
- Separable data
- Manifold / Decision surface
- Hyper cube / volume / plane

机器学习行话

UCL SCHOOL OF MANAGEMENT

- 模型
- 内插 / 外推
- 数据偏差
- 噪声 / 离群值
- 学习算法
- 推断算法
- 监督学习
- 无监督学习
- 分类
- 回归
- 聚类
- 分解
- 参数
- 优化
- 训练数据
- 测试数据
- 误差指标
- 线性模型
- 参数模型
- 模型方差
- 模型偏差
- 模型泛化
- 过拟合
- 拟合优度
- 超参数
- 失败模式
- 混淆矩阵
- 真正例
- 假反例
- 数据密度
- 划分
- 隐藏参数
- 高维空间
- 低维空间
- 可分数据
- 流形/ 决策面
- 超立方体 / 超体积 / 超平面

MODEL / 模型

MÓXÍNG

UCL SCHOOL OF MANAGEMENT

Data + model → prediction

MACHINE LEARNING

DATA DRIVEN AI

UCL SCHOOL OF MANAGEMENT

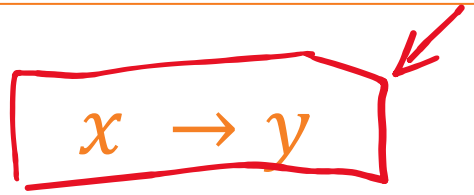
Assume there is enough data to find statistical associations to solve specific tasks

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



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

LEARNING A FUNCTION

UCL SCHOOL OF MANAGEMENT

$$x \rightarrow y$$

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

Measured data

Features

Inferred/Predicted/Estimated value

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

True initial value
(world state)

Learned/Fitted function
From n observations

True target value
(world state)

LEARNING A FUNCTION

UCL SCHOOL OF MANAGEMENT

$$x \rightarrow y$$

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

Measured data

Features

Inferred/Predicted/Estimated value

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

True initial value
(world state)

Learned/Fitted function
From n observations

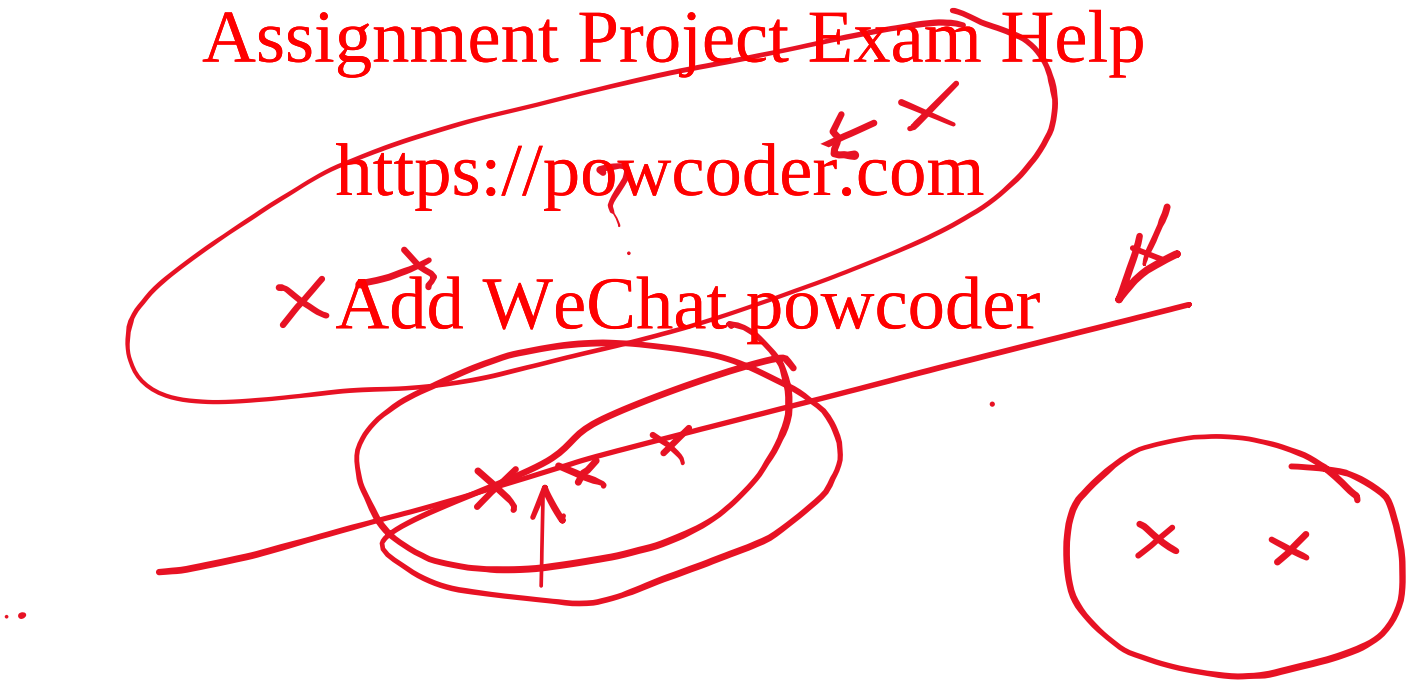
True target value
(world state)

