

LECTURE 5 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

Assignment Project Exam Help

Individual Coursework assignment has

<https://powcoder.com>

been extended by one week

Add WeChat powcoder

to **Friday 5th March 2021** at **10:00 am**

USING OTHER PEOPLE'S CODE

When you copy code from stackoverflow
and GitHub



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
- Model variance
- Model bias
- Model generalization
- Overfitting
- Goodness-of-fit
- Hyper-parameters
- Failure modes
- Confusion matrix
- True Positive
- False Negative
- Partition
- Margin
- 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

A. Classification

Model requirements

Classification

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

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

B. Regression

Model requirements

Regression

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

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

Add WeChat powcoder

C. Clustering

Model requirements

Clustering

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

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

Hidden variables

Density estimation

D. Decomposition

Model requirements

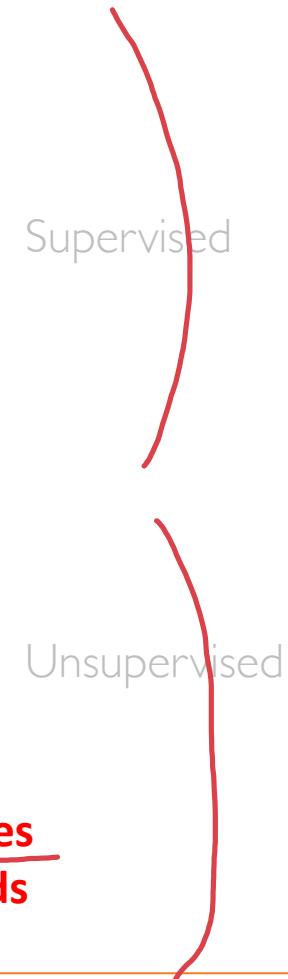
Decomposition

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

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

Subspaces

Manifolds



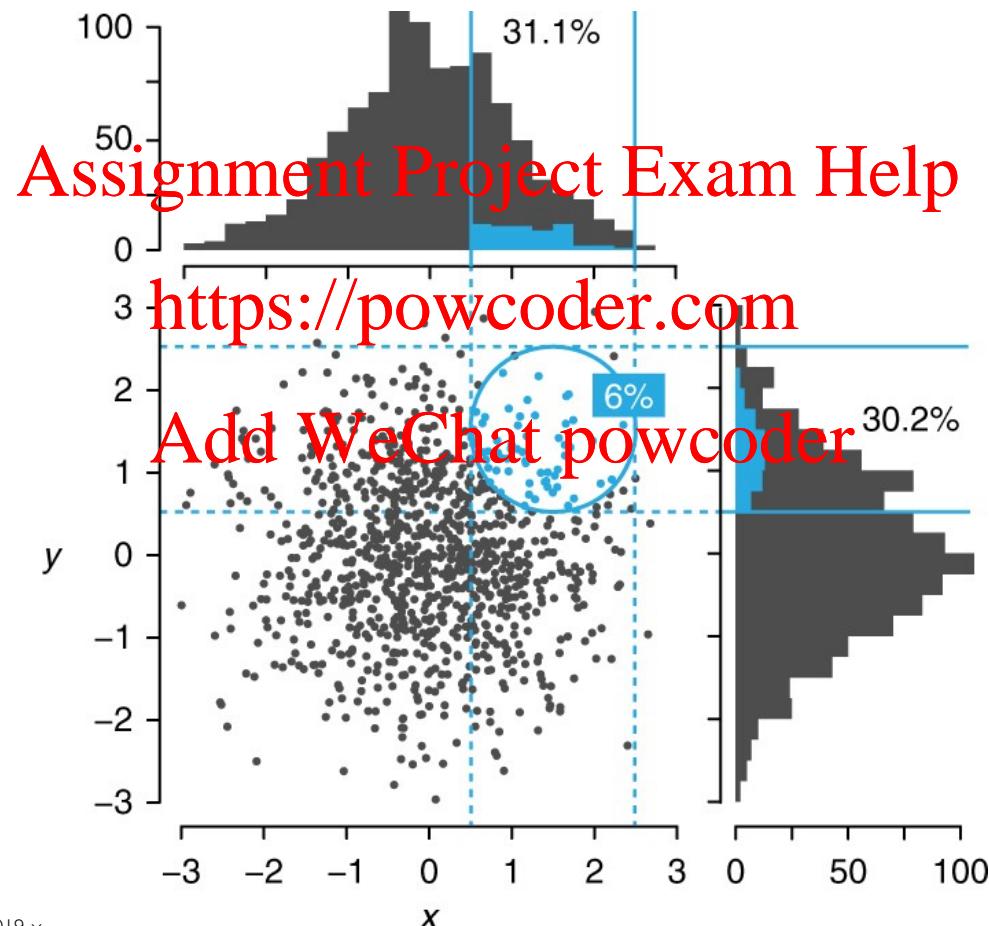
-
- How would I know if my data will be benefitted from a transformation to a higher or lower dimensional space?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

CURSE OF DIMENSIONALITY



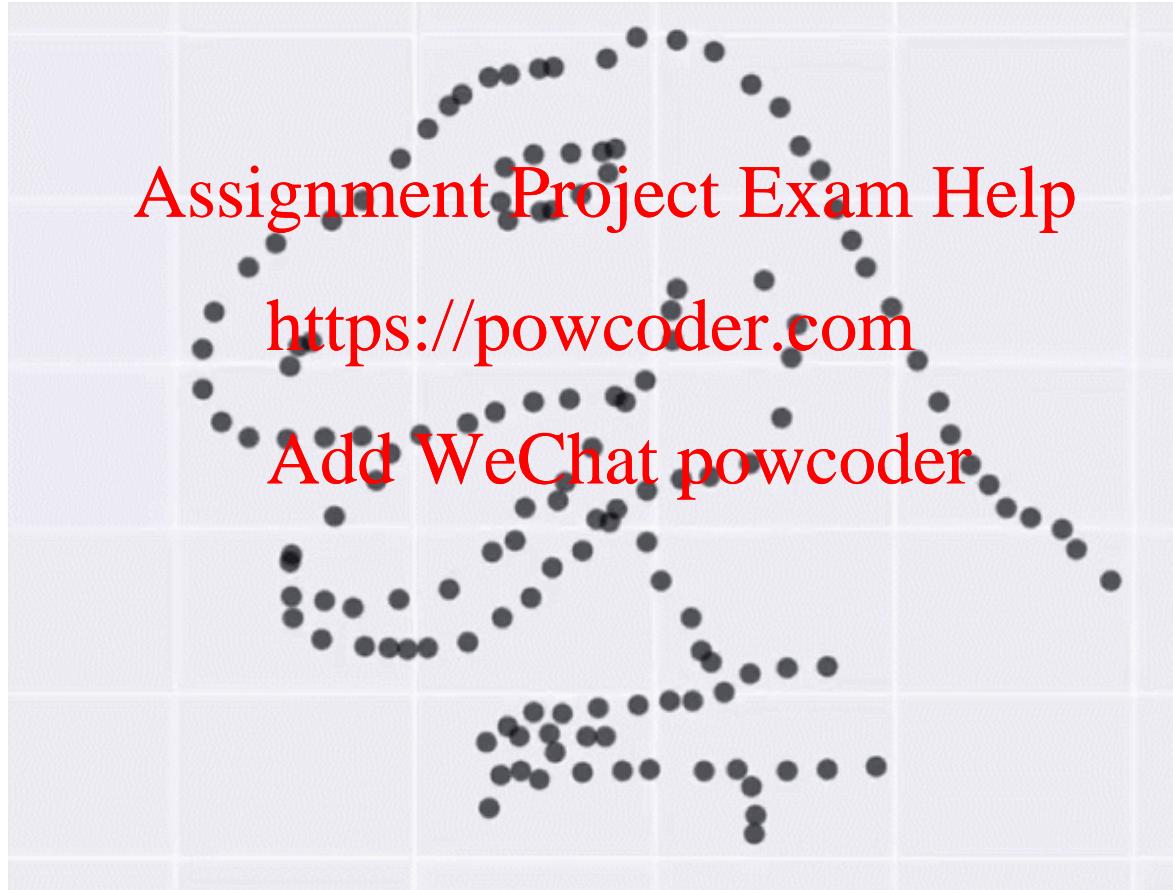
-
- Would I always have to visualize the data at a 2D or 3D level to visually understand if the data can be better separable? (but then this would defeat the idea of going a higher dimensional space which can't be visualized).

Assignment Project Exam Help

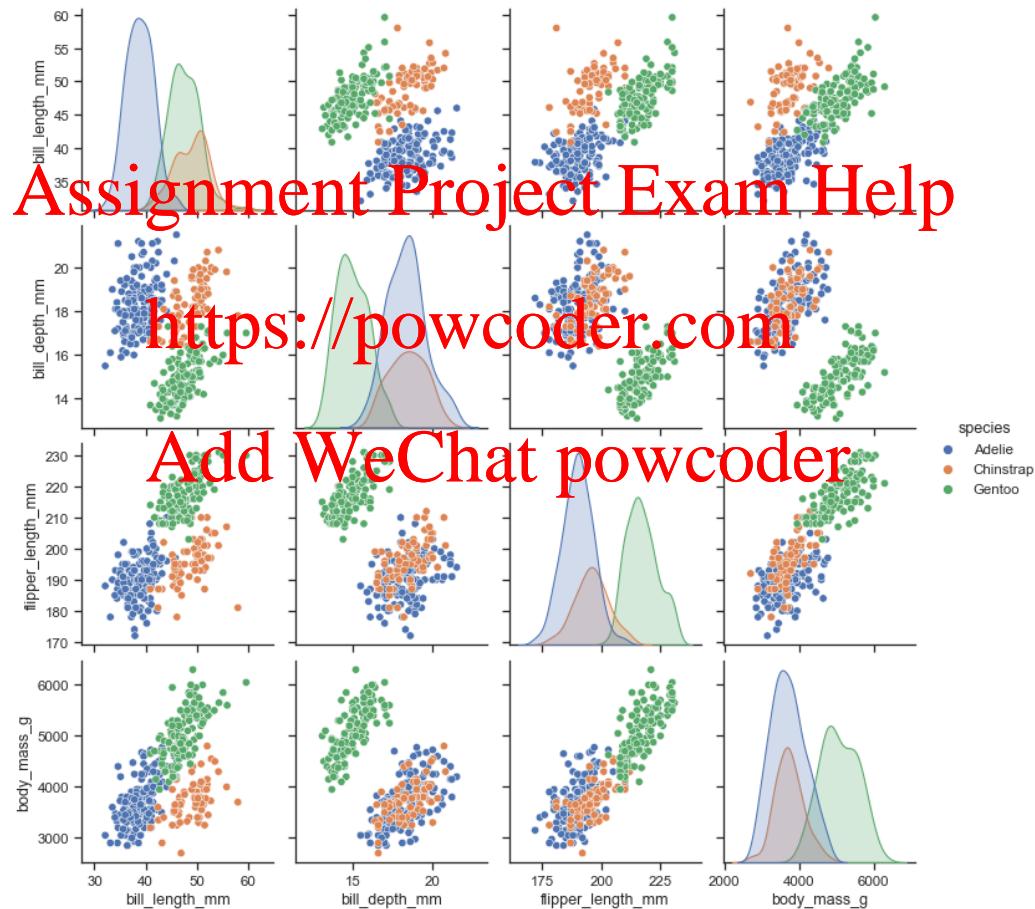
<https://powcoder.com>

Add WeChat powcoder

SUMMARY STATISTICS



SUMMARY STATISTICS



QUESTIONS?

