

LECTURE 3 TERM 2:

MSIN0097

UCL
SCHOOL OF
MANAGEMENT

PREDICTIVE ANALYTICS Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

A P MOORE

Assignment Project Exam Help

Individual coursework
<https://powcoder.com>

Add WeChat powcoder

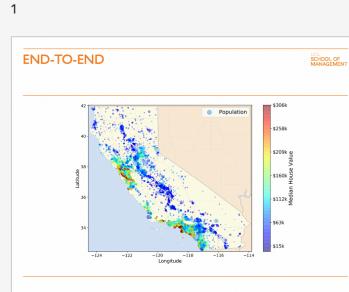
INDIVIDUAL COURSEWORK

Assignment Project Exam Help

MSIN007
SCHOOL OF MANAGEMENT

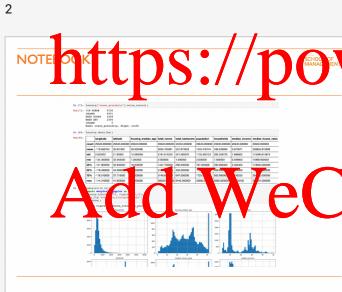
PREDICTIVE ANALYTICS
INDIVIDUAL COURSEWORK

A P MOORE



INDIVIDUAL COURSEWORK
SCHOOL OF MANAGEMENT

— Friday 26th February 2021
— 60% of module mark
— 2000 words

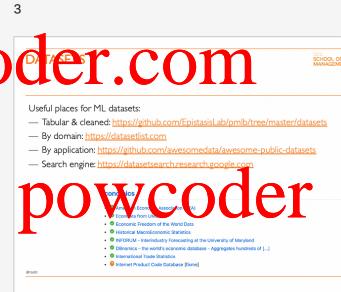


BRIEF
SCHOOL OF MANAGEMENT

The individual coursework task is to identify a dataset and explore building a predictive model using the methods and techniques presented in the first 5 weeks of the course.

There are six main steps:

1. Obtain a dataset and explain the problem you are trying to solve.
This will characterize the type of predictive model you can build
2. Explore the data to gain insights.
Visualize and explain the main trends in the data
3. Prepare the data to better expose the underlying data patterns.
Normalise, remove outliers, encode categorical variables, etc.
4. Train a machine learning model to find better solutions.
Try different models and compare their performance
5. Tune your model to optimise its performance.
6. Present your final solution with any summary conclusions.



GUIDANCE
SCHOOL OF MANAGEMENT

First Pass

Make a first-pass through the project steps as fast as possible. This will give you confidence that you have all the parts that you need and a baseline from which to improve.

Cycles – The process is not linear but cyclic. You will loop between steps, and probably spend most of your time in tight loops between steps 3-4 or 3-4-5 until you achieve a level of accuracy that is sufficient or you run out of time.

The write up in the final submitted Notebook can be more linear – you do not need to include all of your work i.e. including all dead-ends, and it should be concise and consistent.

GUIDANCE
SCHOOL OF MANAGEMENT

Attempt Every Step

It is easy to skip steps, especially if you are not confident or familiar with the tasks that step. Try and do something at each step in the process, even if it does not improve accuracy. You can always build upon it later. Don't skip steps, just reduce their contribution to your final submission as necessary.

Ratchet Accuracy

The goal of the project is to achieve good model performance (which ever metric you use to measure this). Every step contributes towards this goal. Set some simple benchmarks early on. Treat changes that you make as experiments that potentially increase accuracy.

Performance is a ratchet that can only move in one direction (better, not worse).

GUIDANCE
SCHOOL OF MANAGEMENT

Adapt As Needed

Modify the steps as you need on a project, especially as you become more experienced with using the Notebook.

The final submitted Notebook does not need to preserve the suggested structure if you think something else is more appropriate.

GUIDANCE
SCHOOL OF MANAGEMENT

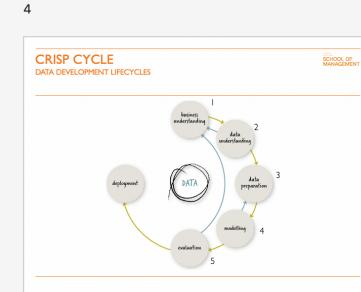
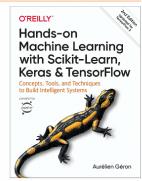
Appendix B. Machine Learning Project Checklist

The checklist can guide you through a Machine Learning project. There are eight main steps:

1. Define the problem and the hypothesis.
2. Collect the data.
3. Explore the data and gain insights.
4. Prepare the data to better expose the underlying data patterns.
5. Train a machine learning model to find better solutions.
6. Tune your model to optimise its performance.
7. Present your solution.
8. Evaluate and refine the model to work.

From the Problem and Look at the Big Picture

- 1. Define the objective & hypothesis.
- 2. What did you collect for data?
- 3. What are the main challenges in the project?



KEY DATES
SCHOOL OF MANAGEMENT

— Submission Friday 26th February 2021, 10 am

Assignment Project Exam Help

<https://powcoder.com>
Group coursework
Add WeChat powcoder

COURSEWORK / INDUSTRY REPORT

Assignment Project Exam Help

The grid contains 24 numbered screenshots of a coursework report:

- 1: Warner Music Group logo and title page.
- 2: Table of Contents.
- 3: List of Figures.
- 4: Introduction section with red annotations pointing to the first few paragraphs.
- 5: Part I - Preparation and Evaluation section.
- 6: Feature Engineering section.
- 7: Part II - Prediction section.
- 8: Another part of the Prediction section.
- 9: Data visualization with bar charts.
- 10: Data visualization with scatter plots.
- 11: Data visualization with histograms.
- 12: Data visualization with scatter plots.
- 13: Data visualization with histograms.
- 14: Data visualization with scatter plots.
- 15: Data visualization with box plots.
- 16: Data visualization with scatter plots.
- 17: A section titled "A. Warner Artist Cleaning".
- 18: A section titled "A.1 Cleaning Methods".
- 19: A section titled "A.2 Cleaning Results".
- 20: A section titled "B. Target Analysis".
- 21: A section titled "C. Model Selection".
- 22: Annexes section.
- 23: Annexes section.
- 24: Annexes section.

MACHINE LEARNING JARGON

- 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
 - Partition
 - Data density
 - Hidden parameter
 - High dimensional space
 - Low dimensional space
 - Separable data
 - Manifold / Decision surface
 - Hyper cube / volume / plane
- Assignment Project Exam Help**
<https://powcoder.com>
- Add WeChat powcoder