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

INTERPOLATING / 内插

NÈI CHĀ

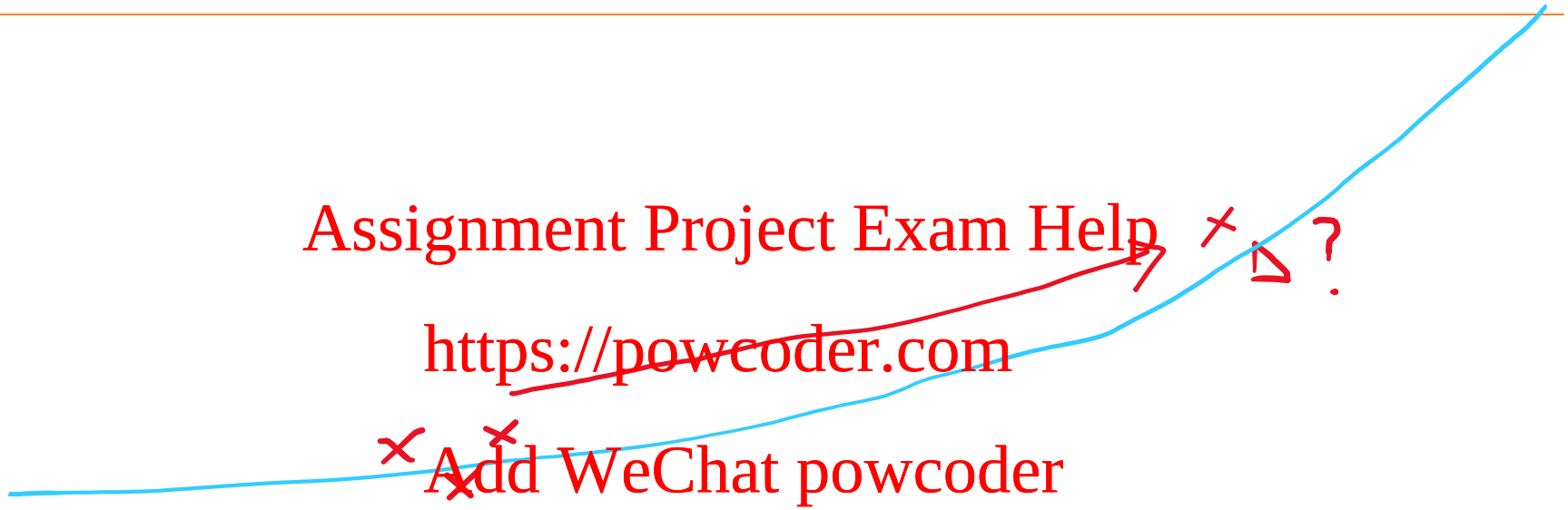
UCL SCHOOL OF MANAGEMENT



EXTRAPOLATING / 外推

WÀI TUĪ

UCL SCHOOL OF MANAGEMENT



NOISE, OUTLIERS / 噪声，离群值

ZÀOSHĒNG , LÍ QÚN ZHÍ

UCL SCHOOL OF MANAGEMENT

LEARNING ALGORITHM / 学习算法

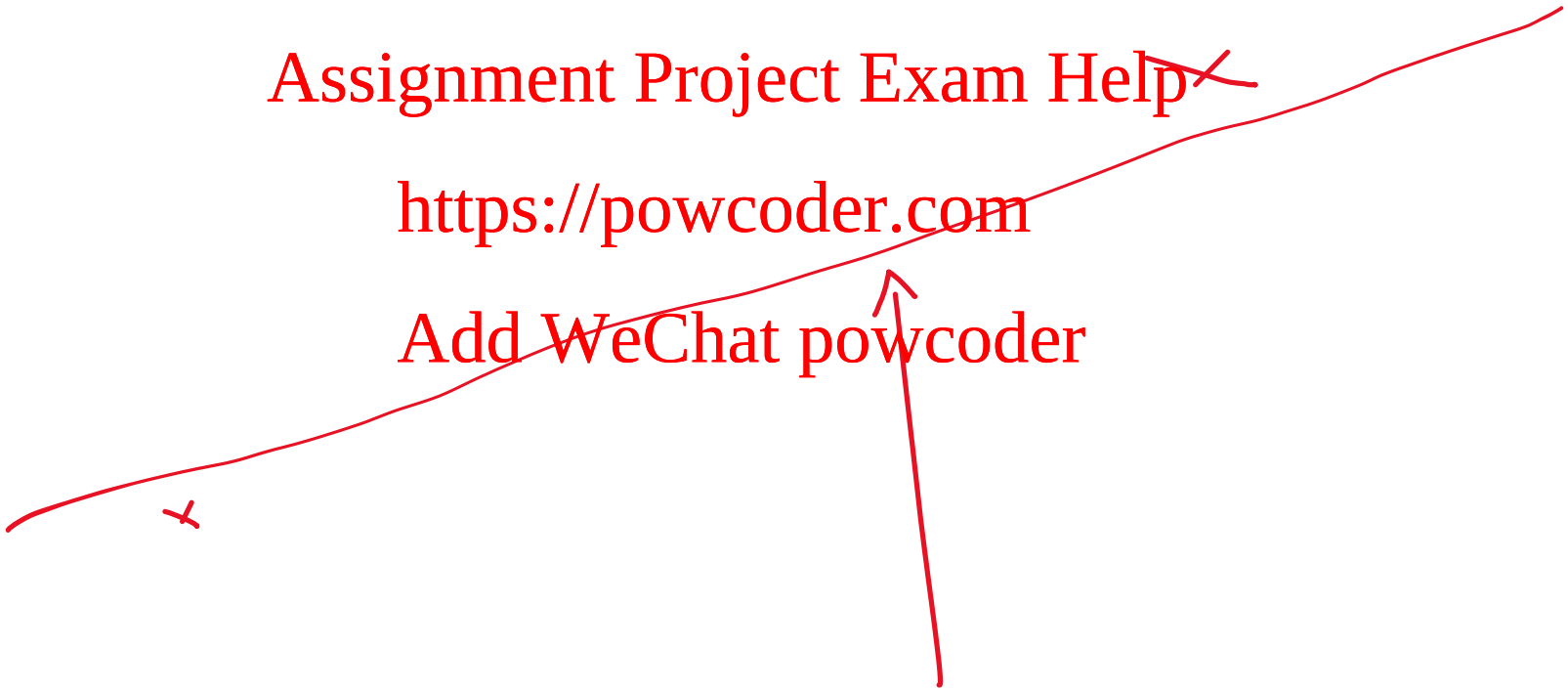
XUÉXÍ SUÀNFǍ

UCL SCHOOL OF MANAGEMENT

INFERENCE ALGORITHM / 推断算法

TUĪDUÀN SUÀNFǍ

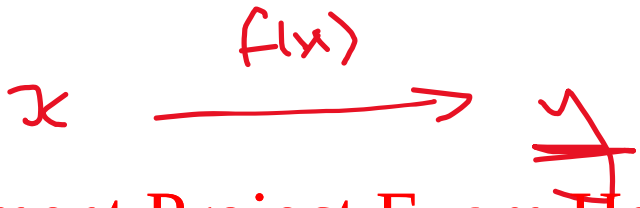
UCL SCHOOL OF MANAGEMENT