-
- Should I have to go all the way through modelling (e.g. classification) and evaluate a metric such as the Gini coefficient and then go back to comparing different Gini scores from (addition of) extra dimensions?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

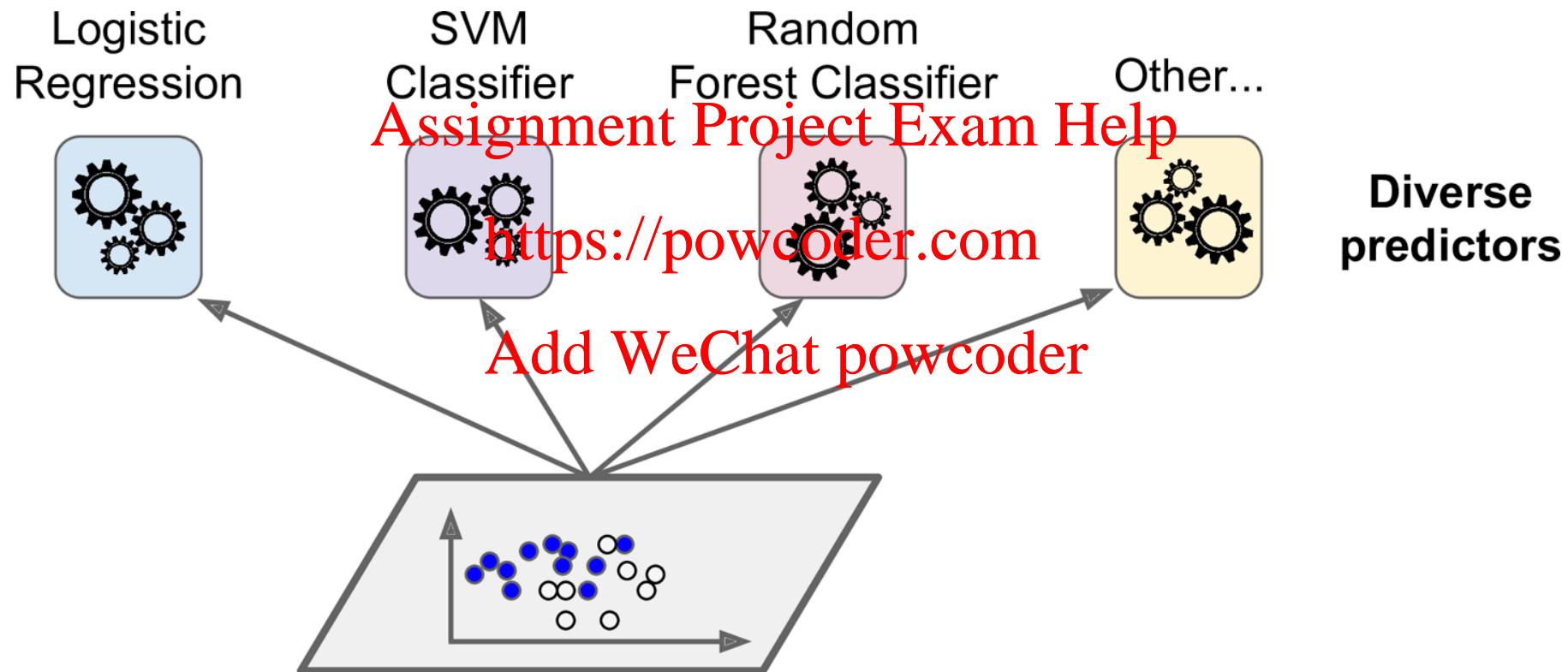
-
- I understand that it might be better to go up a dimension in certain cases and other cases it will be better to go lower a dimension?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