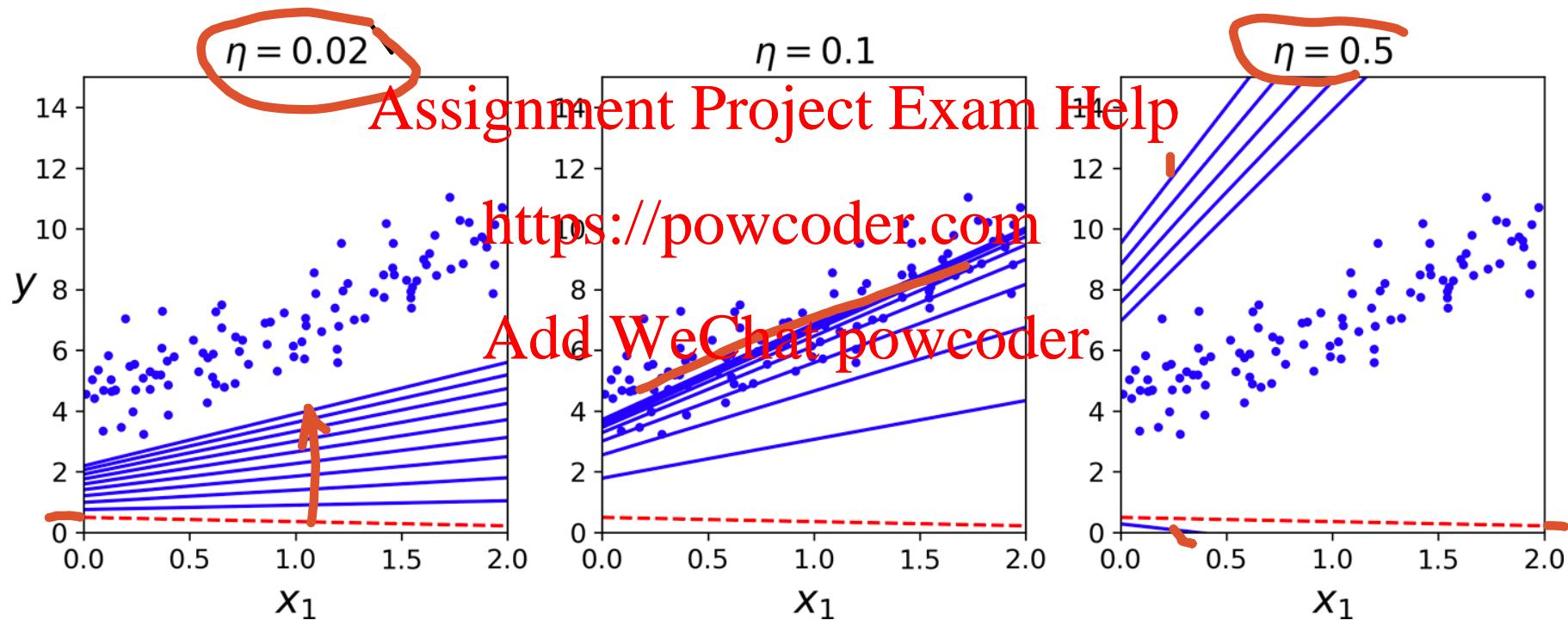
Assignment Project Exam Help

<https://powcoder.com>

Homework...

Add WeChat powcoder

LEARNING RATES



Assignment Project Exam Help

<https://powcoder.com>
Reviewing notebooks

Add WeChat powcoder

```
1 theta_path_bgd = []
2
3 def plot_gradient_descent(theta, eta, theta_path=None):
4     m = len(y)
5     plt.plot(X, y, "b.")
6     n_iterations = 1000
7     for iteration in range(n_iterations):
8         if iteration < 10:
9             y_predict = X_new_b.dot(theta)
10            style = "b--" if iteration > 0 else "r--"
11            plt.plot(X_new, y_predict, style)
12            gradients = 2/m * X_b.T.dot(X_b.dot(theta) - y)
13            theta = theta - eta * gradients
14            if theta_path is not None:
15                theta_path.append(theta)
16            plt.xlabel("$x_1$", fontsize=18)
17            plt.axis([0, 2, 0, 15])
18            plt.title(r"$\eta = {}$".format(eta), fontsize=16)
```