SUPERVISED LEARNING / 监督学习

JIĀNDŪ XUÉXÍ

UCL SCHOOL OF MANAGEMENT



train
test

UNSUPERVISED LEARNING / 无监督学习

WÚ JIĀNDŪ XUÉXÍ

UCL SCHOOL OF MANAGEMENT

goodness

- similar
- closeness

CLASSIFICATION / 分类

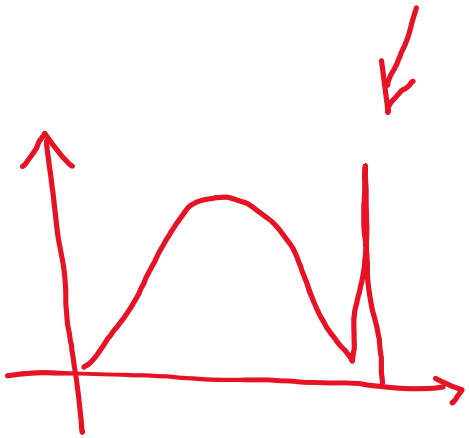
FĒNLÈI

UCL
SCHOOL OF
MANAGEMENT

CLUSTERING / 聚类

JÙ LÈI

UCL SCHOOL OF MANAGEMENT

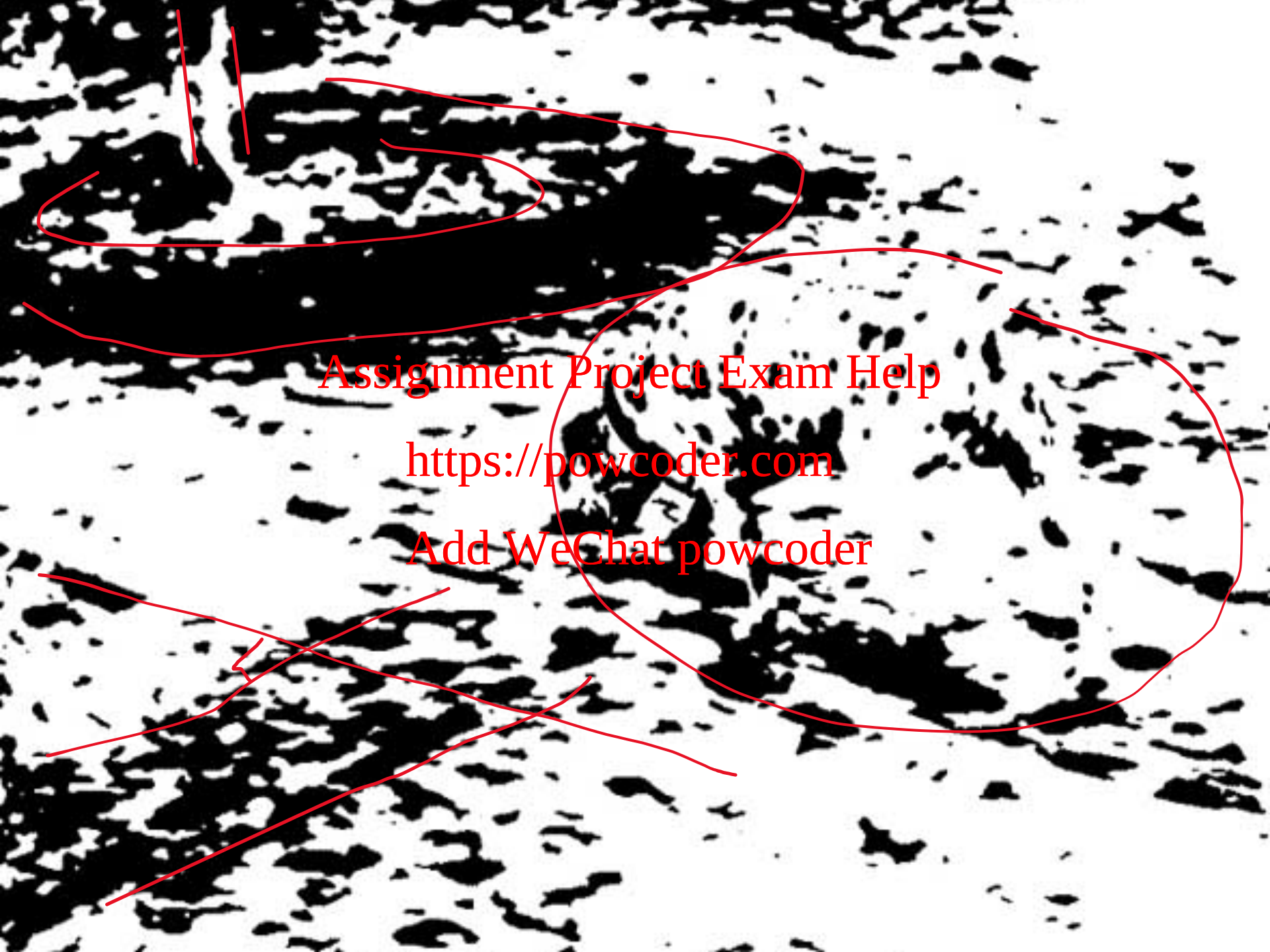


DECOMPOSITION / 分解

FĒNJIĚ

UCL
SCHOOL OF
MANAGEMENT

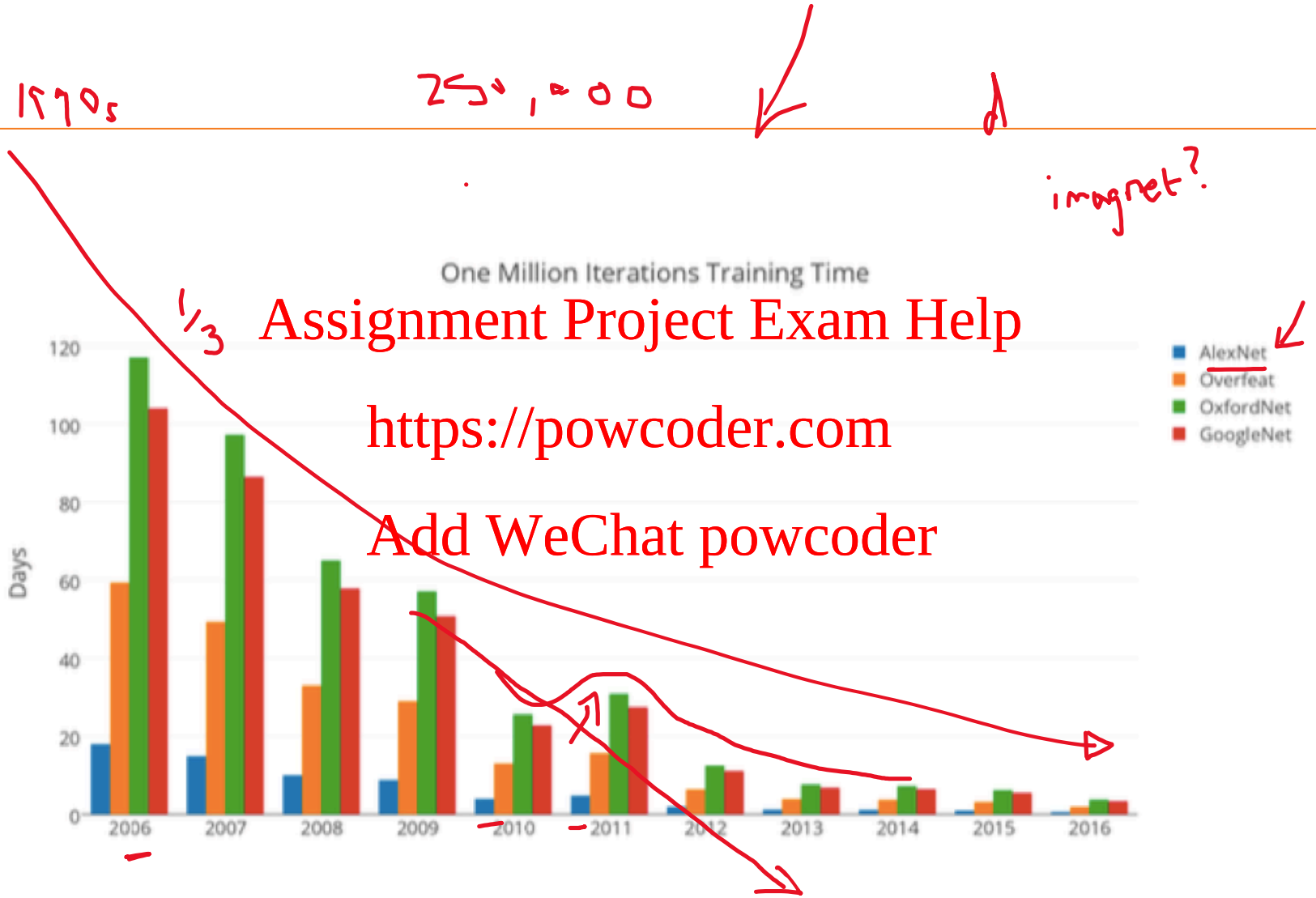