MULTIPLE MODELS



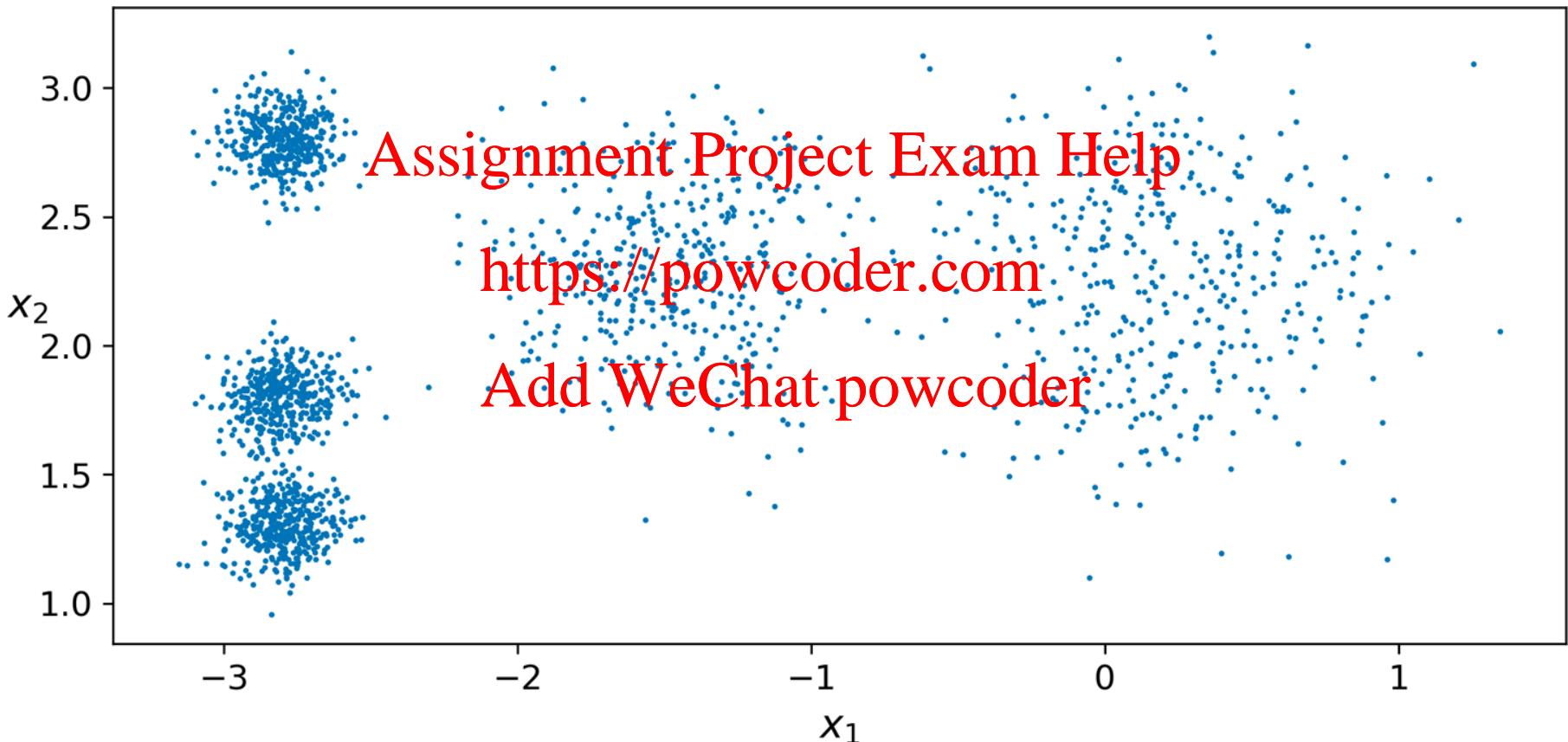
Assignment Project Exam Help

<https://powcoder.com>
Kmeans

Add WeChat powcoder

K-MEANS

LLOYD-FORGY ALGORITHM



— Advantages

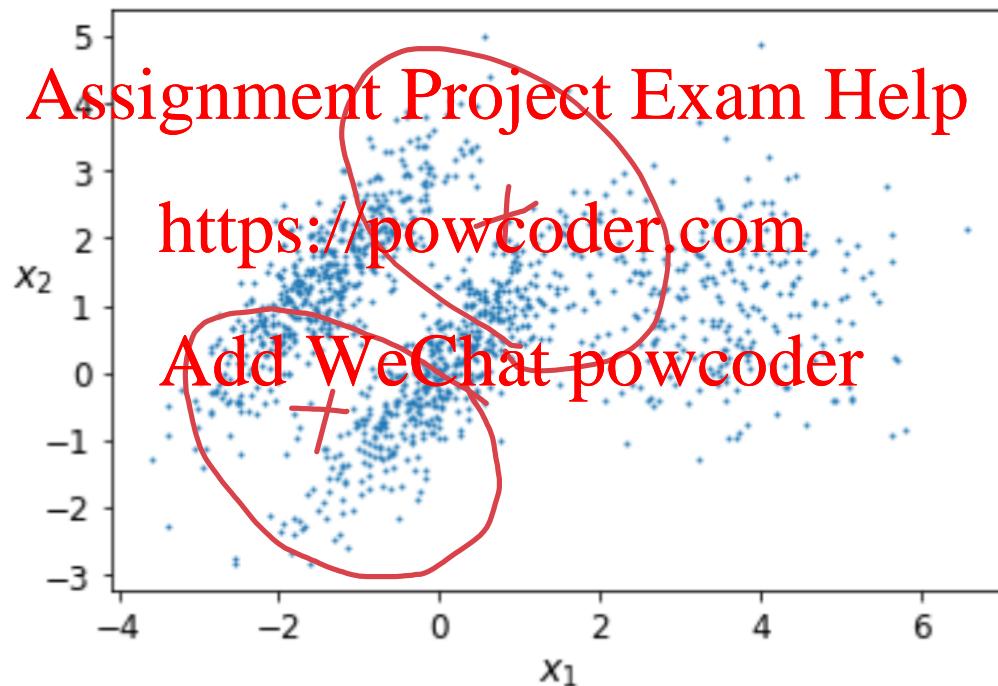
— Disadvantages

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

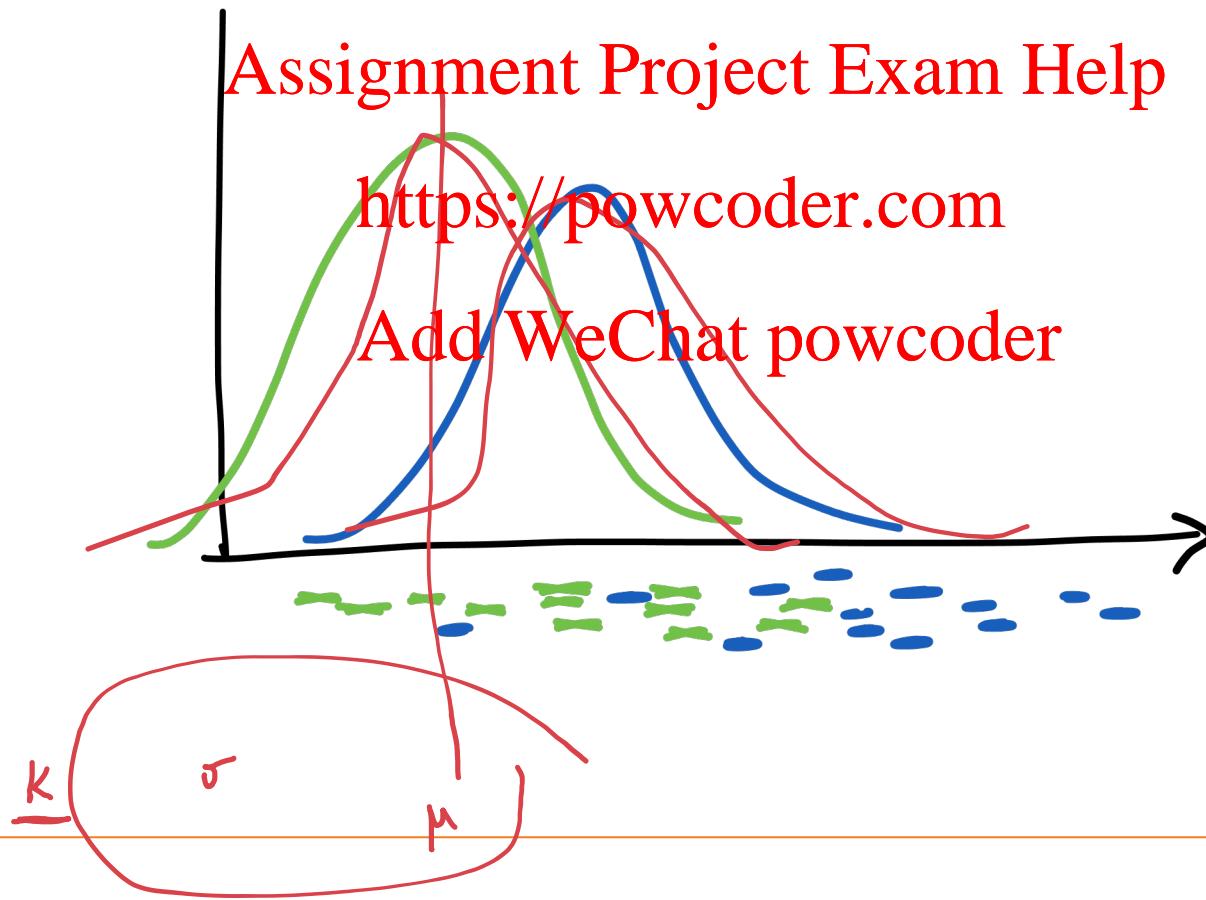
ELLIPSOIDAL DISTRIBUTED DATA



Assignment Project Exam Help

<https://powcoder.com>
Gaussian mixtures

Add WeChat powcoder

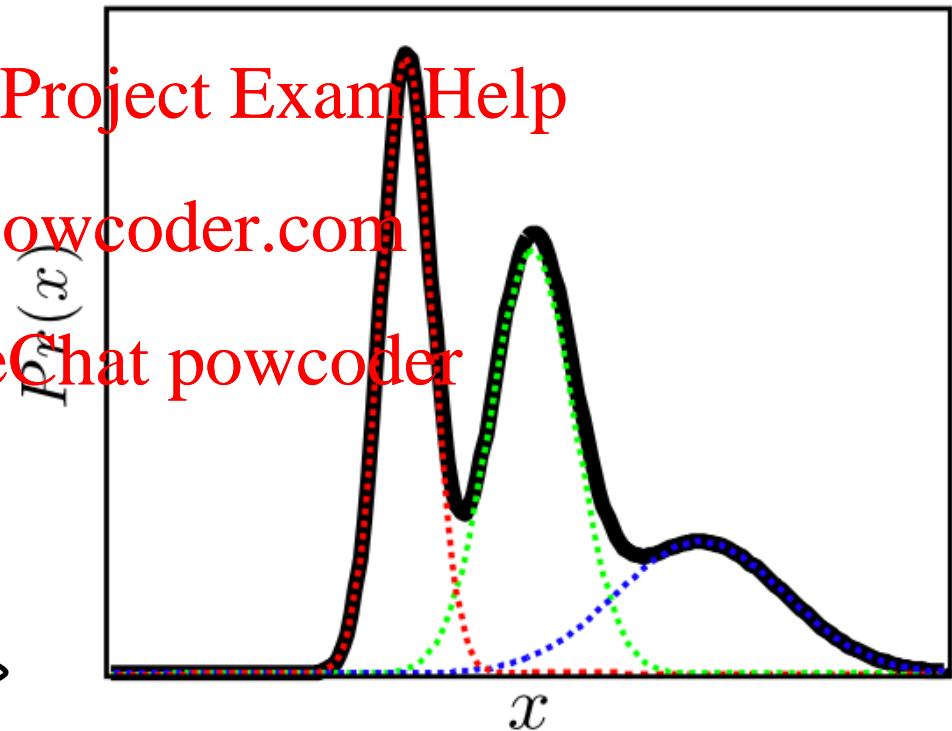
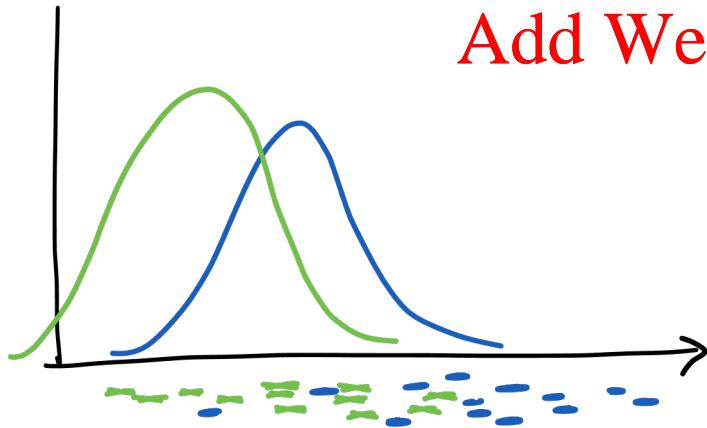


MIXTURE OF GAUSSIANS (1D)

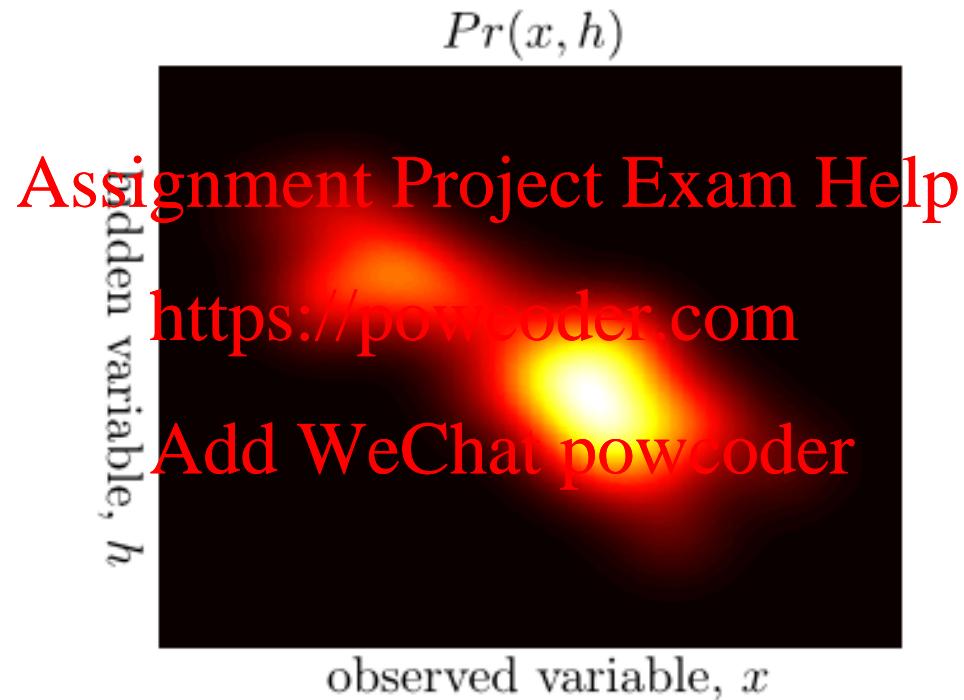
Assignment Project Exam Help

<https://powcoder.com>

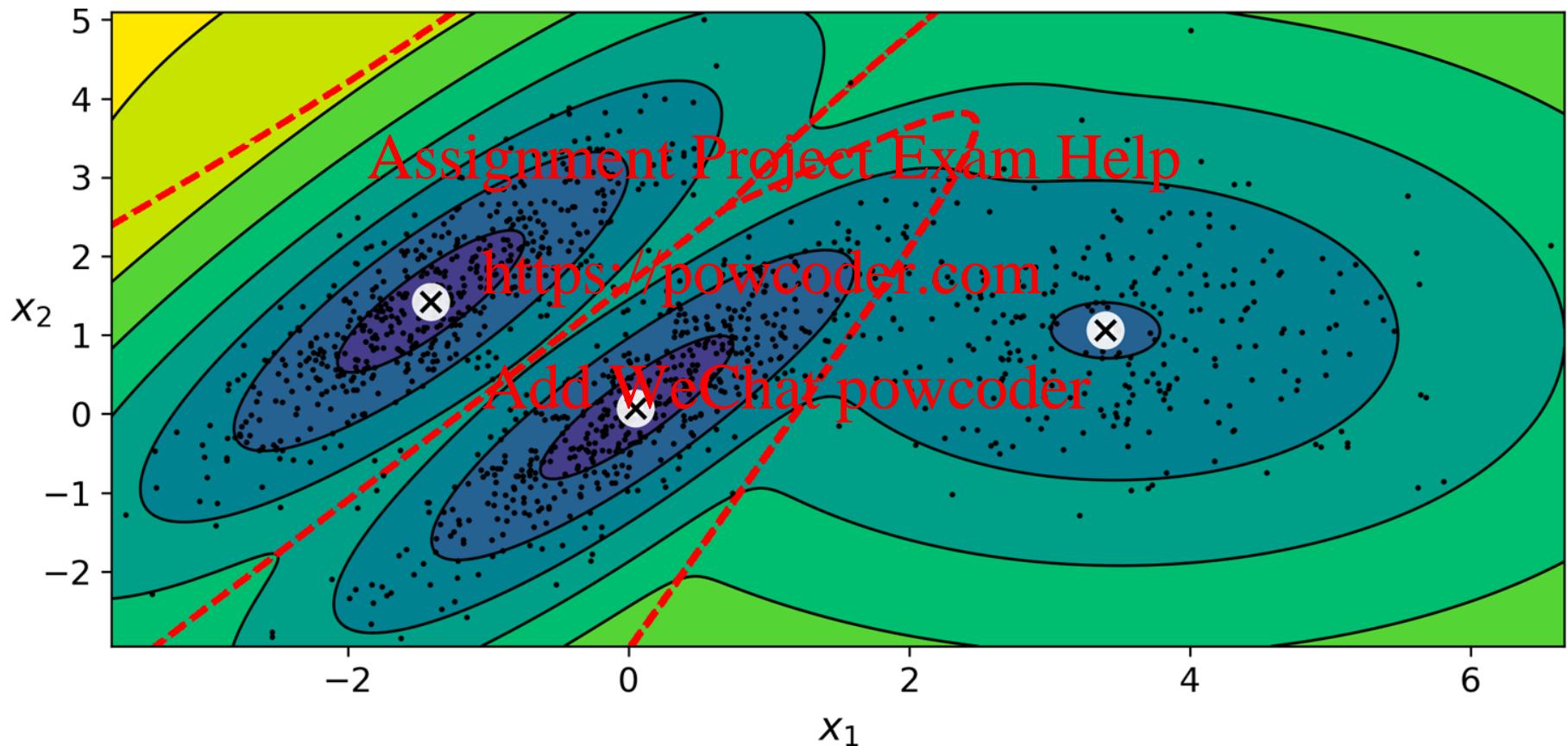
Add WeChat powcoder



HIDDEN (LATENT) VARIABLES

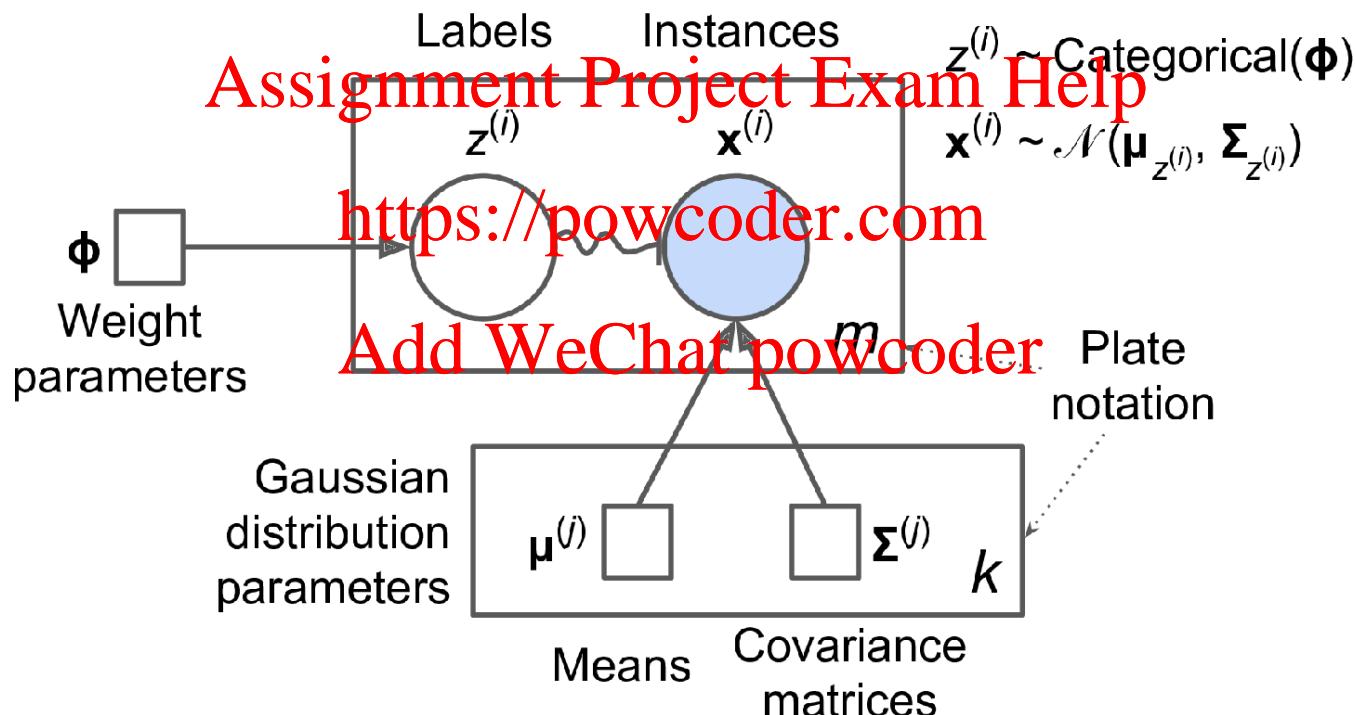


MIXTURE OF GAUSSIANS (2D)