ALTERING CODE

```
1 theta_path_bgd = []
2
3 def plot_gradient_descent(theta, eta, theta_path=None):
4     m = len(X_b)
5     plt.plot(X, y, "b.")
6     n_iterations = 1000
7     n_alpha = 0.1;
8     for iteration in range(n_iterations):
9         if iteration < 20:
10             y_predict = X_new_b.dot(theta)
11             if iteration > 0: # changed alpha setting to make direction of iterations clear
12                 style = "b-"
13                 plt.plot(X_new, y_predict, style, alpha=n_alpha)
14             n_alpha = n_alpha + 0.01
15         else:
16             style = "r--"
17             plt.plot(X_new, y_predict, style, alpha=1)
18
19         gradients = 2/m * X_b.T.dot(X_b.dot(theta) - y)
20         theta = theta - eta * gradients
21         if theta_path is not None:
22             theta_path.append(theta)
23     plt.xlabel("$x_1$", fontsize=18)
24     plt.axis([0, 2, 0, 15])
25     plt.title(r"$\eta = {}$".format(eta), fontsize=16)
26     plt.text(0.2, 13, r"$m = {}$".format(float(theta[0])), fontsize=12)
27     plt.text(0.2, 12, r"$c = {}$".format(float(theta[1])), fontsize=12)
```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```
1 theta_path_bgd = []
2
3 def plot_gradient_descent(theta, eta, theta_path=None):
4     m = len(X_b)
5     plt.plot(X, y, "b.")