Perception, Patterns and Gestalt


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

The increase in computing power

TRAINING TIMES

UCL SCHOOL OF MANAGEMENT



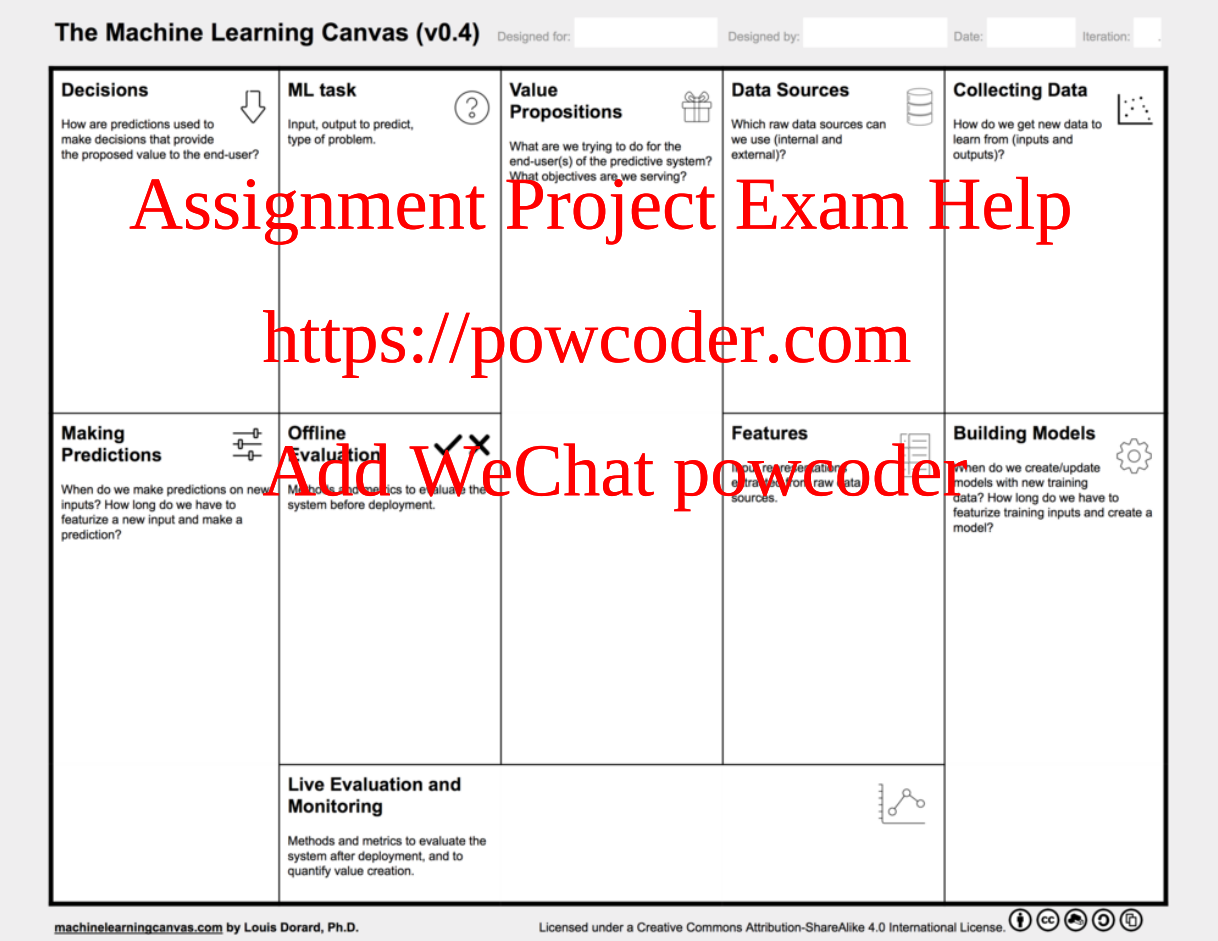
Problem 1

~15 mins group work

~15 mins discussion

PRACTICAL TOOLS
ML CANVAS

UCL SCHOOL OF MANAGEMENT



The Machine Learning Canvas (v0.4) Designed for: Designed by: Date: Iteration:

Decisions
How are predictions used to make decisions that provide the proposed value to the end-user?