GRAPHICAL MODELS

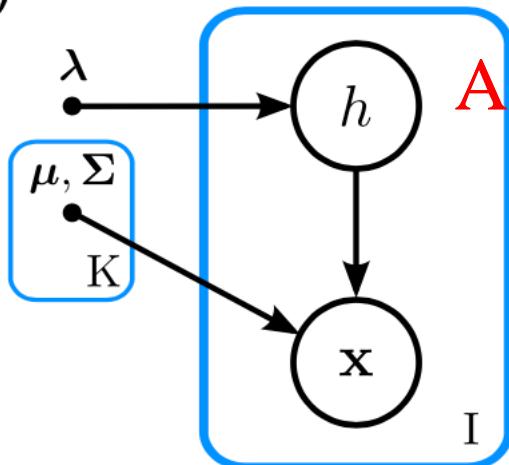
GAUSSIAN MIXTURES



- including its parameters (squares, solid circles, bullet)
- random variables (circles)
- conditional dependencies (solid arrows)

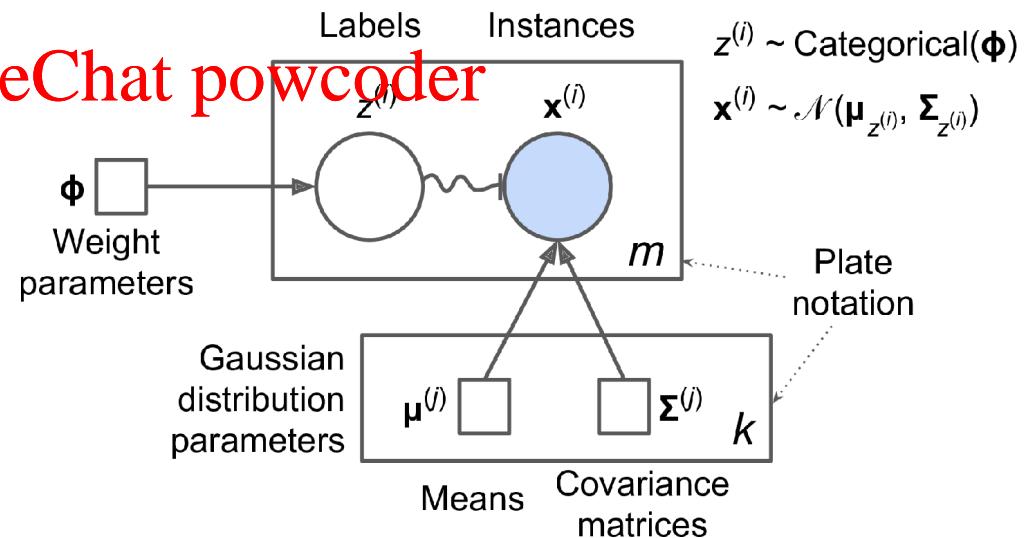
Assignment Project Exam Help

a)



<https://powcoder.com>

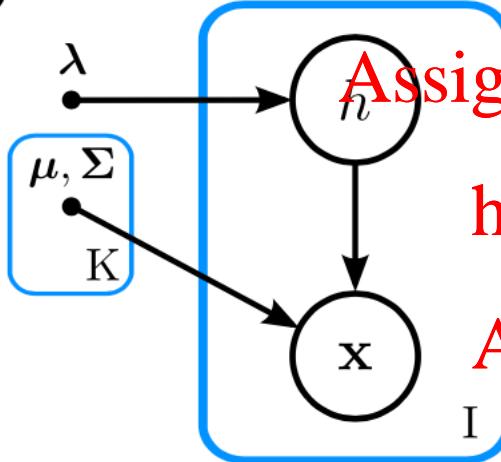
Add WeChat powcoder



FAMILIES OF MODELS

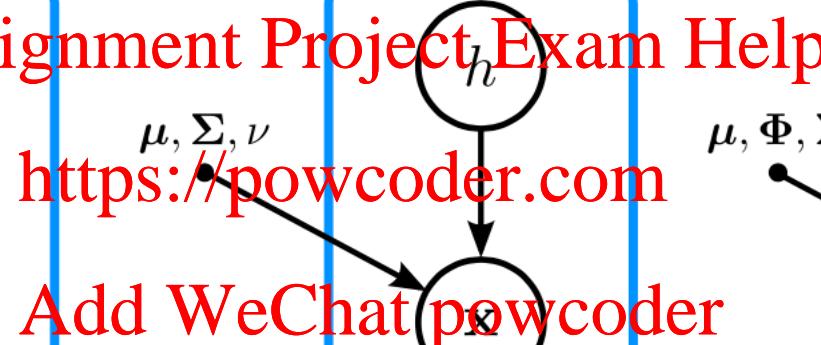
Gaussian mixture

a)



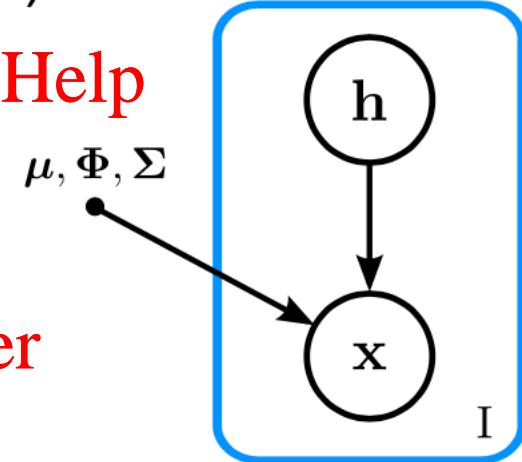
T-distribution mixture

b)

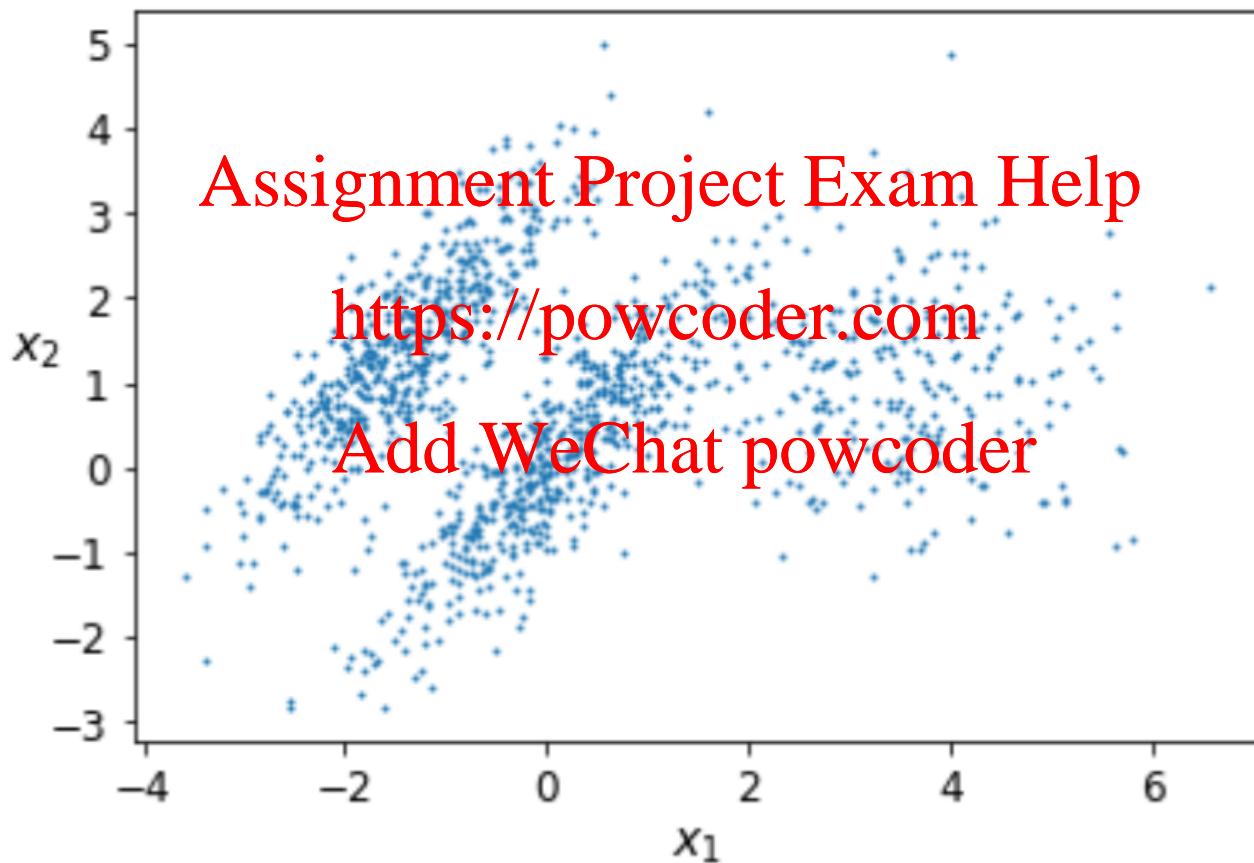


Factor Analysis

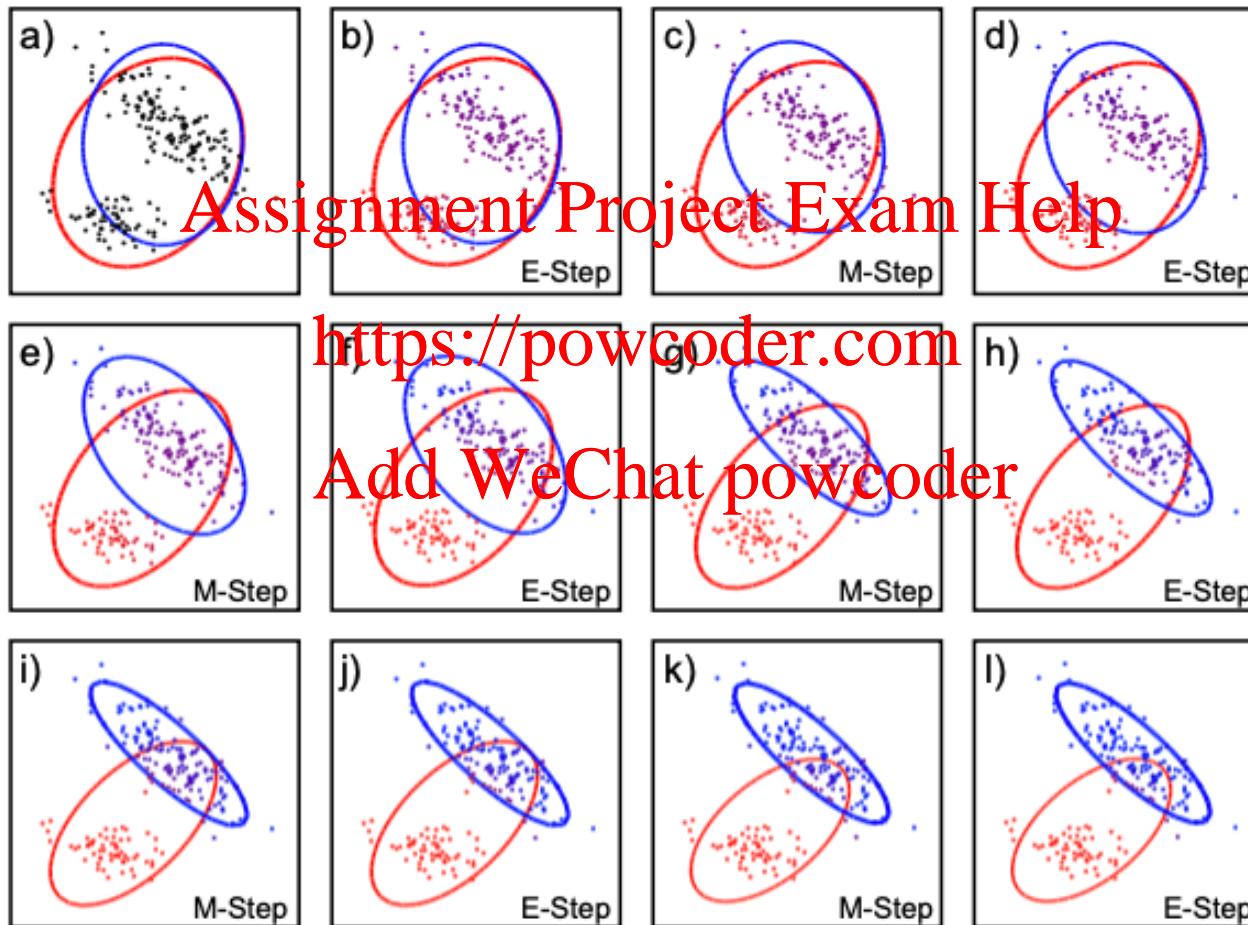
c)