6     n_iterations = 1000
7     for iteration in range(n_iterations):
8         if iteration < 10:
9             y_predict = X_new_b.dot(theta)
10            style = "b-" if iteration > 0 else "r--"
11            plt.plot(X_new, y_predict, style)
12            gradients = 2/m * X_b.T.dot(X_b.dot(theta) - y)
13            theta = theta - eta * gradients
14            if theta_path is not None:
15                theta_path.append(theta)
16            plt.xlabel("$x_{1\$}", fontsize=18)
17            plt.axis([0, 2, 0, 15])
18            plt.title(r"$\eta = {}$".format(eta), fontsize=16)
```

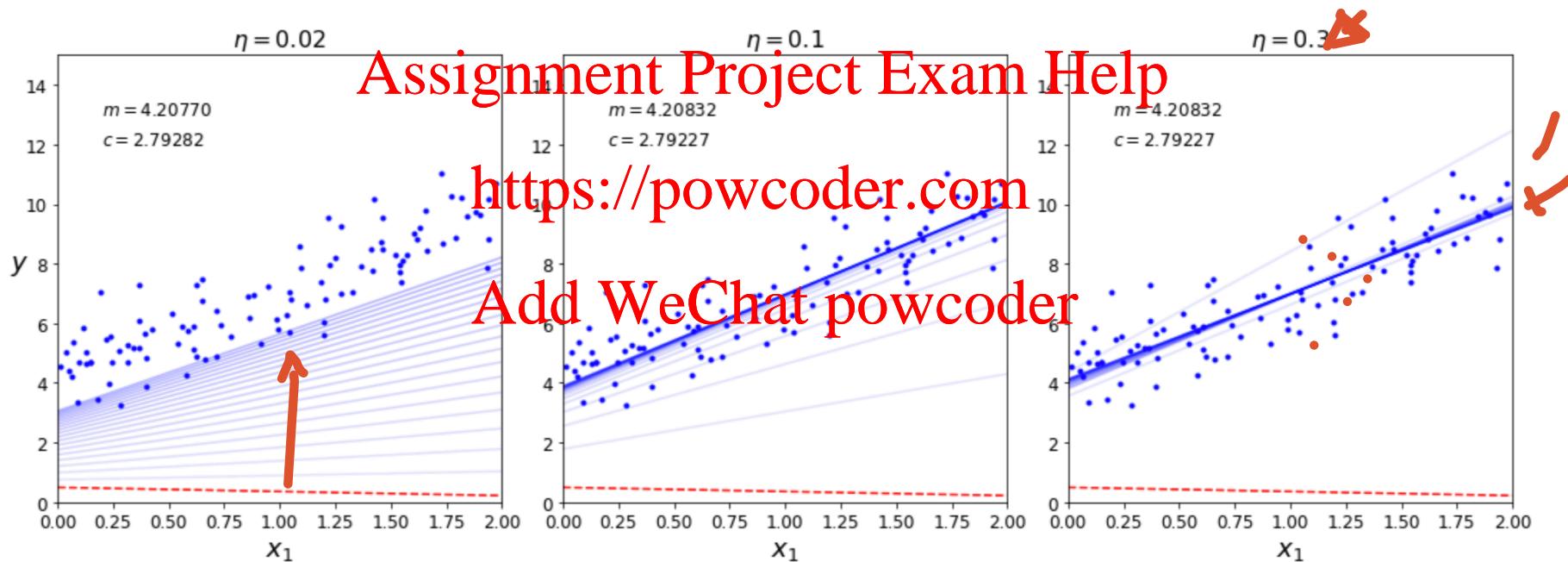
```
1 theta_path_bgd = []
2
3 def plot_gradient_descent(theta, eta, theta_path=None):
4     m = len(X_b)
5     plt.plot(X, y, "b.")
6     n_iterations = 1000
7     n_alpha = 0.1;
8     for iteration in range(n_iterations):
9         if iteration < 20:
10            y_predict = X_new_b.T.dot(theta)
11            if iteration > 0: # changed alpha setting to make direction of iterations clear
12                style = "b-"
13                plt.plot(X_new, y_predict, style, alpha=n_alpha)
14                n_alpha = n_alpha + 0.01
15            else:
16                style = "r--"
17                plt.plot(X_new, y_predict, style, alpha=1)