ML task
Input, output to predict, type of problem.

Value Propositions
What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources
Which raw data sources can we use (internal and external)?

Collecting Data
How do we get new data to learn from (inputs and outputs)?

Making Predictions
When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation
Methods and metrics to evaluate the system before deployment.

Features
Input representation, extraction from raw data sources.

Building Models
When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring
Methods and metrics to evaluate the system after deployment, and to quantify value creation.

machinelearningcanvas.com by Louis Dorard, Ph.D.

Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

The Machine Learning Canvas (v0.4)

Designed for: | Designed by: | Date: | Iteration:

Learn

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

ML task

Input, output to predict, type of problem.

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources

Which raw data sources can we use (internal and external)?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Features

Input representations extracted from raw data sources.

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

The Machine Learning Canvas (v0.4)

Designed for: | Designed by: | Date: | Iteration:

Prediction

Learning

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

ML task

Input, output to predict, type of problem.

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources

Which raw data sources can we use (internal and external)?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Features

Input representations extracted from raw data sources.

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

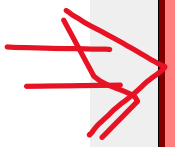
Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

The Machine Learning Canvas (v0.4)

Designed for: | Designed by: | Date: | Iteration:

Domain



Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

ML task

Input, output to predict, type of problem.

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources

Which raw data sources can we use (internal and external)?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Features

Input representations extracted from raw data sources.

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

The Machine Learning Canvas (v0.4)

Designed for: | Designed by: | Date: | Iteration:

Prediction Engine

Decisions	ML task	Value Propositions	Data Sources	Collecting Data
How are predictions used to make decisions that provide the proposed value to the end-user?	Input, output to predict, type of problem.	What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?	Which raw data sources can we use (internal and external)?	How do we get new data to learn from (inputs and outputs)?
Making Predictions When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?	**Offline Evaluation** Methods and metrics to evaluate the system before deployment.		**Features** Input representations extracted from raw data sources.	**Building Models** When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?
	Live Evaluation and Monitoring Methods and metrics to evaluate the system after deployment, and to quantify value creation.			