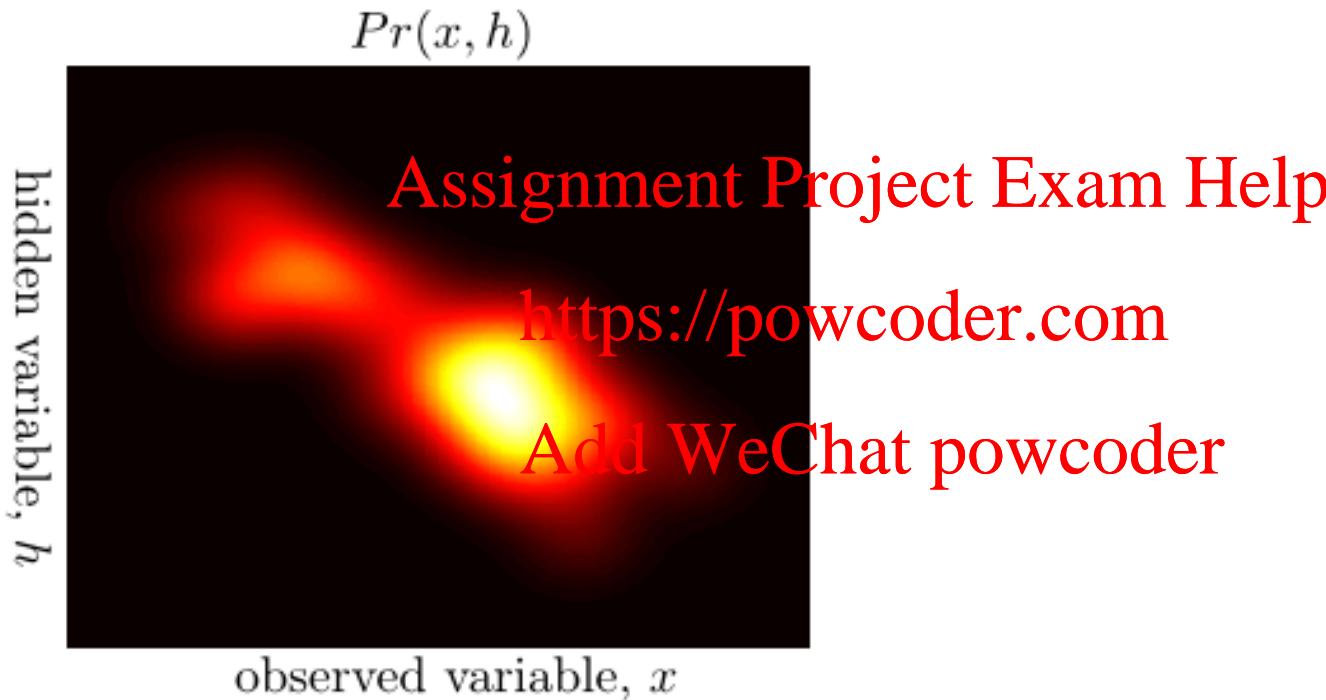
TWO STEP – EM ALGORITHM



EM ALGORITHM

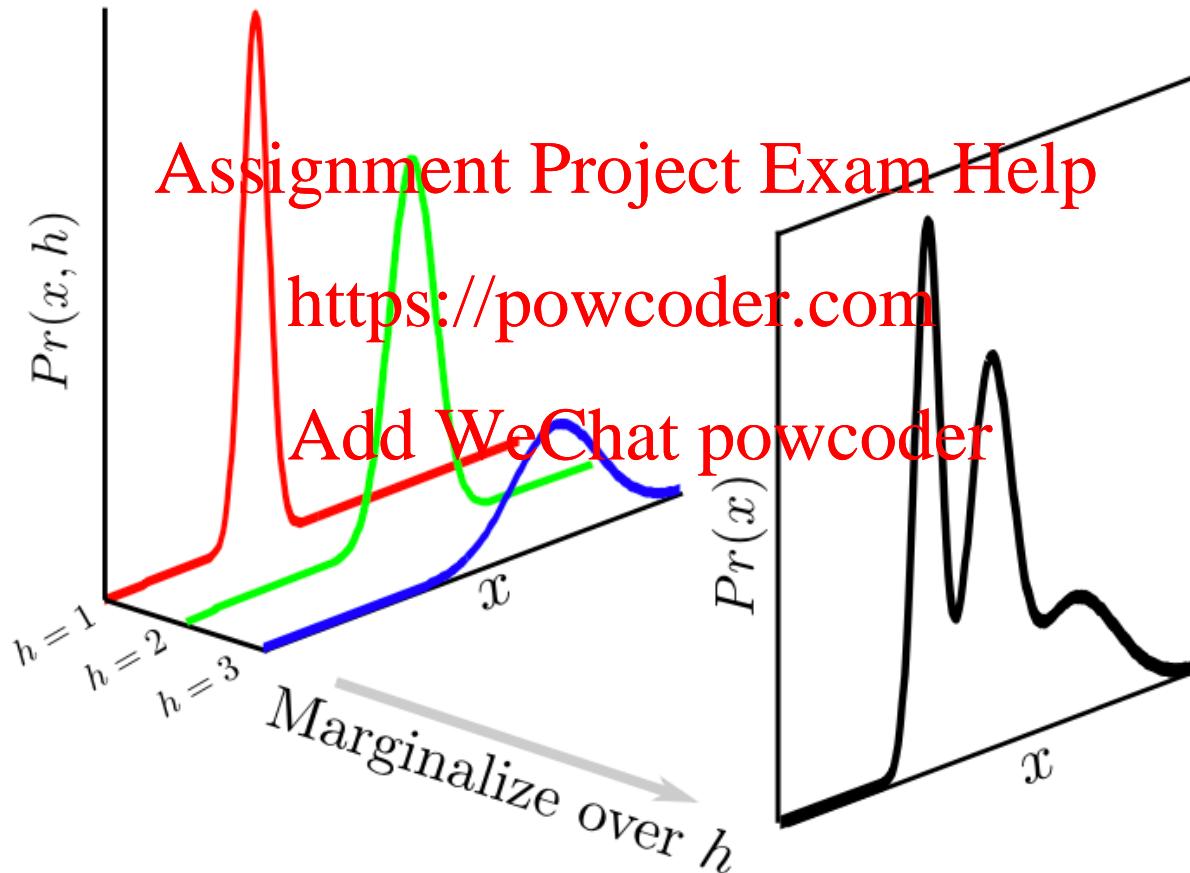


EXPECTATION MAXIMIZATION

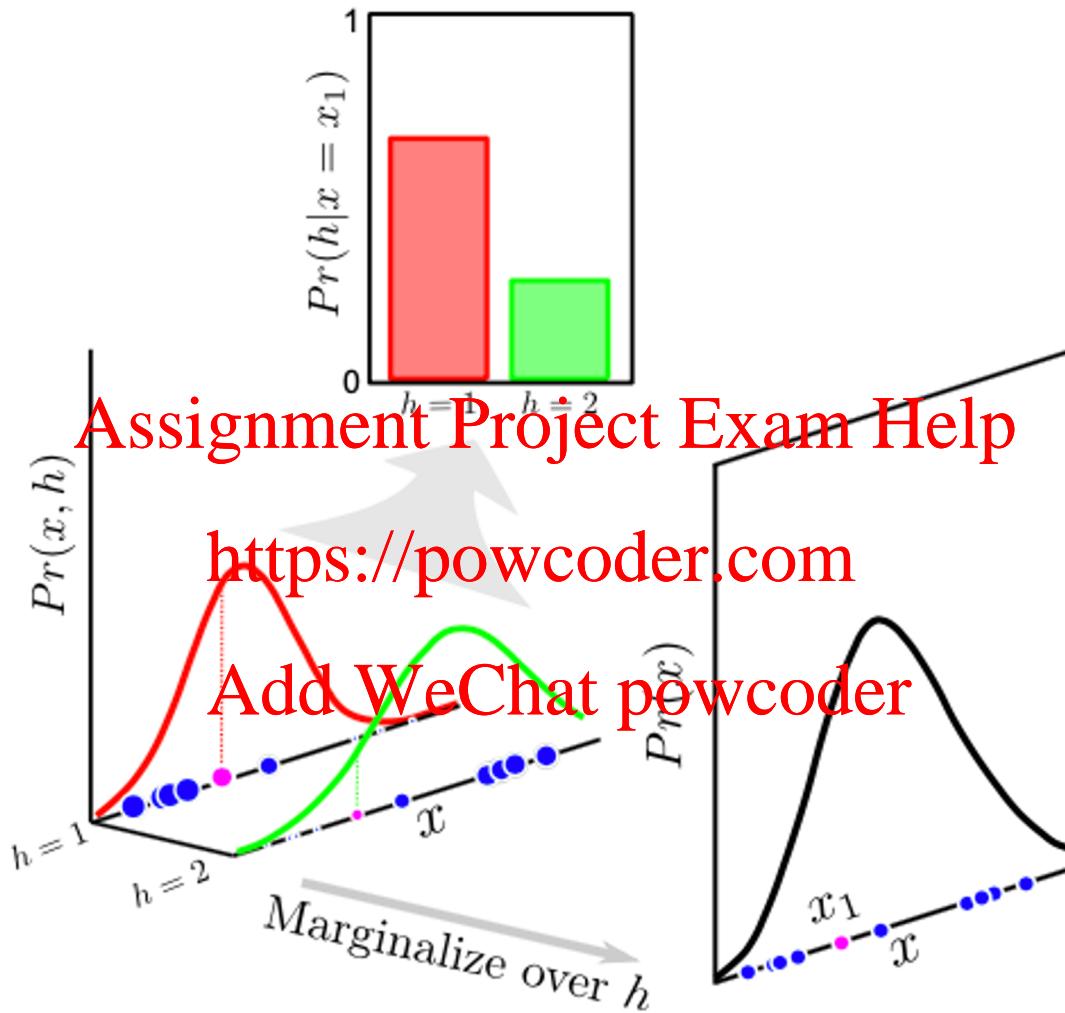


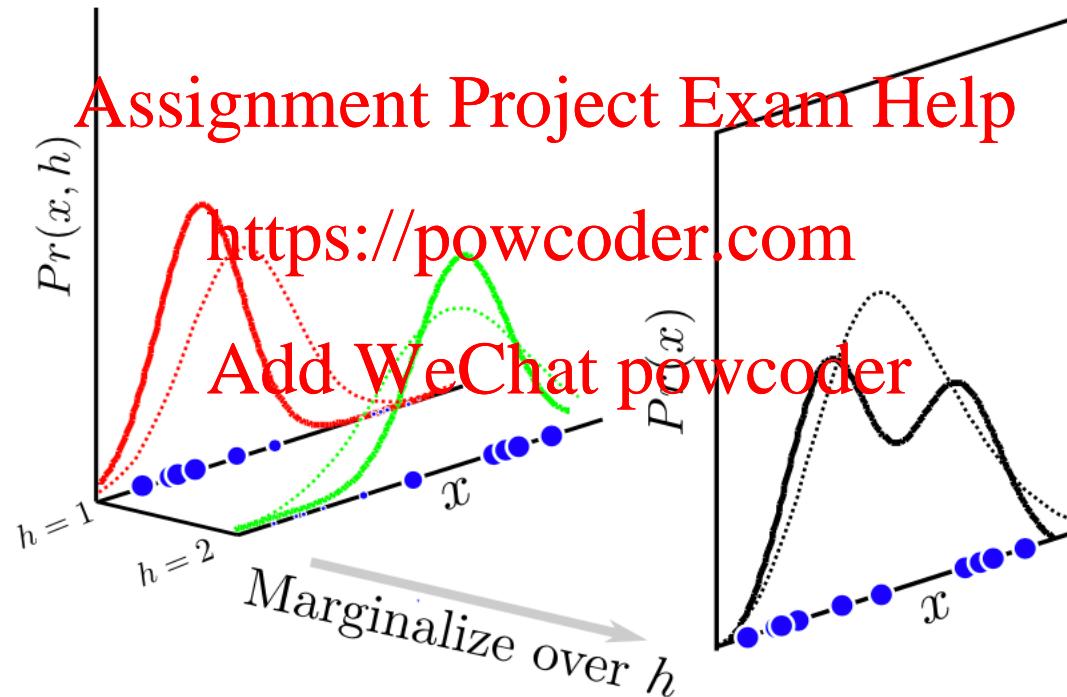
$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} \left[\sum_{i=1}^I \log \left[\int Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) d\mathbf{h}_i \right] \right]$$