18
19            gradients = 2/m * X_b.T.dot(X_b.dot(theta) - y)
20            theta = theta - eta * gradients
21            if theta_path is not None:
22                theta_path.append(theta)
23            plt.xlabel("$x_{1\$}", fontsize=18)
24            plt.axis([0, 2, 0, 15])
25            plt.title(r"$\eta = {}$".format(eta), fontsize=16)
26            plt.text(0.2, 13, r"$m = {:.5f}$".format(float(theta[0])), fontsize=12)
27            plt.text(0.2, 12, r"$c = {:.5f}$".format(float(theta[1])), fontsize=12)
```

Assignment Project Exam Help

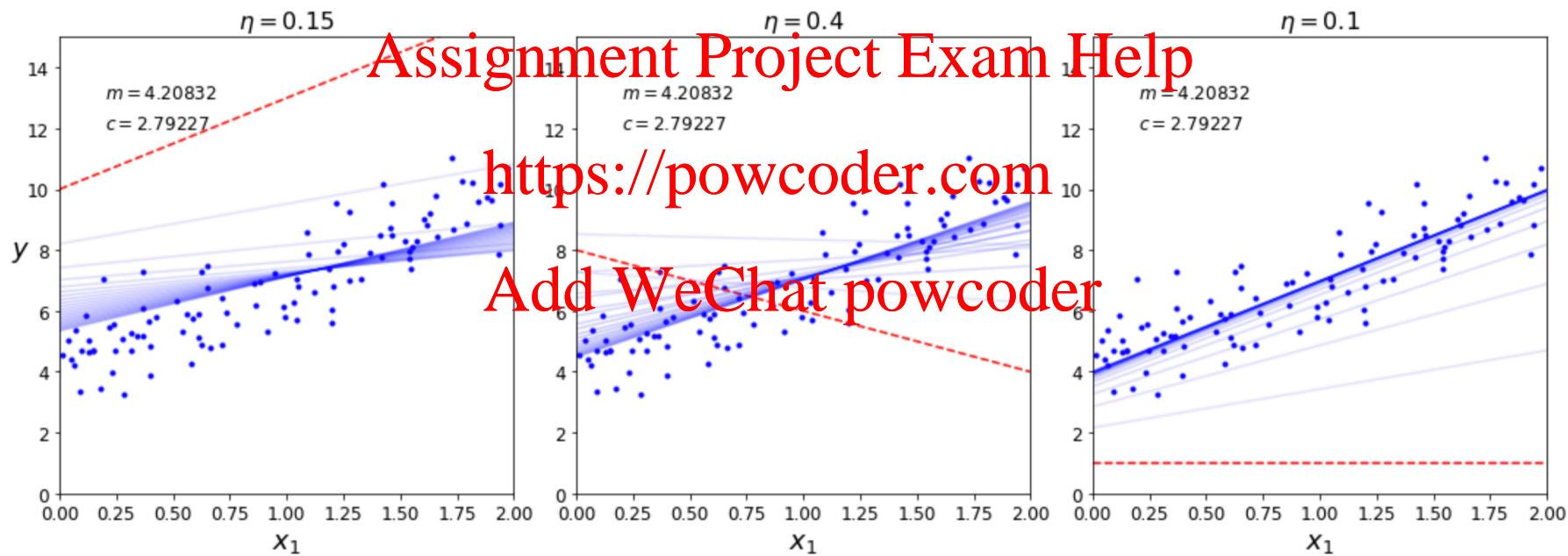
<https://powcoder.com>

Add WeChat powcoder

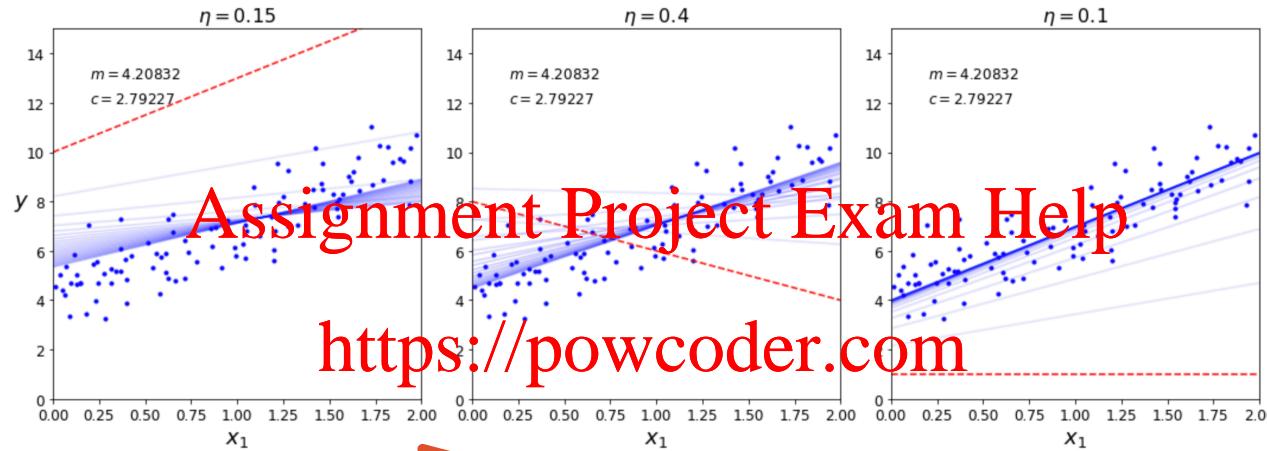
LEARNING RATES



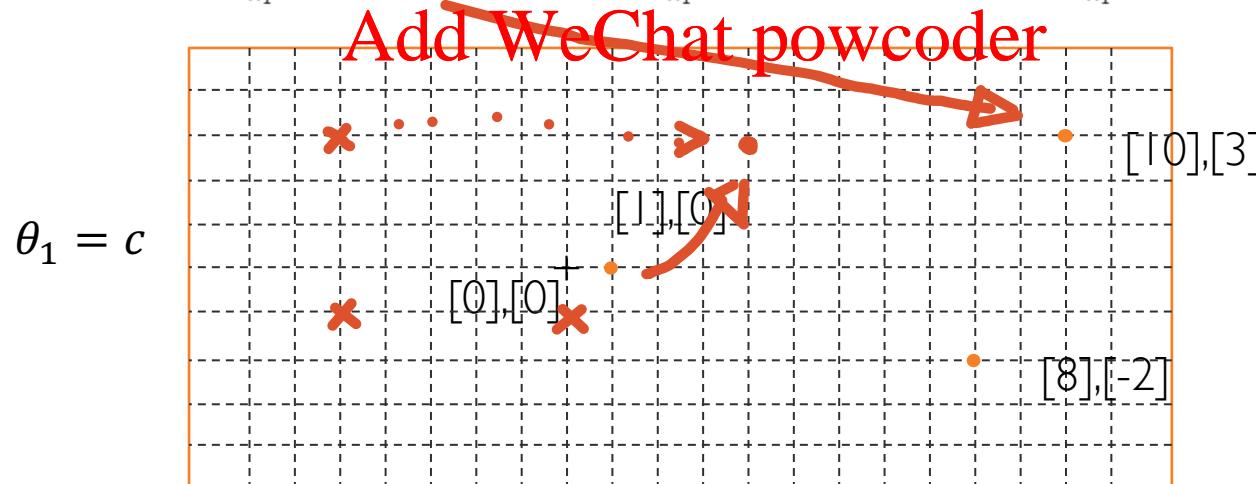
LEARNING RATES



LEARNING RATES

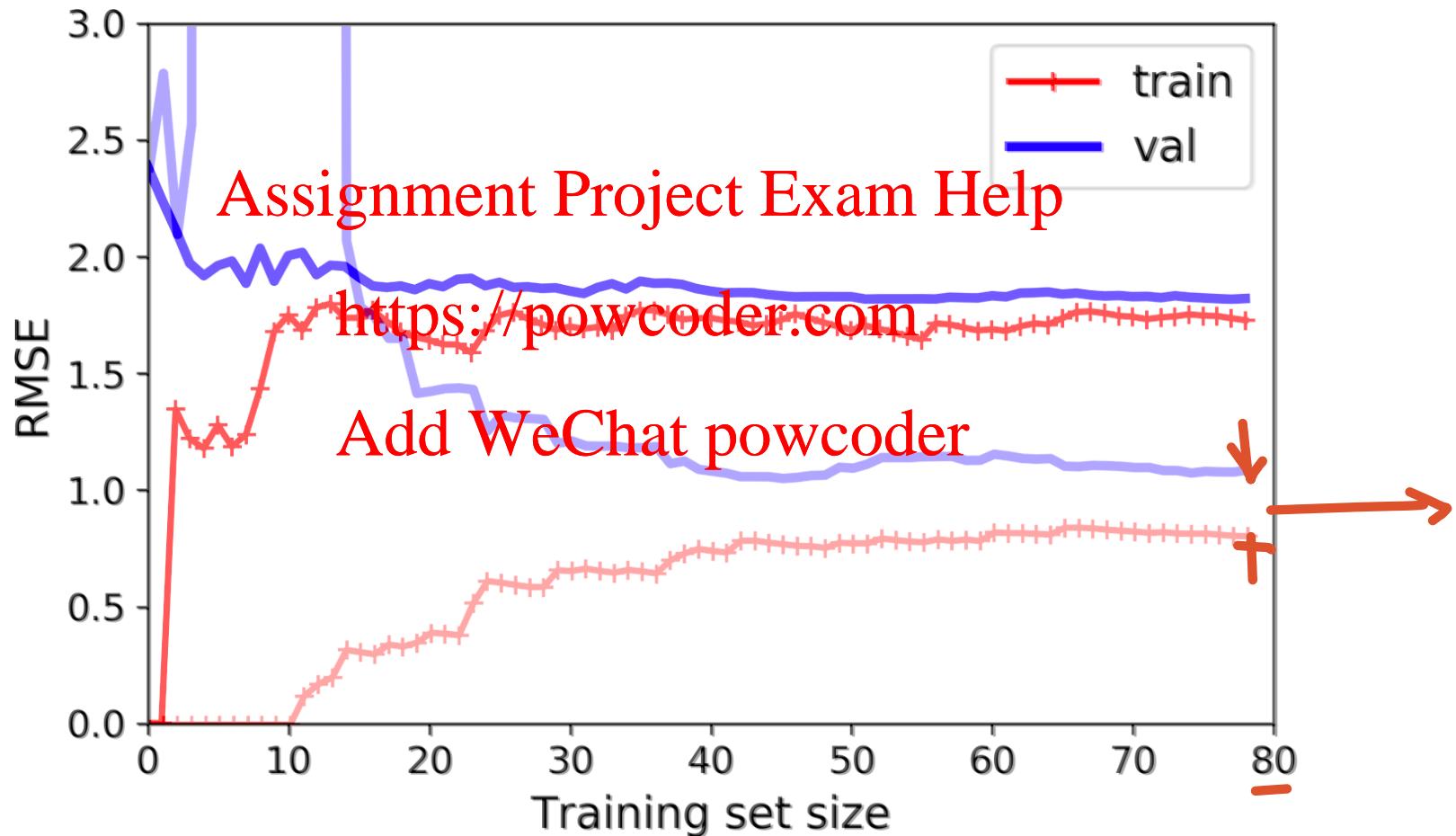


Assignment Project Exam Help
<https://powcoder.com>



$$\theta_0 = m$$

LEARNING CURVES



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

A visual introduction Assignment Project Exam Help machine learning

<https://powcoder.com>

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.

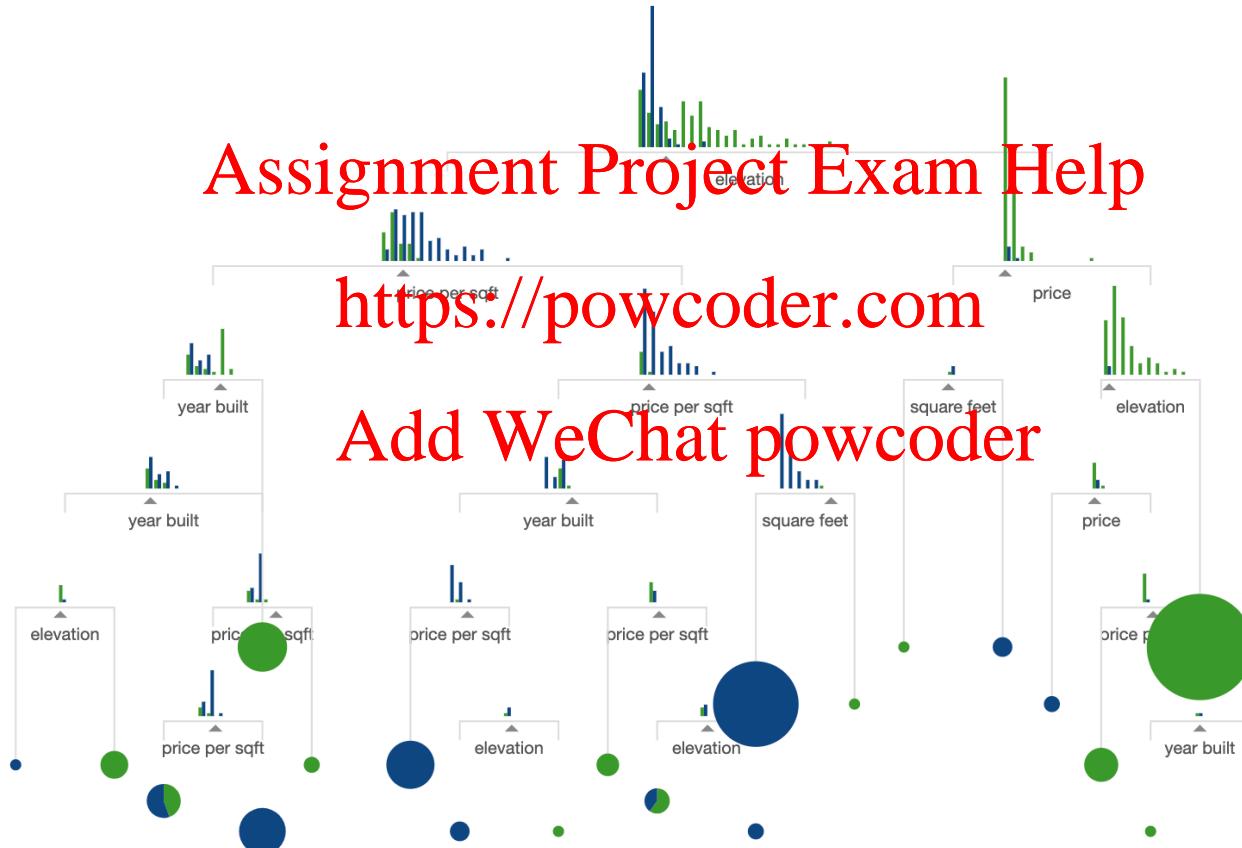
Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



MODEL VARIANCE / 模型方差

MÓXÍNG FĀNGCHĀ

UCL
SCHOOL OF
MANAGEMENT