HOMEWORK

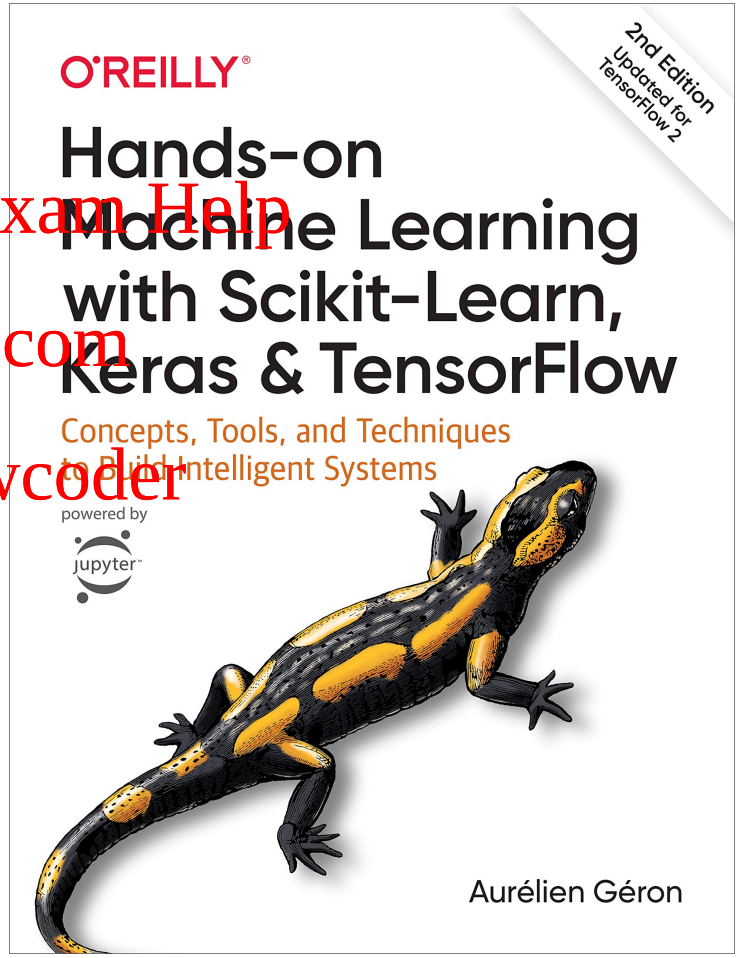
UCL SCHOOL OF MANAGEMENT

Hands-on Machine Learning

Chapter 2: End-to-End Machine Learning Project

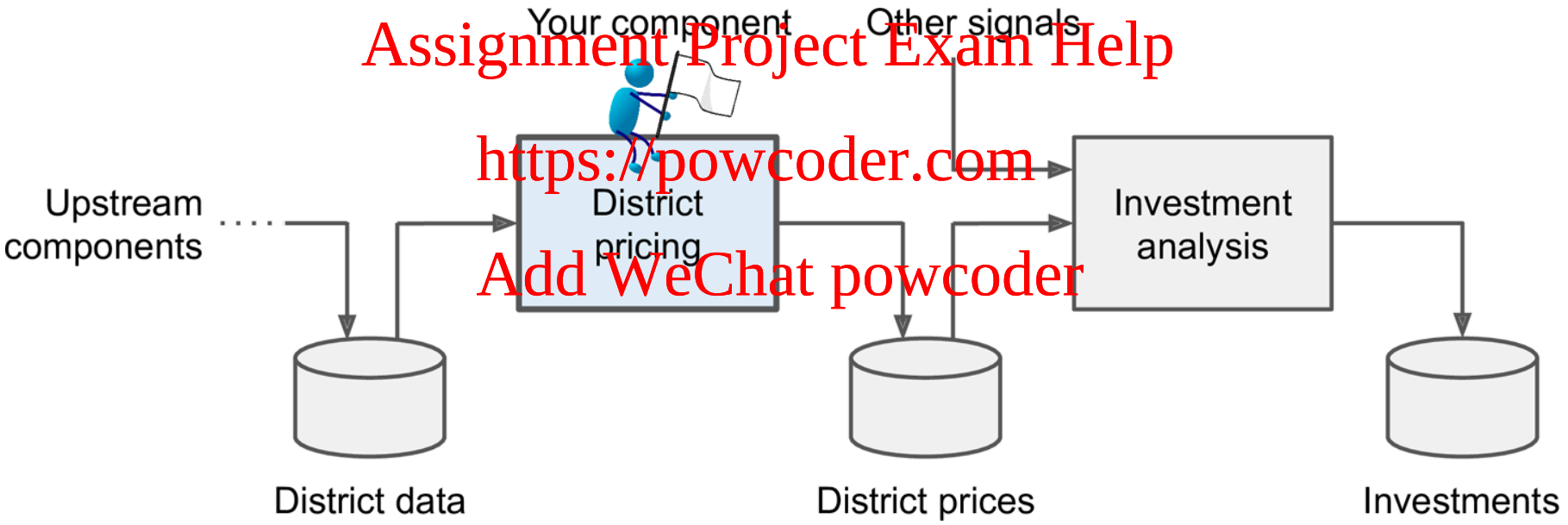
Try reading the Chapter from start to finish. We will work through the problem in class but please come prepared to discuss the case study.

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



PIPELINES

UCL
SCHOOL OF
MANAGEMENT



CALIFORNIA HOUSING DATASET

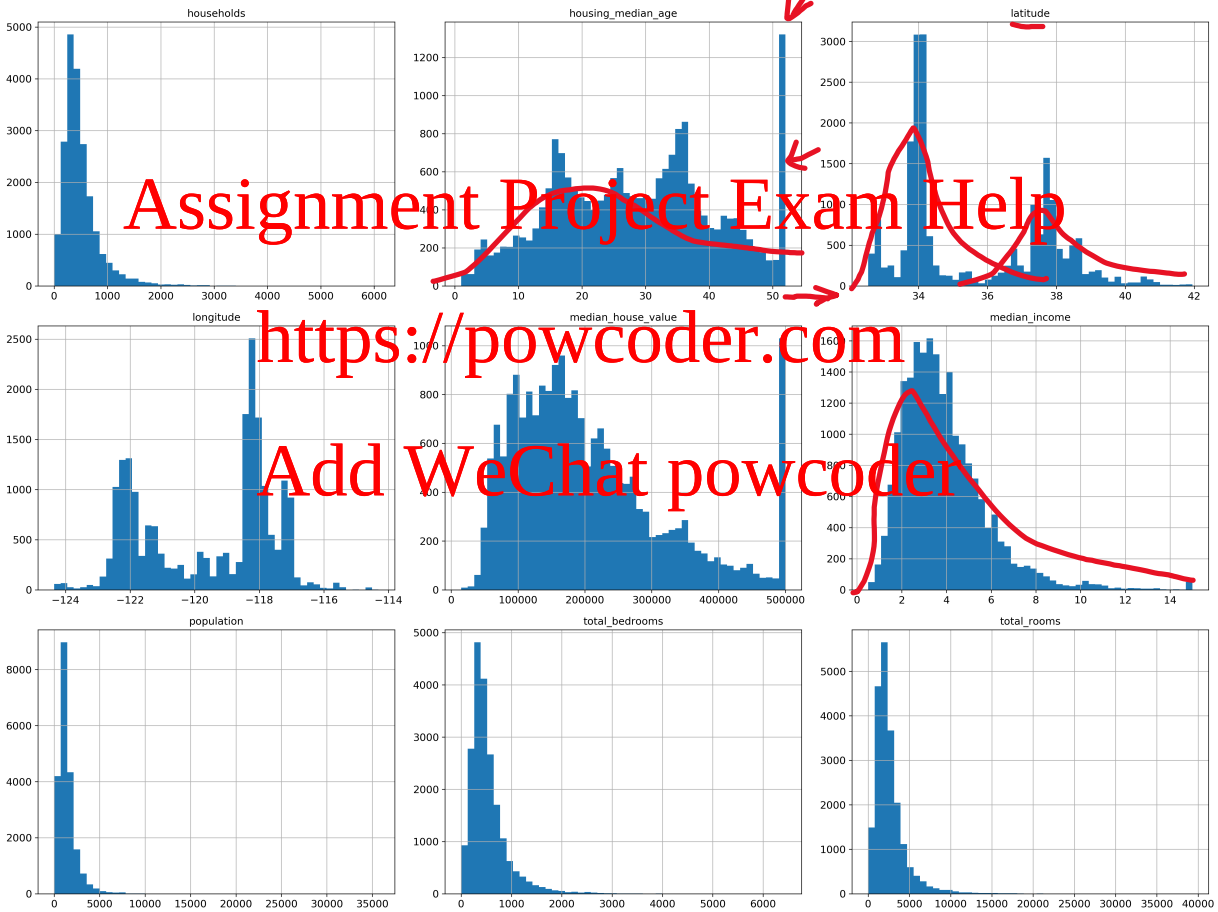
HMLST – CHAPTER 2

UCL SCHOOL OF MANAGEMENT

	longitude	latitude	housing_median_age	total_rooms
count	20640.000000	20640.000000	20640.000000	20640.000000
mean	-119.569704	35.631861	28.639486	2635.763081
std	2.003532	2.135952	12.585558	2181.615252
min	-124.350000	32.540000	1.000000	2.000000
25%	-121.800000	33.930000	18.000000	1447.750000
50%	-118.490000	34.260000	29.000000	2127.000000
75%	-118.010000	37.710000	37.000000	3148.000000
max	-114.310000	41.950000	52.000000	39320.000000