MIXTURE OF GAUSSIANS AS MARGINALIZATION

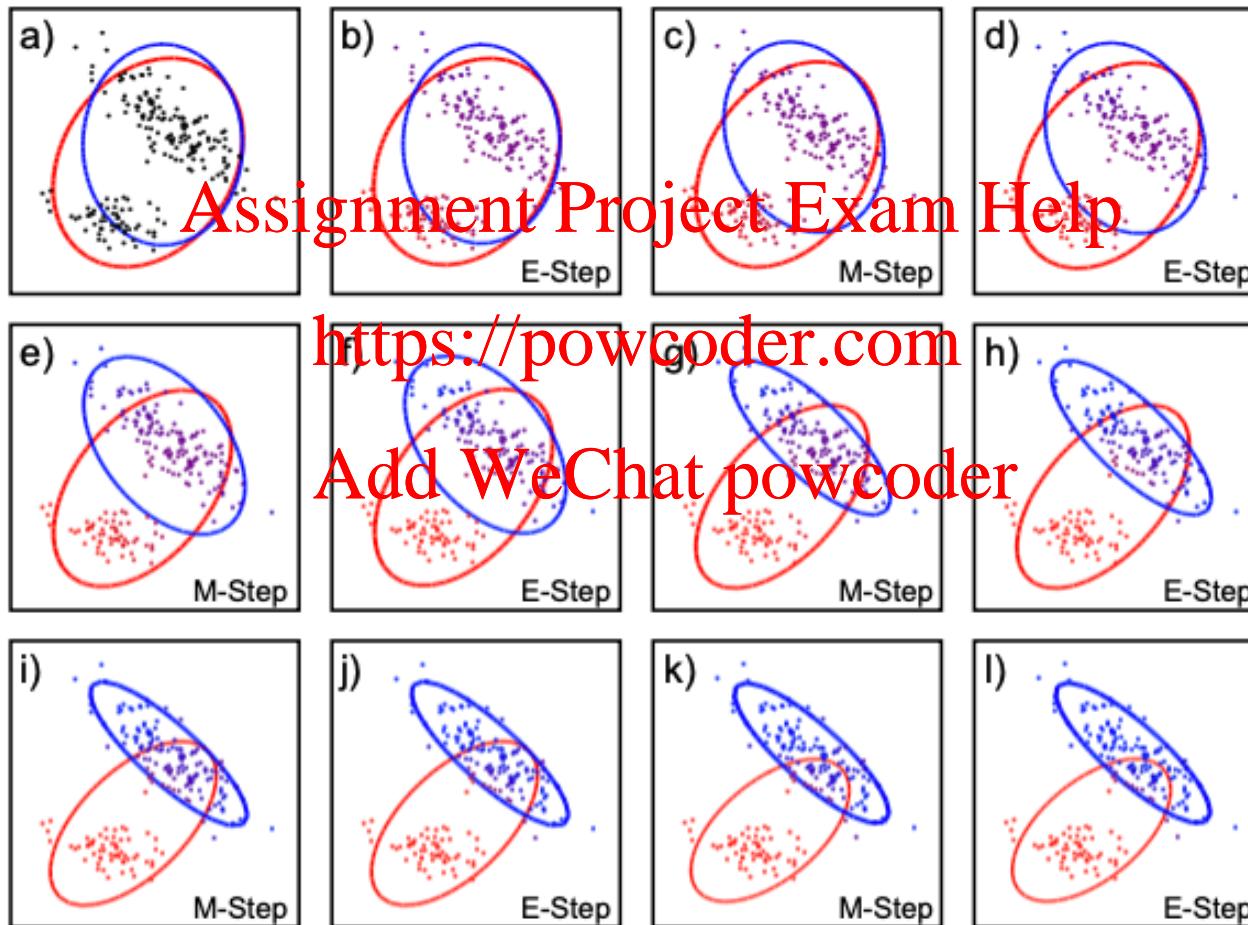


E-STEP





EM ALGORITHM



EXPECTATION MAXIMIZATION

$$p(\mathbf{z}|\mathbf{X}) = \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} = \frac{p(\mathbf{X}|\mathbf{z}) p(\mathbf{z})}{p(\mathbf{X})}$$

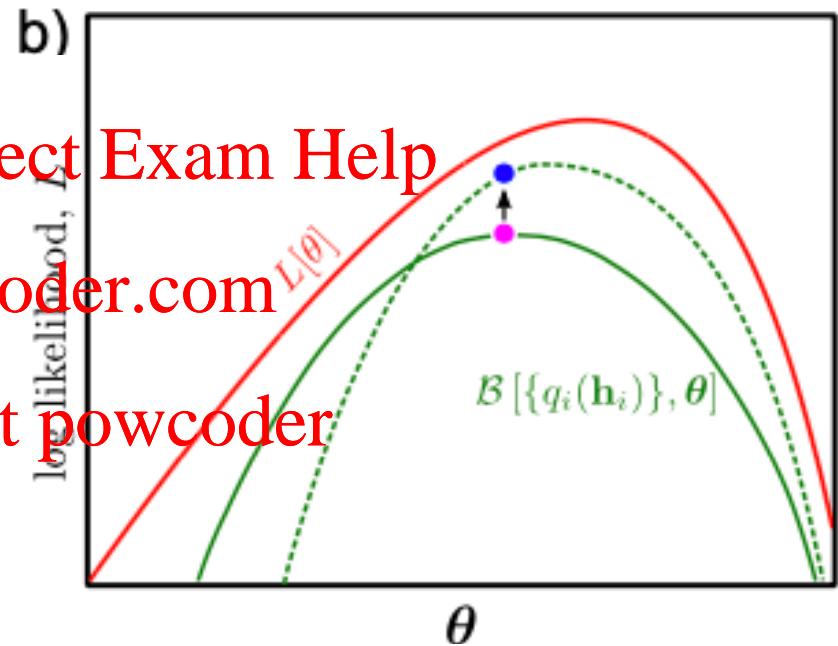
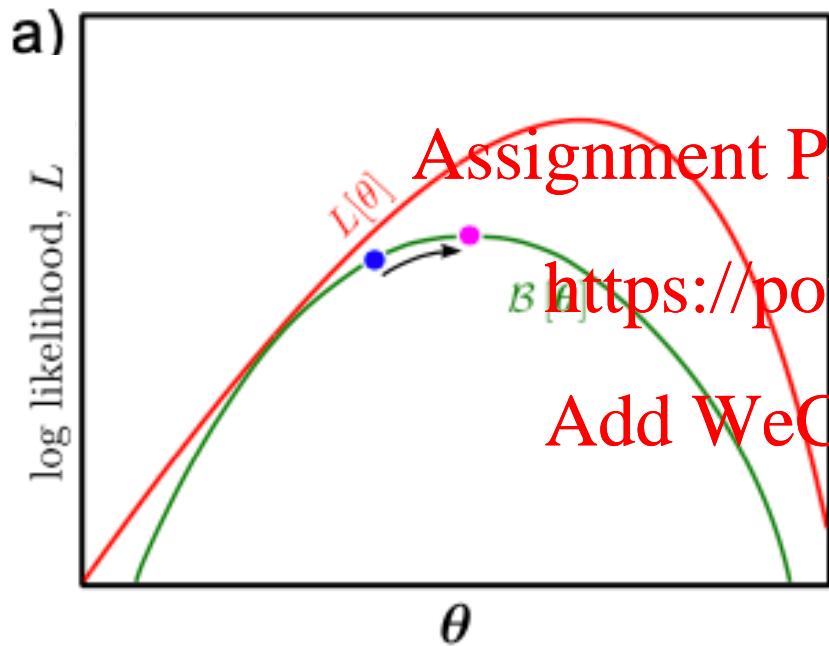
Assignment Project Exam Help

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\sum_{i=1}^I \log \left[\int Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) d\mathbf{h}_i \right] \right]$$

Add WeChat powcoder

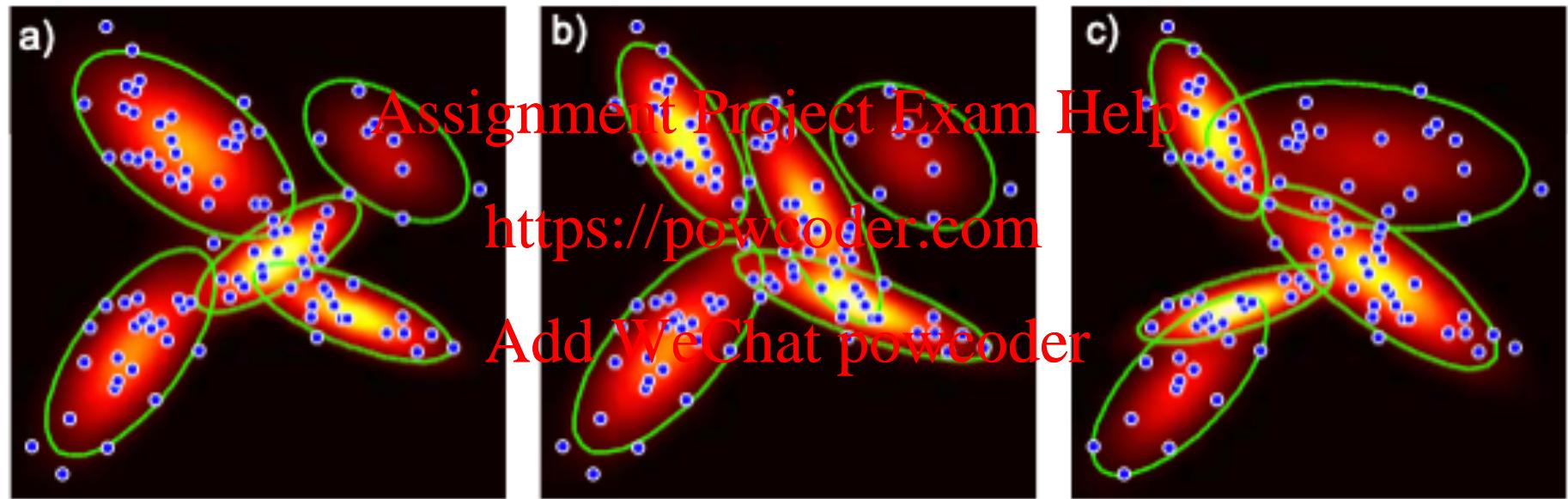
$$\begin{aligned} \mathcal{B} [\{q_i(\mathbf{h}_i)\}, \boldsymbol{\theta}] &= \sum_{i=1}^I \int q_i(\mathbf{h}_i) \log \left[\frac{Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \\ &\leq \sum_{i=1}^I \log \left[\int Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) d\mathbf{h}_i \right]. \end{aligned}$$