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

MODEL GENERALIZATION / 模型泛化

MÓXÍNG FÀN HUÀ

UCL
SCHOOL OF
MANAGEMENT

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

OVERFITTING / 过拟合

GUÒ NÍ HÉ

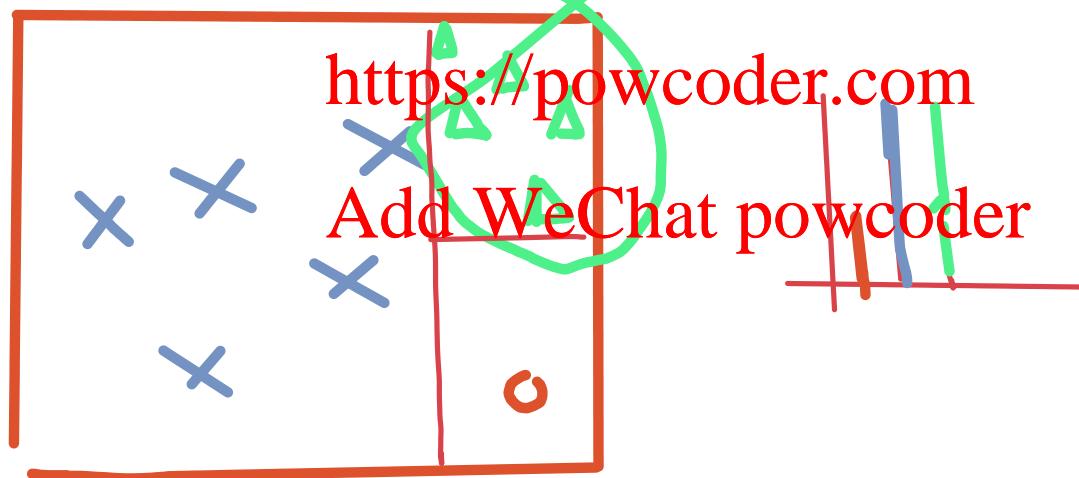
Assignment Project Exam Help
<https://powcoder.com>

Add WeChat powcoder

FAILURE MODES / 失败模式

SHĪBÀI MÓSHÌ

Assignment Project Exam Help



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

— What is the benefit of out-of-bag evaluation? ✗

— If a Decision Tree is overfitting the training set, is it a good idea to try decreasing `max_depth`? ✓

<https://powcoder.com>

— If a Decision Tree is underfitting the training set, is it a good idea to try scaling the input features? ✓ Add WeChat powcoder

— What problems might we have if we try to grow a tree with high class imbalance? ✓

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- **Domain expertise**

- determined by talking to subject matter experts.

- **Tuning**

Assignment Project Exam Help

- determined by a hyperparameter search such as a grid search.

- **Heuristic**

<https://powcoder.com>

- specified using a general best practice

Add WeChat powcoder

Applications

- Fraud Detection
- Claim Prediction
- Churn Prediction
- Spam Detection