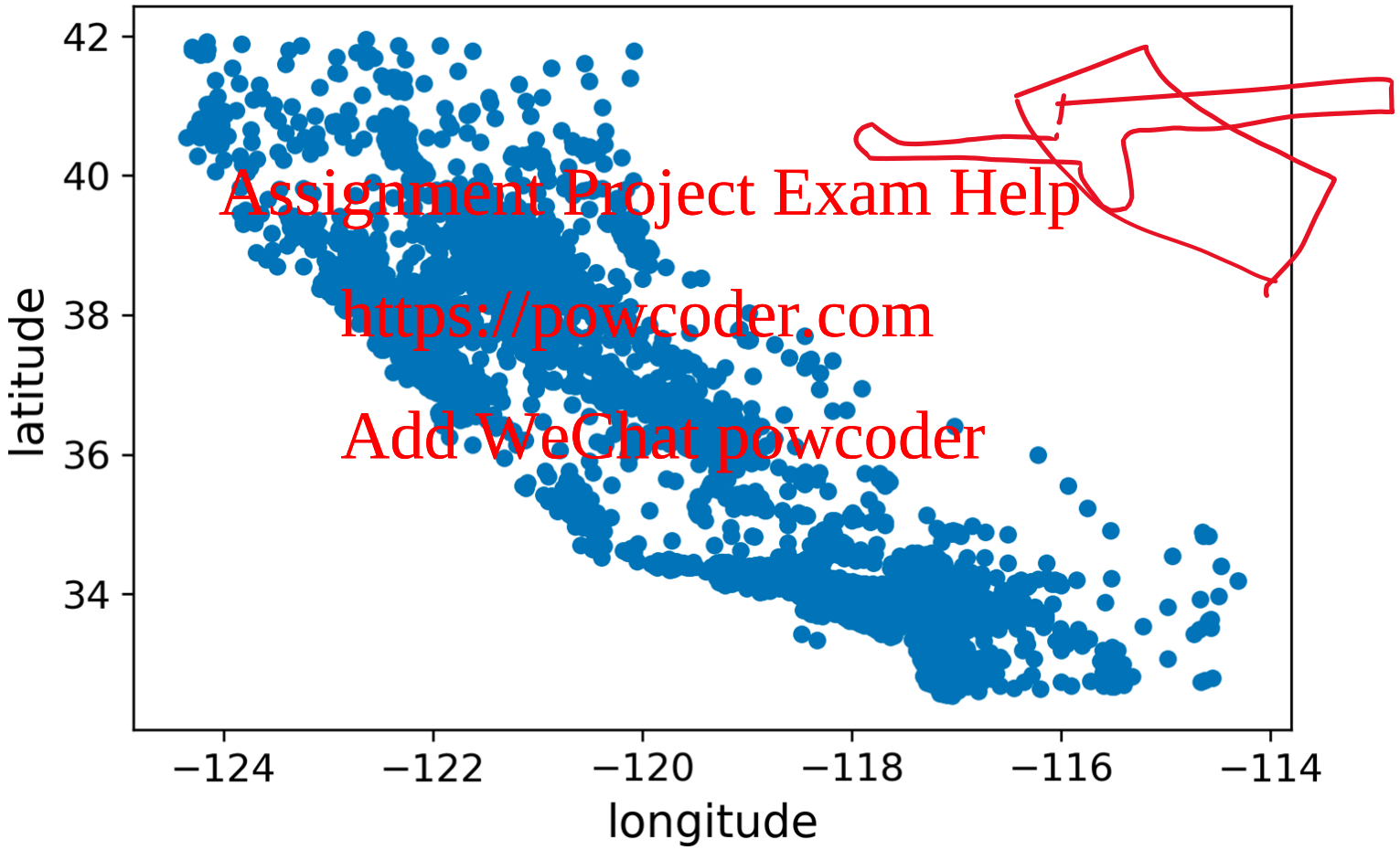
DISTRIBUTIONS

UCL SCHOOL OF MANAGEMENT



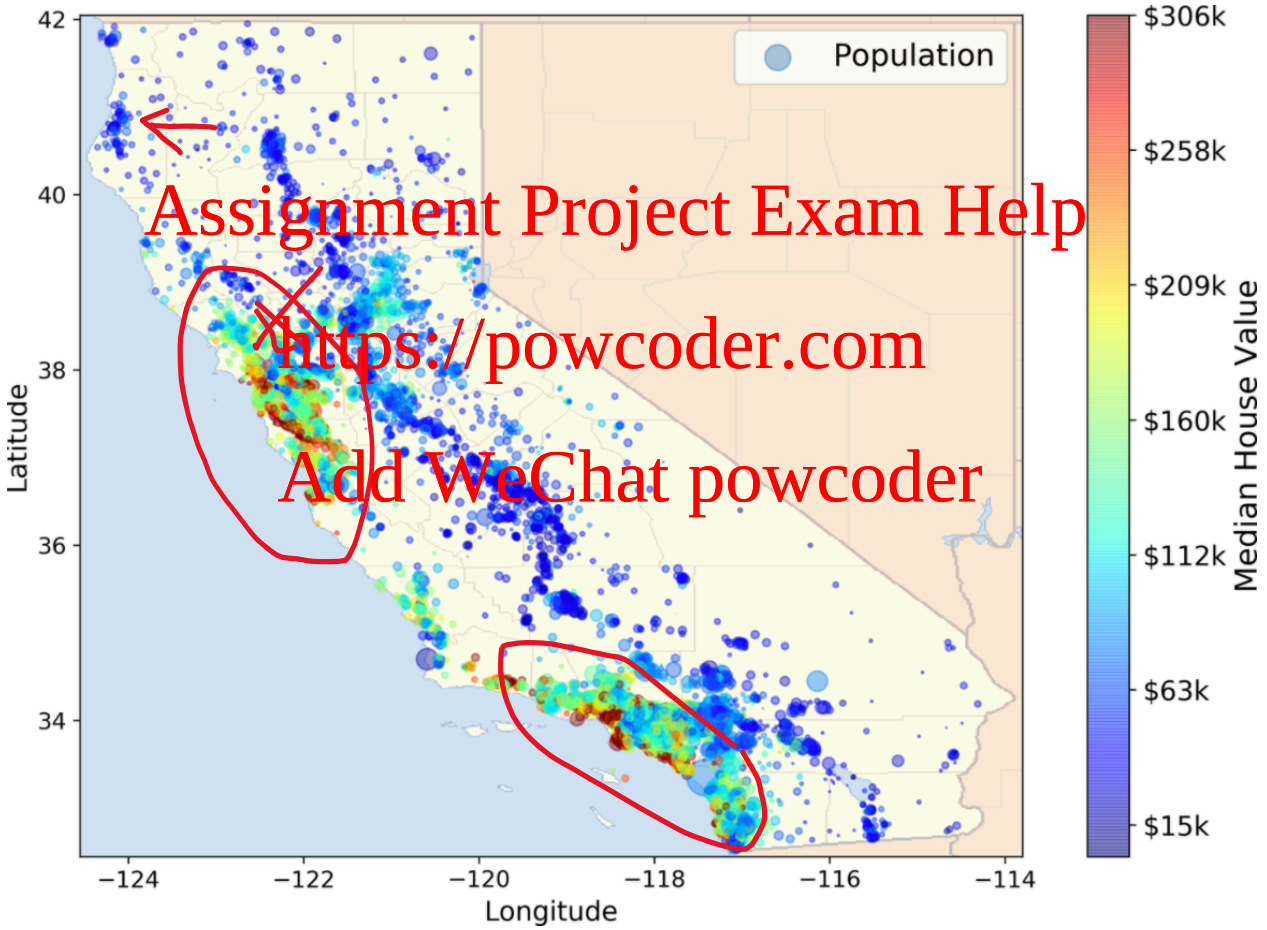
VISUALIZATION

UCL
SCHOOL OF
MANAGEMENT



CALIFORNIA HOUSING PRICES

UCL
SCHOOL OF
MANAGEMENT



END-TO-END

UCL SCHOOL OF MANAGEMENT

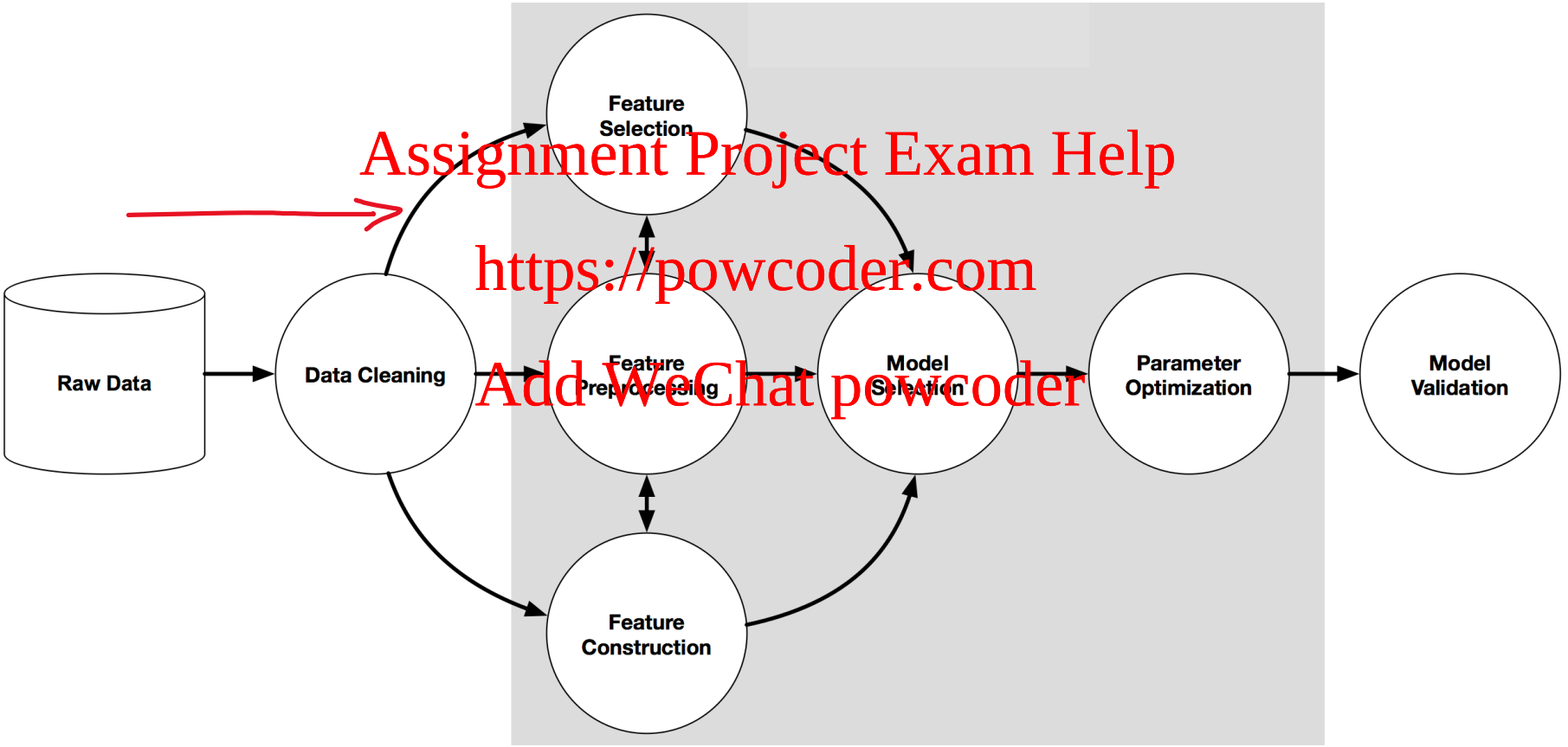
- Discover
- Explore
- Visualize

- Clean
- Sample
- Impute
- Encode
- Transform
- Scale
- Features
- Pipelines
- Training/Validation splits
- Modeling
- Tuning
- Error Analysis

- Documentation
- Presentation
- Launch
- Monitor
- Maintain

ML PIPELINES

UCL SCHOOL OF MANAGEMENT



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

The Machine Learning Canvas (v0.4)

Designed for: | Designed by: | Date: | Iteration:

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

ML task

Input, output to predict, type of problem.

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Data Sources

Which raw data sources can we use (internal and external)?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Features

Input representations extracted from raw data sources.

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

The Machine Learning Canvas (v0.4)

Designed for: Machine Learning Housing Corp. Designed by: A P MOORE Date: 23/01/2020 Iteration: 1

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

- model output to be used in another ML system with other signals.
- inform investment decision for area

ML task

Input, output to predict, type of problem.

- supervised
- multivariate regression
- batch training

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

- build a model of housing prices in California
- predict the median housing price in any district
- reduce manual error rate estimate

Data Sources

Which raw data sources can we use (internal and external)?

- California census data

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Block groups - 300-6000 "districts" people

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

- RMSE
- MAE

Features

Input representations extracted from raw data sources.

population
median income (10)
...
rooms per household
bedrooms per room
population per household

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

LECTURE 1 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

PREDICTIVE ANALYTICS

A P MOORE