MANIPULATING THE LOWER BOUND



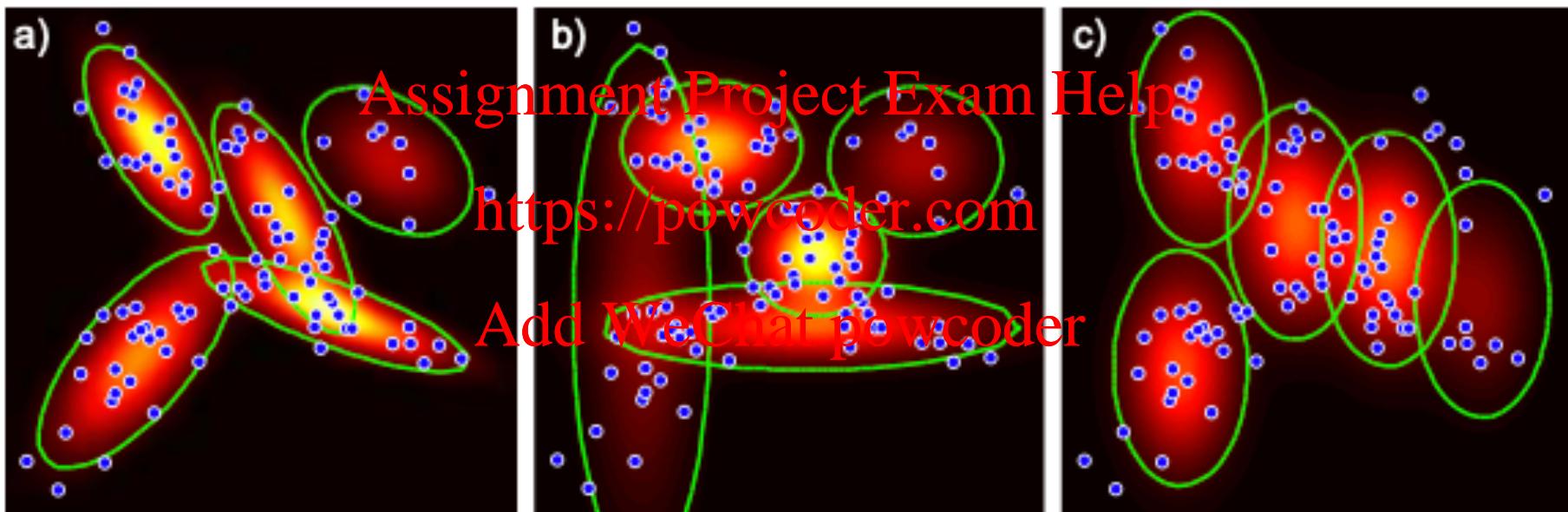
$$\begin{aligned} \mathcal{B} [\{q_i(\mathbf{h}_i)\}, \theta] &= \sum_{i=1}^I \int q_i(\mathbf{h}_i) \log \left[\frac{Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \\ &\leq \sum_{i=1}^I \log \left[\int Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) d\mathbf{h}_i \right]. \end{aligned}$$

LOCAL MAXIMA



Repeated fitting of mixture of Gaussians model with different starting points results in different models as the fit converges to different local maxima.
Log likelihoods are a) 98.76 b) 96.97 c) 94.35, respectively, indicating that (a) is the best fit.

COVARIANCE COMPONENTS



- a) Full covariances.
- b) Diagonal covariances.
- c) Identical diagonal covariances.

LEARNING GMM

PSEUDO CODE

```
Input : Training data  $\{\mathbf{x}_i\}_{i=1}^I$ , number of clusters  $K$ 
Output: ML estimates of parameters  $\boldsymbol{\theta} = \{\lambda_{1\dots K}, \boldsymbol{\mu}_{1\dots K}, \boldsymbol{\Sigma}_{1\dots K}\}$ 
```

```
begin
```

```
    Initialize  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  a
```

```
repeat
```

```
    // Expectation Step
```

```
    for  $i=1$  to  $I$  do
```

```
        for  $k=1$  to  $K$  do
```

```
            |  $l_{ik} = \lambda_k \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]$  // numerator of Bayes' rule
```

```
        end
```

```
        // Compute posterior (responsibilities) by normalizing
```

```
        for  $k=1$  to  $K$  do
```

```
            |  $r_{ik} = l_{ik}/(\sum_{k=1}^K l_{ik})$ 
```

```
        end
```

```
    end
```

```
    // Maximization Step b
```

```
    for  $k=1$  to  $K$  do
```

```
        |  $\lambda_k^{[t+1]} = (\sum_{i=1}^I r_{ik})/(\sum_{k=1}^K \sum_{i=1}^I r_{ik})$ 
```

```
        |  $\boldsymbol{\mu}_k^{[t+1]} = (\sum_{i=1}^I r_{ik} \mathbf{x}_i)/(\sum_{i=1}^I r_{ik})$ 
```

```
        |  $\boldsymbol{\Sigma}_k^{[t+1]} = (\sum_{i=1}^I r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]}) (\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})^T)/(\sum_{i=1}^I r_{ik})$ .
```

```
    end
```

```
    // Compute Data Log Likelihood and EM Bound
```

```
     $L = \sum_{i=1}^I \log \left[ \sum_{k=1}^K \lambda_k \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k] \right]$ 
```

```
     $B = \sum_{i=1}^I \sum_{k=1}^K r_{ik} \log [\lambda_k \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]/r_{ik}]$ 
```

```
    until No further improvement in  $L$ 
```