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

— Performance Metrics

- F-measure
- G-mean

Assignment Project Exam Help

— Data Sampling

- SMOTE (Synthetic Minority OverSampling Technique)
- ENN (Edited nearest neighbours)

Add WeChat powcoder

— Cost-Sensitive Algorithms

- Decision Trees

— Post-Processing

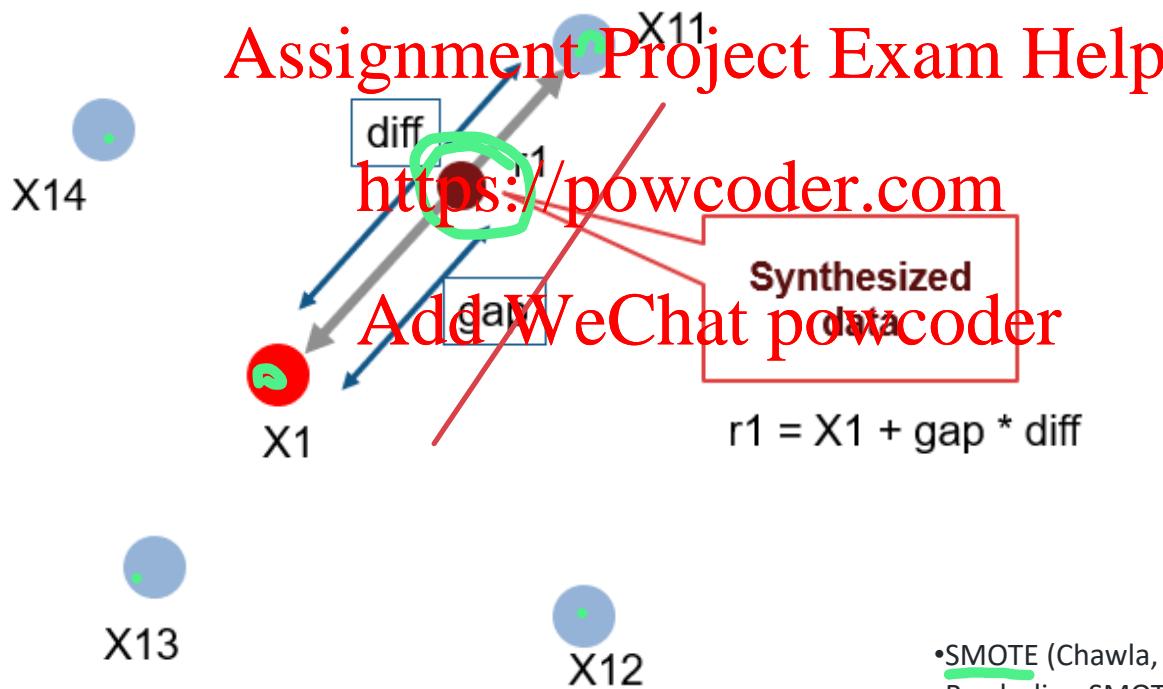
- Threshold Moving
- Calibration

SYNTHETIC DATA SAMPLING



Minority class

Example case with k = 4



- SMOTE (Chawla, NV. et al. 2002)
- Borderline SMOTE (Han, H. et al. 2005)
- ADASYN (He, H. et al. 2008)
- Safe-level SMOTE (Bunkhumpornpat, C. et al. 2009)

Assignment Project Exam Help

<https://powcoder.com>

Ensembles

Add WeChat powcoder

If you have trained five different models on the exact same training data, and they all achieve 95% precision, is there any chance that you can combine these models to get better results?

Assignment Project Exam Help

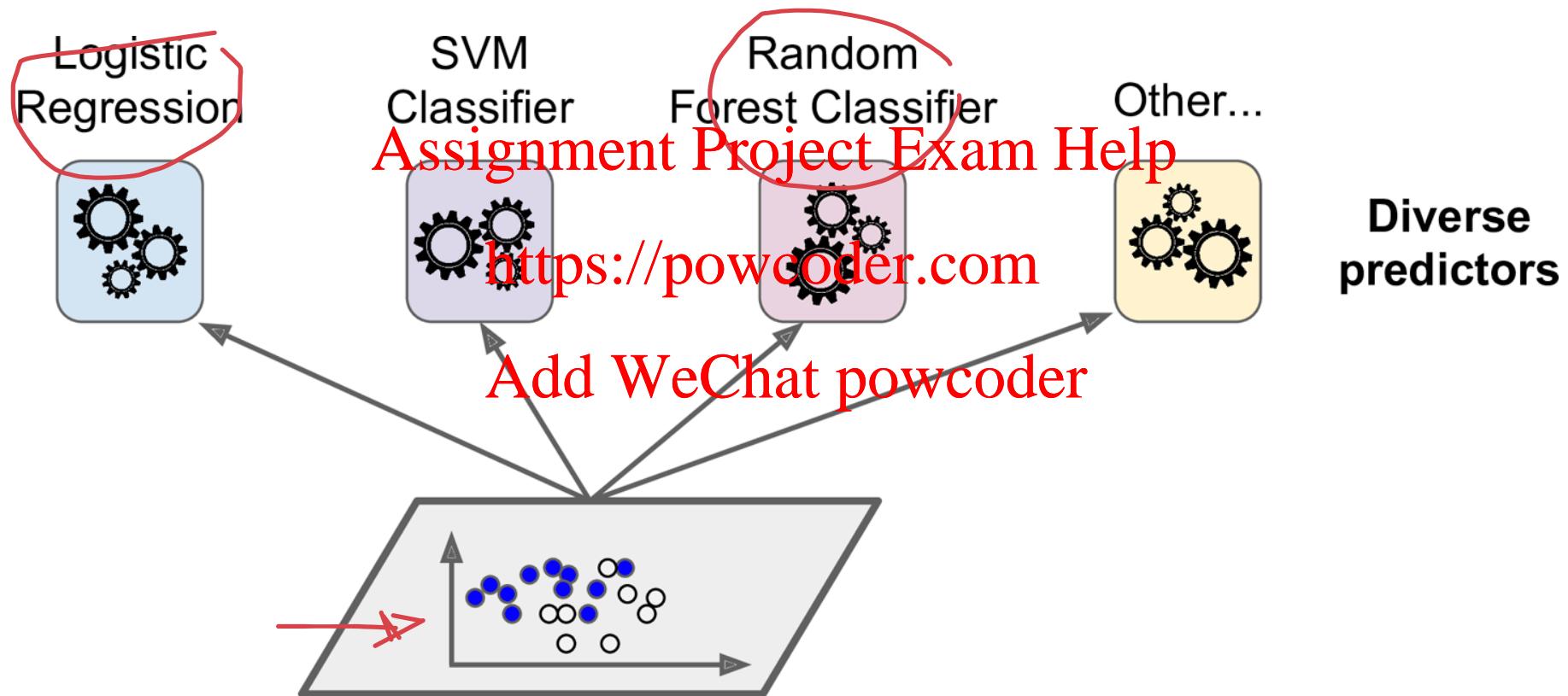
If so, how?

<https://powcoder.com>

If not, why?

Add WeChat powcoder

MULTIPLE MODELS



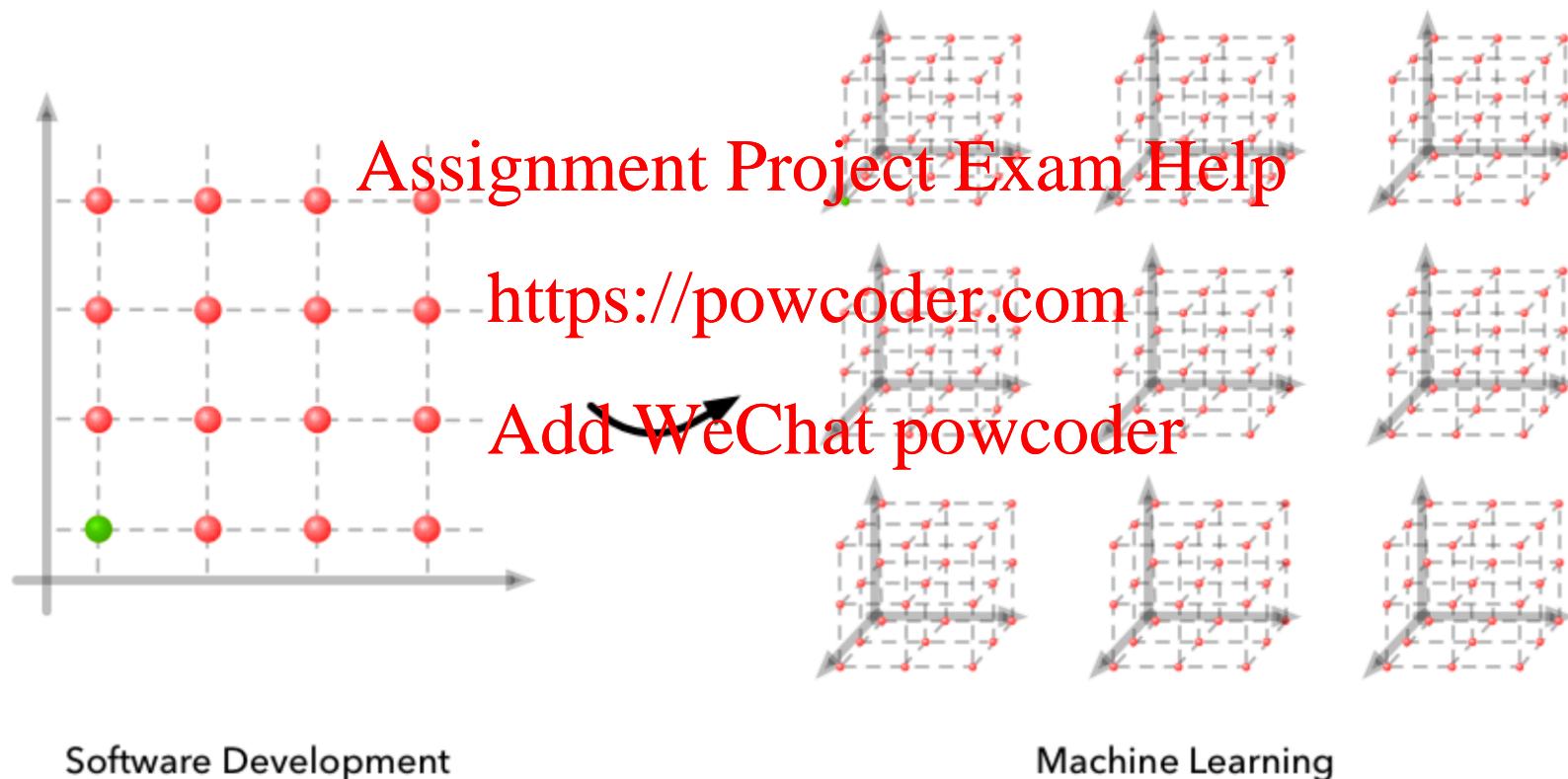


[Leipzig–Dresden Railway Company](#) in 1852

Assignment Project Exam Help

<https://powcoder.com>
Why is ML hard?
Add WeChat powcoder

WHY IS ML HARD?



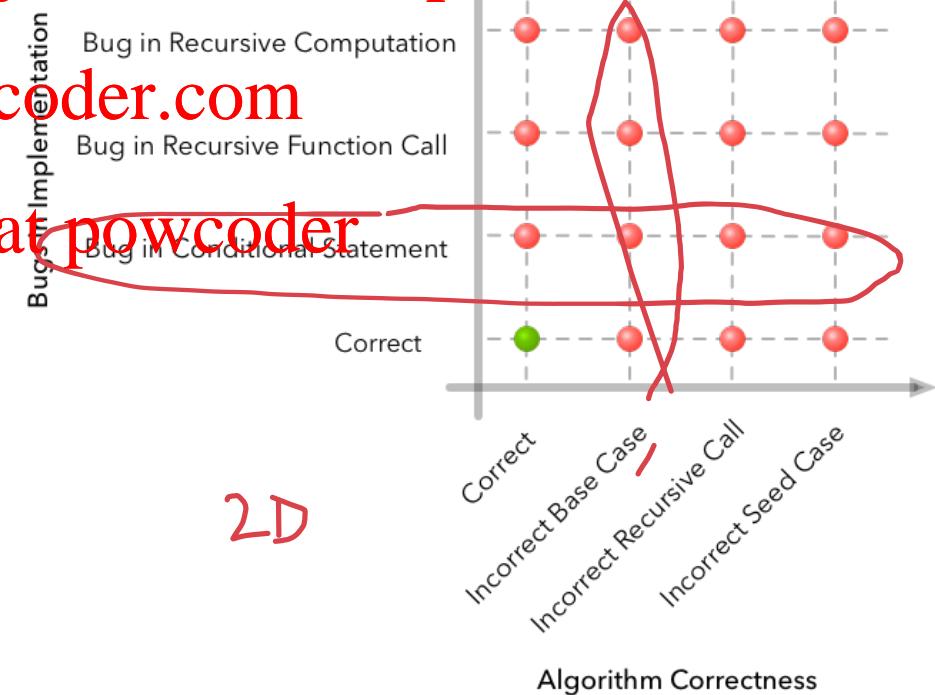
DEBUGGING

```
def recursion(input):  
    if input is endCase:  
        return transform(input)  
    else:  
        return recursion(transform(input))
```

Assignment Project Exam Help

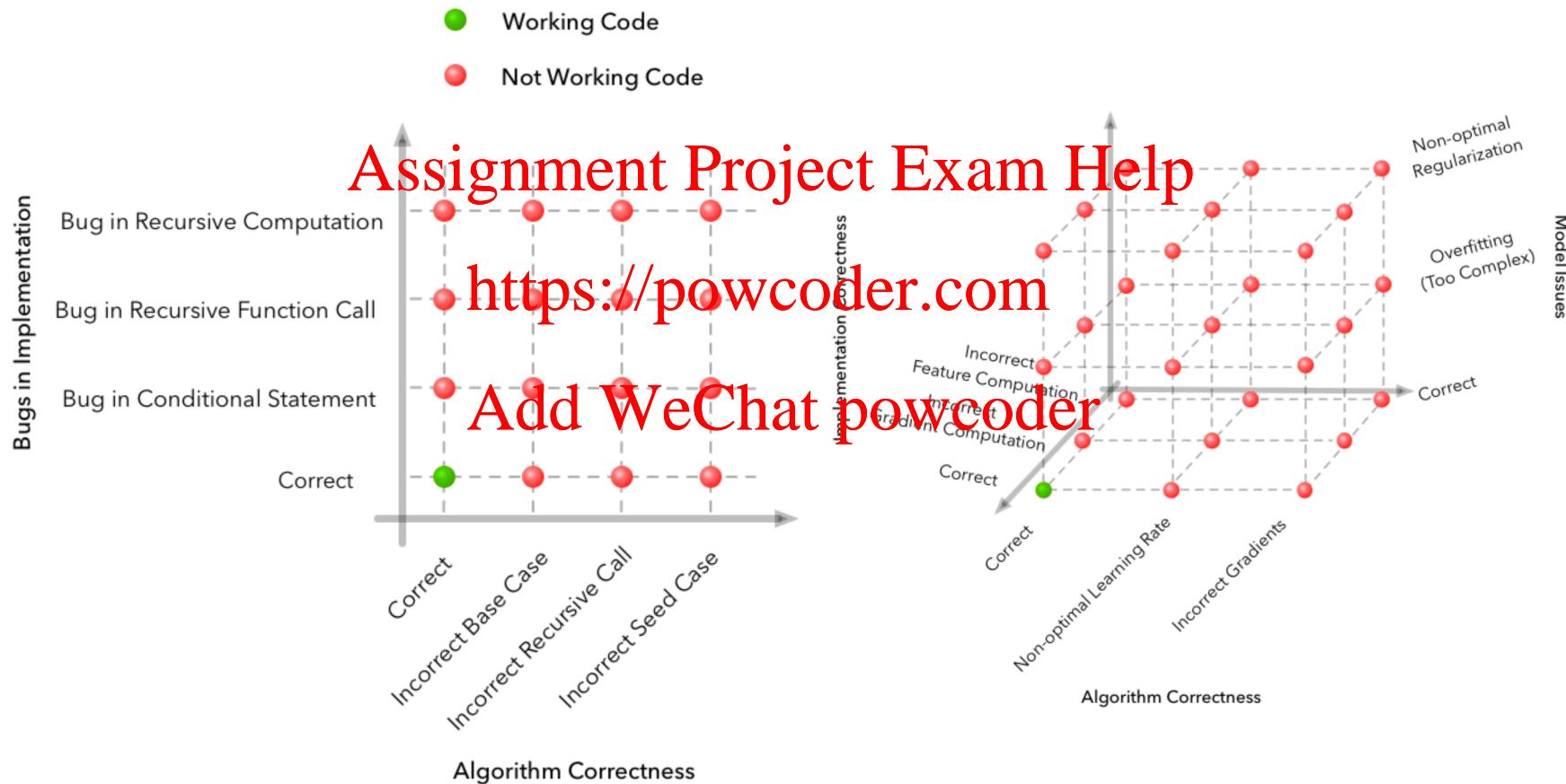
<https://powcoder.com>

Add WeChat powcoder



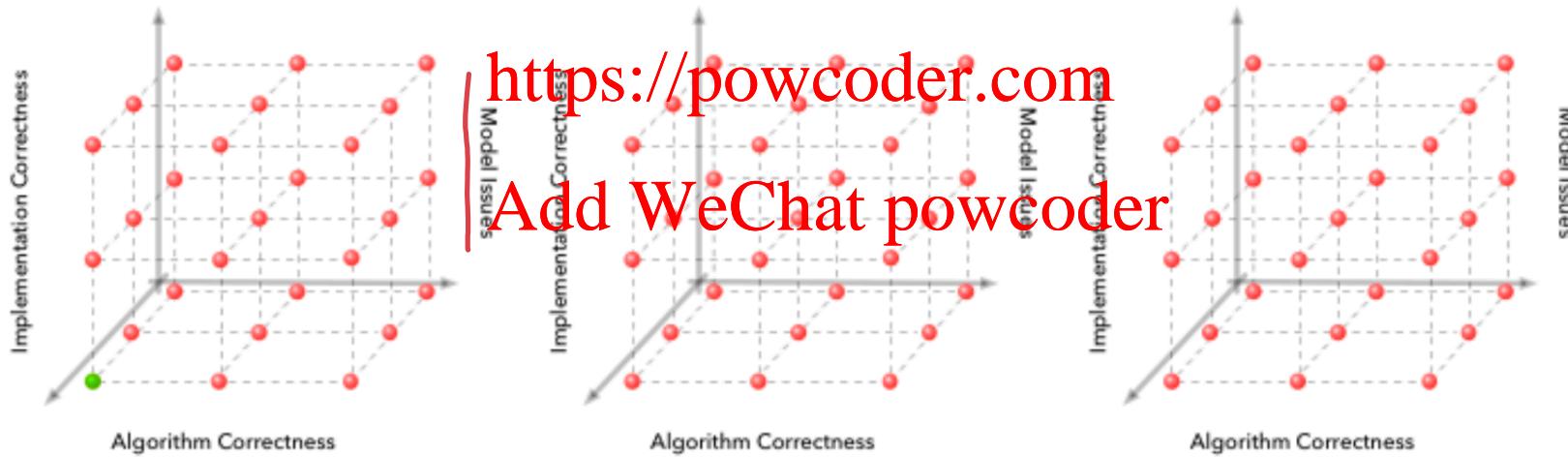
WHY IS ML HARD?

ALGORITHM, IMPLEMENTATION, DATA, MODEL



WHY IS ML HARD?

Enough Correct Data Assignment Project Exam Help Not Enough Data Weak Labels





Zachary Lipton
@zacharylipton

The dominant practice of the applied machine learnist
has shifted from ad-hoc feature hacking (2000s) to ad-
hoc architecture hacking (2010s) to ad-hoc pre-training
hacking (2020s).

4:35 am · 28 Jan 2021 · Twitter Web App

~~Assignment Project Exam Help~~

<https://powcoder.com>



Judea Pearl @yudapearl · 2h
Replying to @zacharylipton

But is ad-hoc-ness a curable weakness of ML? I have a theory that it can
be cured only by having a model of right and wrong, namely, a model of
the world outside the data themselves. Such a model, unfortunately, is not
in the vocabulary of applied ML.

2

2

29

1

Tip

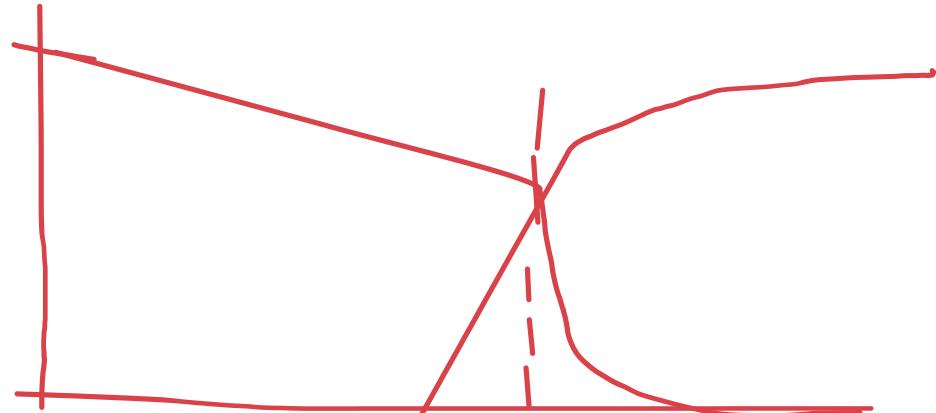
~~Add WeChat powcoder~~

@chipro
@random_forests
@zacharylipton
@yudapearl

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



HAVE YOUR SAY!

The screenshot shows a Moodle course page with the following structure:

- Section 1:** "Have your say" (with edit icon) contains:
 - A gear icon.
 - A large red title: "Assignment Project Exam Help".
 - A "Have your say" button.
 - A text block: "With the move to Connected Learning, we are keen to gather early module feedback from students, to check you have been able to access the relevant resources; that you understand the purpose of the teaching and content; that you understand how you are assessed; and that you know where to get help if you need it."
 - A note: "By answering 5 simple yes/no questions, we will be able to identify areas for improvement early enough to take positive action whilst you are studying the module."
 - A "Thanks for your cooperation!" message.
 - A large red text: "Add WeChat powcoder".
- Section 2:** "Early Module Feedback Questionnaire MSIN0097" (with edit icon) contains:
 - A gear icon.
 - A "Edit" button.
 - A "Add an activity or resource" button.

"The questionnaires are very short and will take less than a minute for them to complete."

VOTING IN ACTION!

Best module of this term so far. Engaging lectures and comprehensive Async materials.

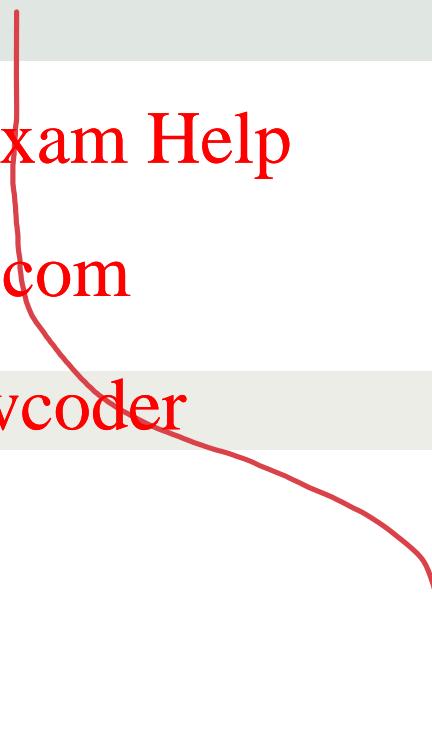
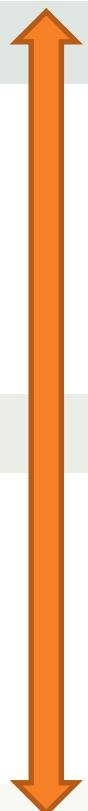
Assignment Project Exam Help

<https://powcoder.com>

Many complex things to learn from scratch. A little bit overwhelming.

Add WeChat powcoder

The concept of giving us a book to learn by ourself is against visiting a university which explains the context in a good way.



TEACHING TEAM



Dr Alastair Moore
Senior Teaching Fellow
a.p.moore@ucl.ac.uk
@latticecut



Kamil Tylinski
Teaching Assistant
kamil.tylinski.16@ucl.ac.uk



Jiangbo Shangguan
Teaching Assistant
j.shangguan.17@ucl.ac.uk

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

LECTURE 3 TERM 2:

MSIN0097

UCL
SCHOOL OF
MANAGEMENT

PREDICTIVE ANALYTICS Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

A P MOORE