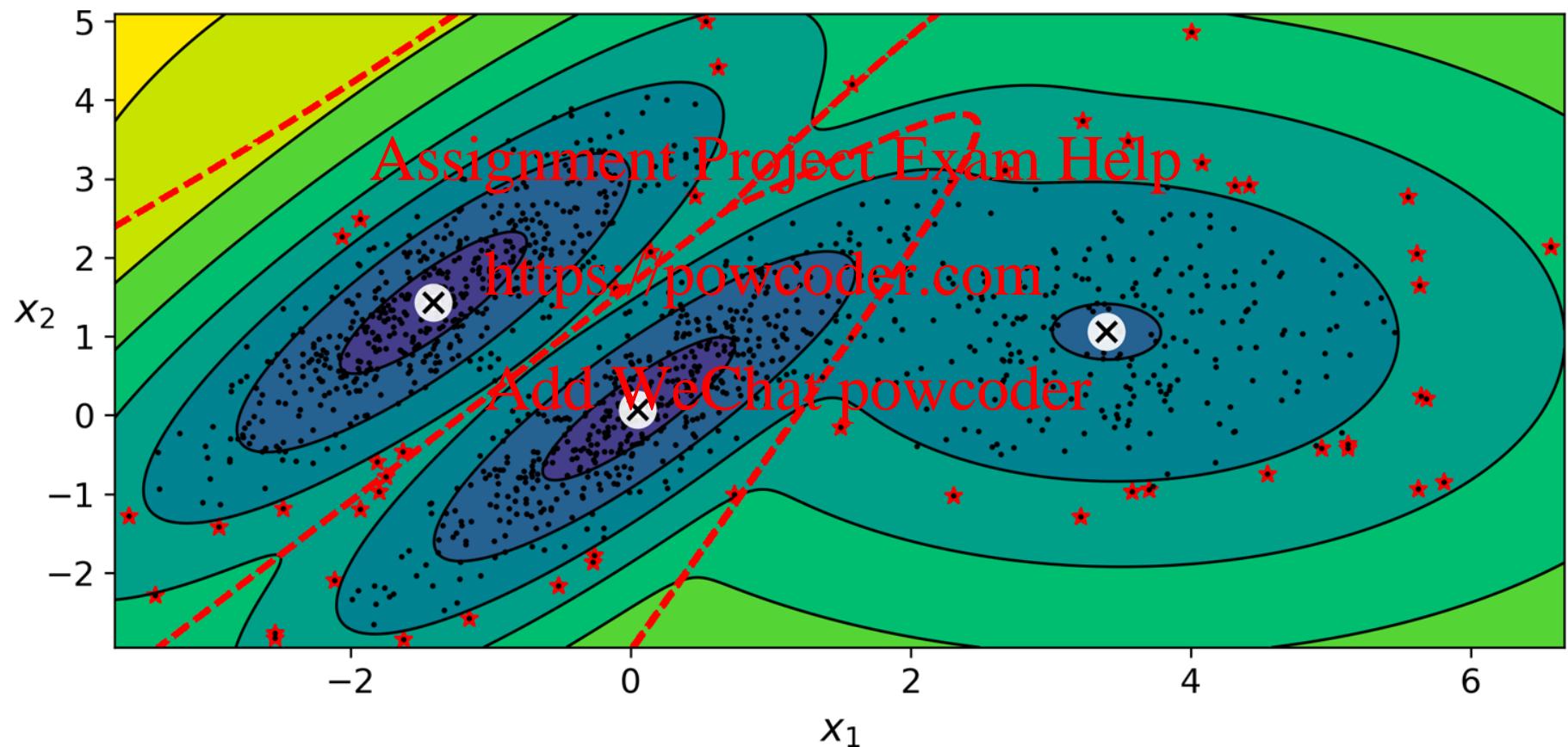
```
end
```

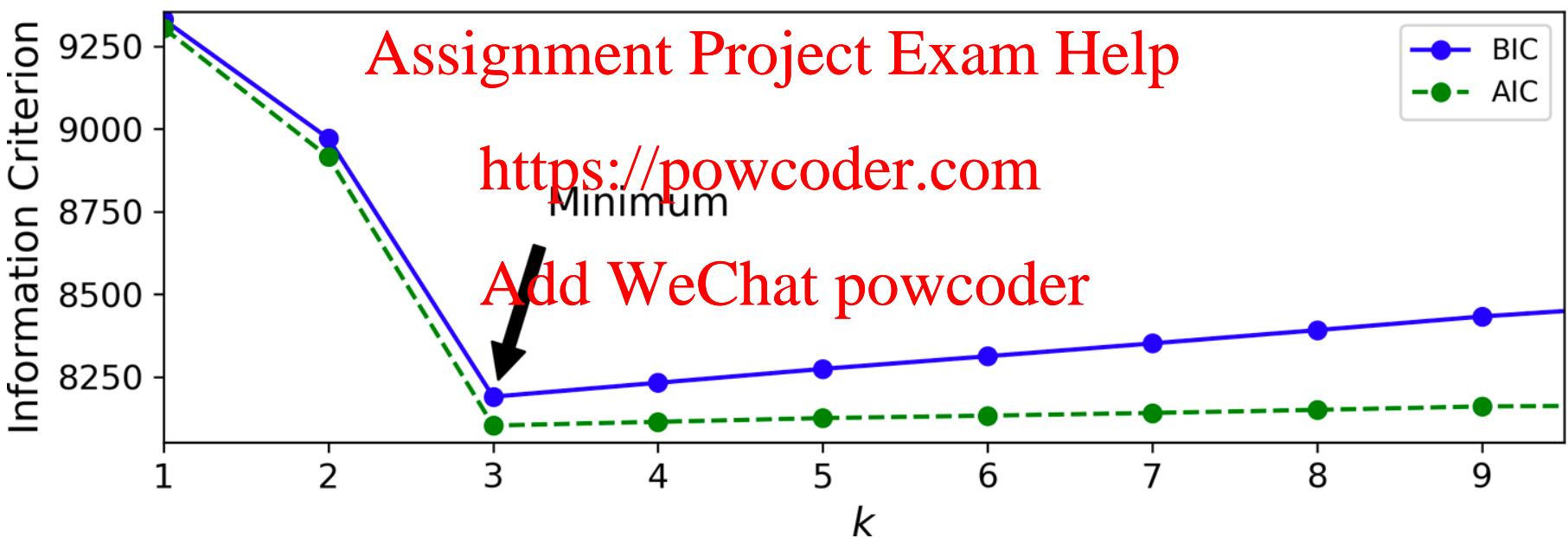
Assignment Project Exam Help

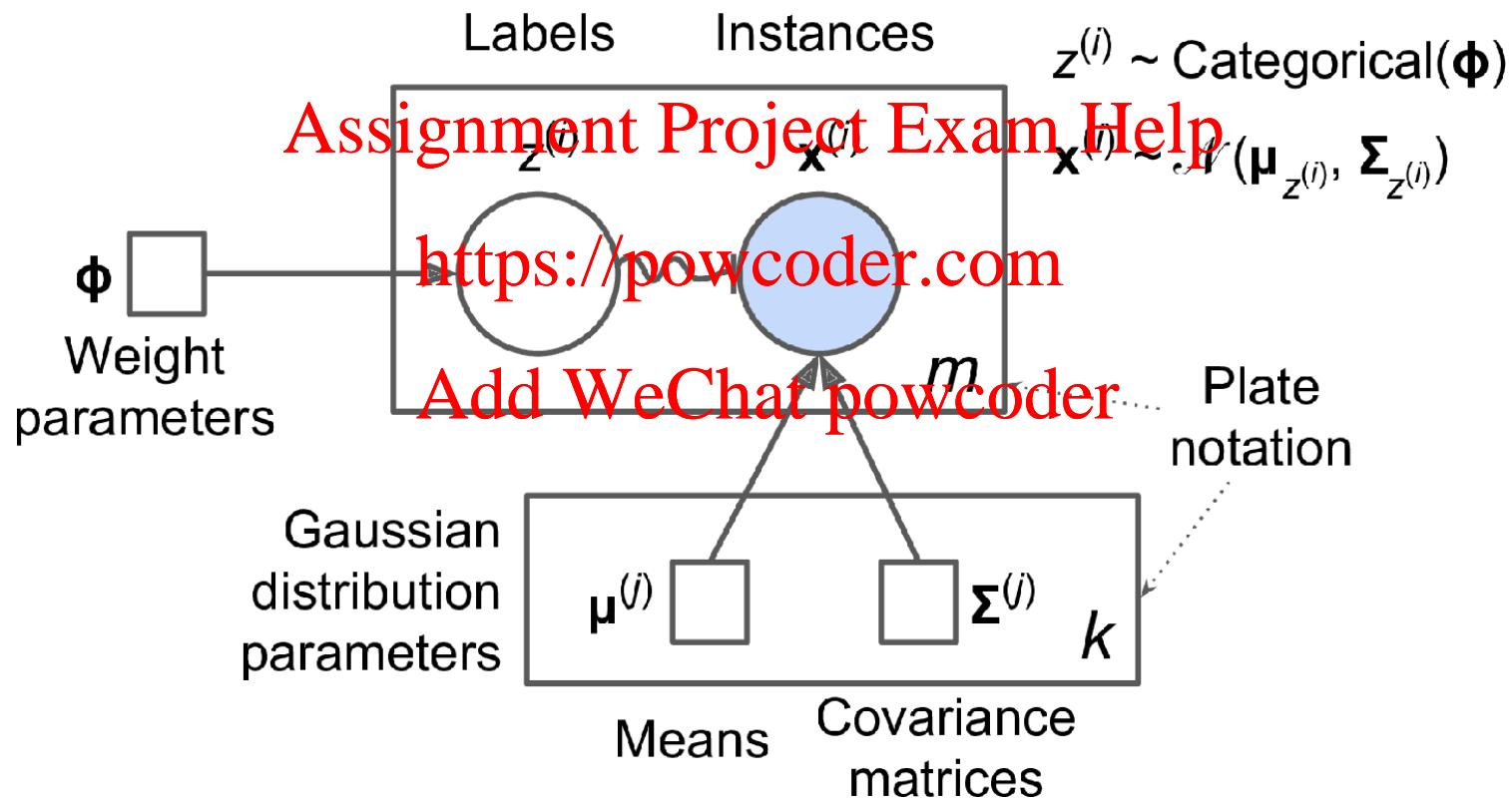
<https://powcoder.com>

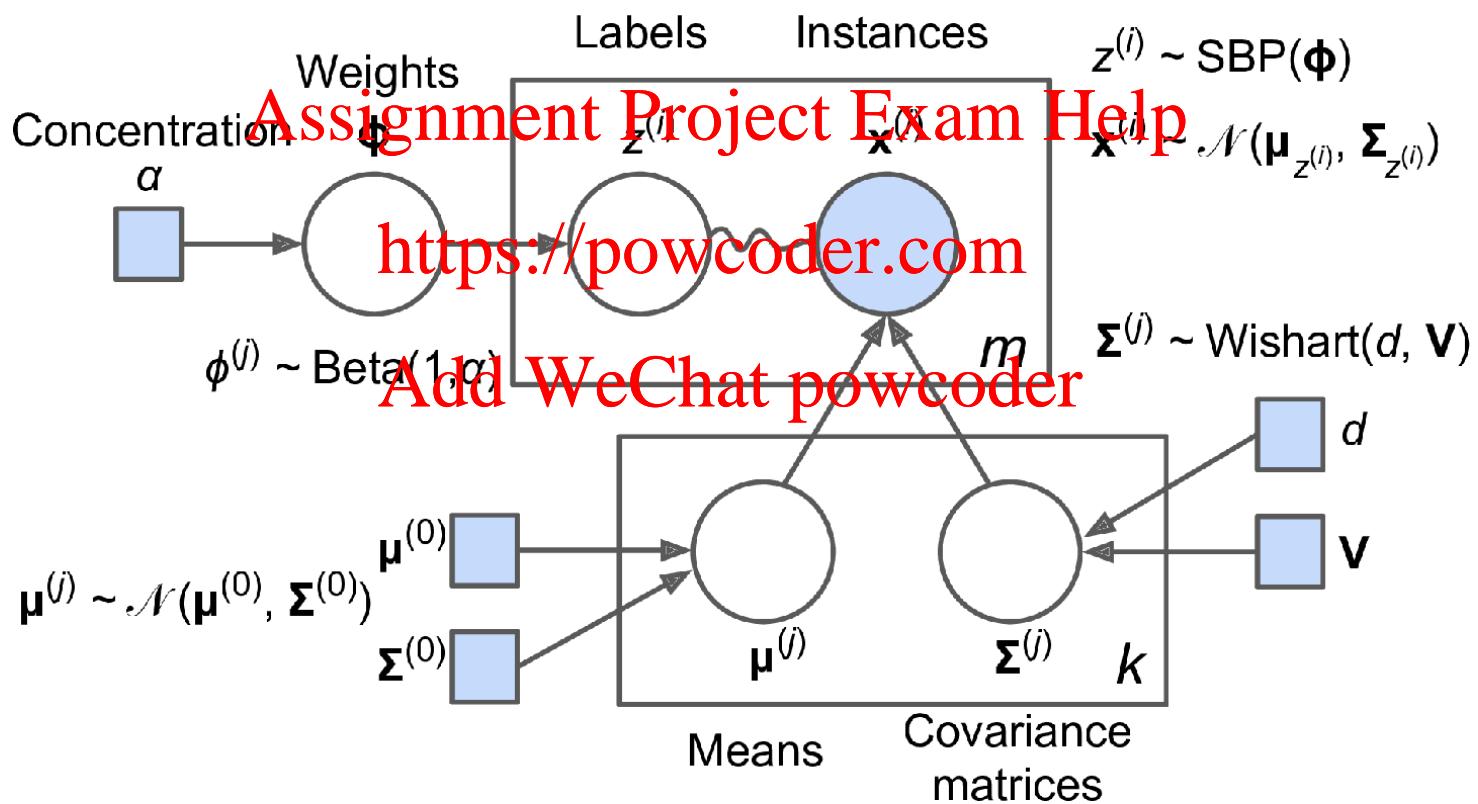
Add WeChat powcoder

ANOMALY DETECTION

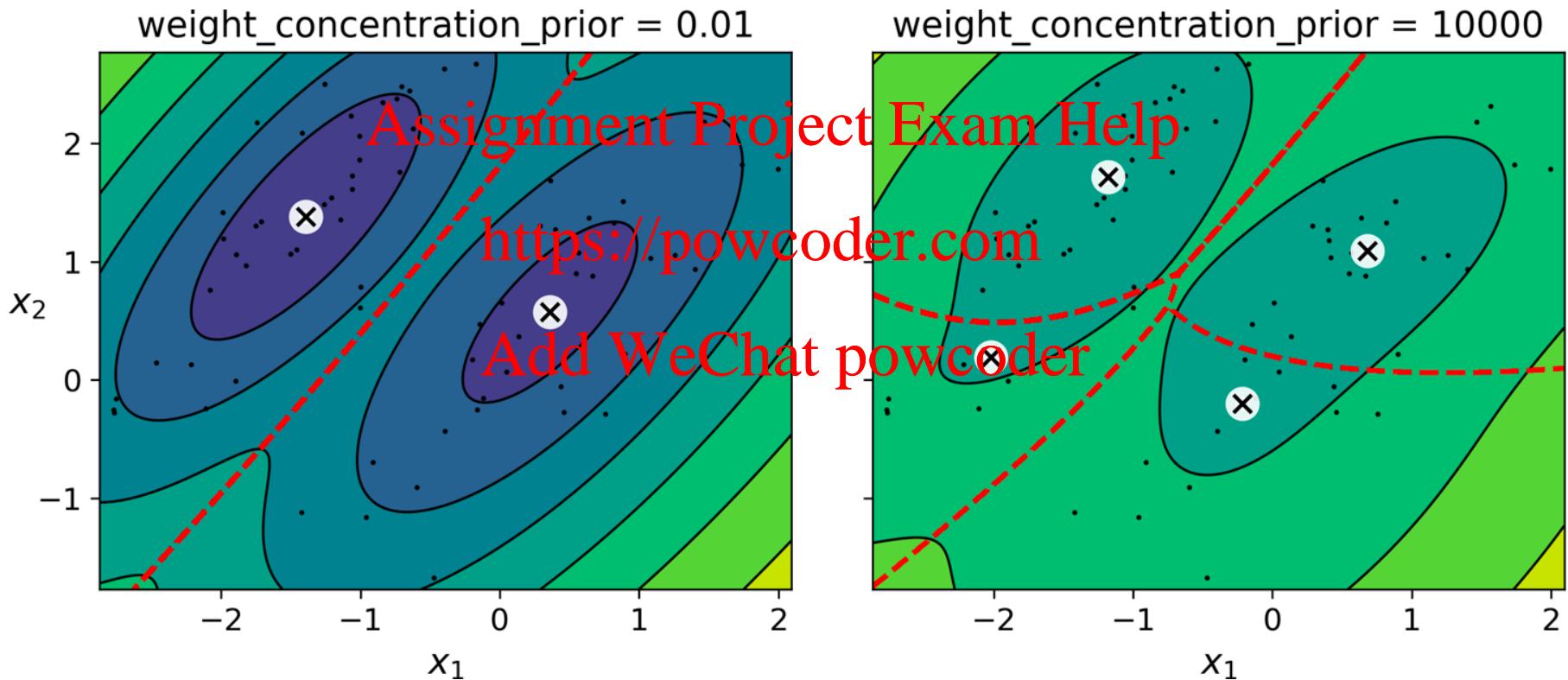






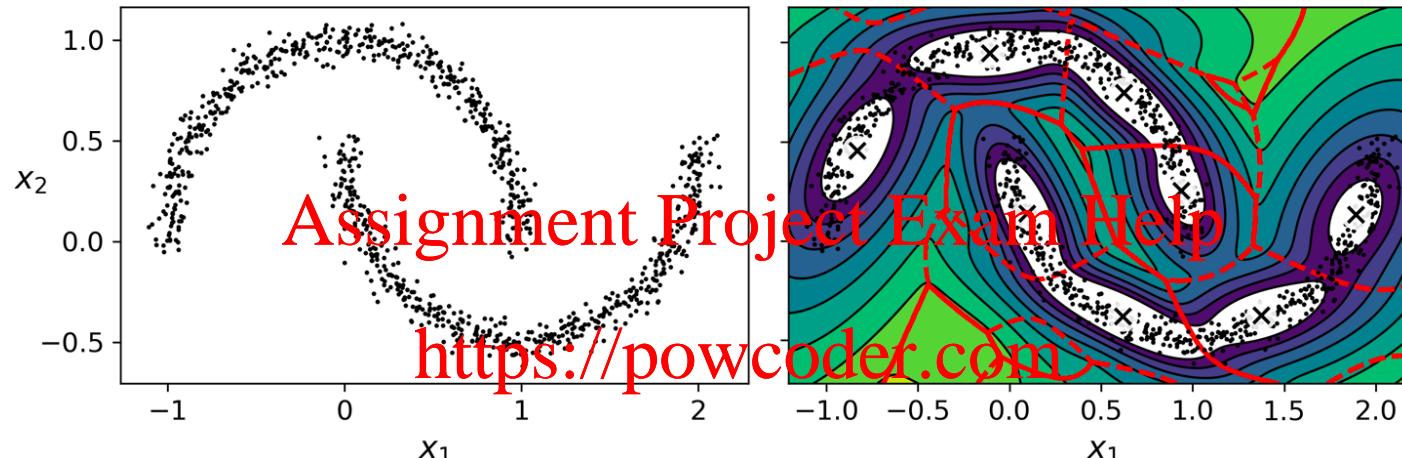


CONCENTRATION PRIORS



The more data we have, however, the less the priors matter. In fact, to plot diagrams with such large differences, you must use very strong priors and little data.

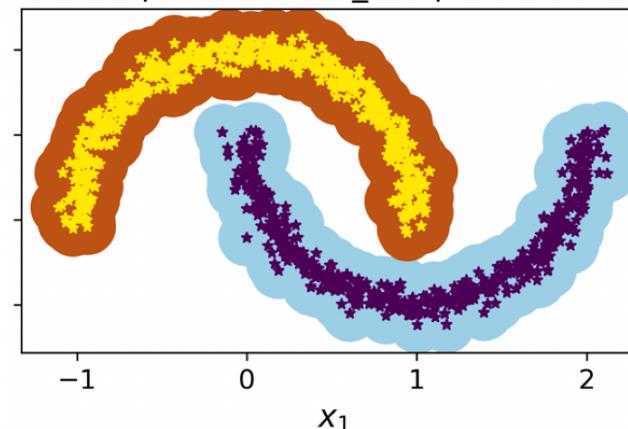
TWO MOONS DATA



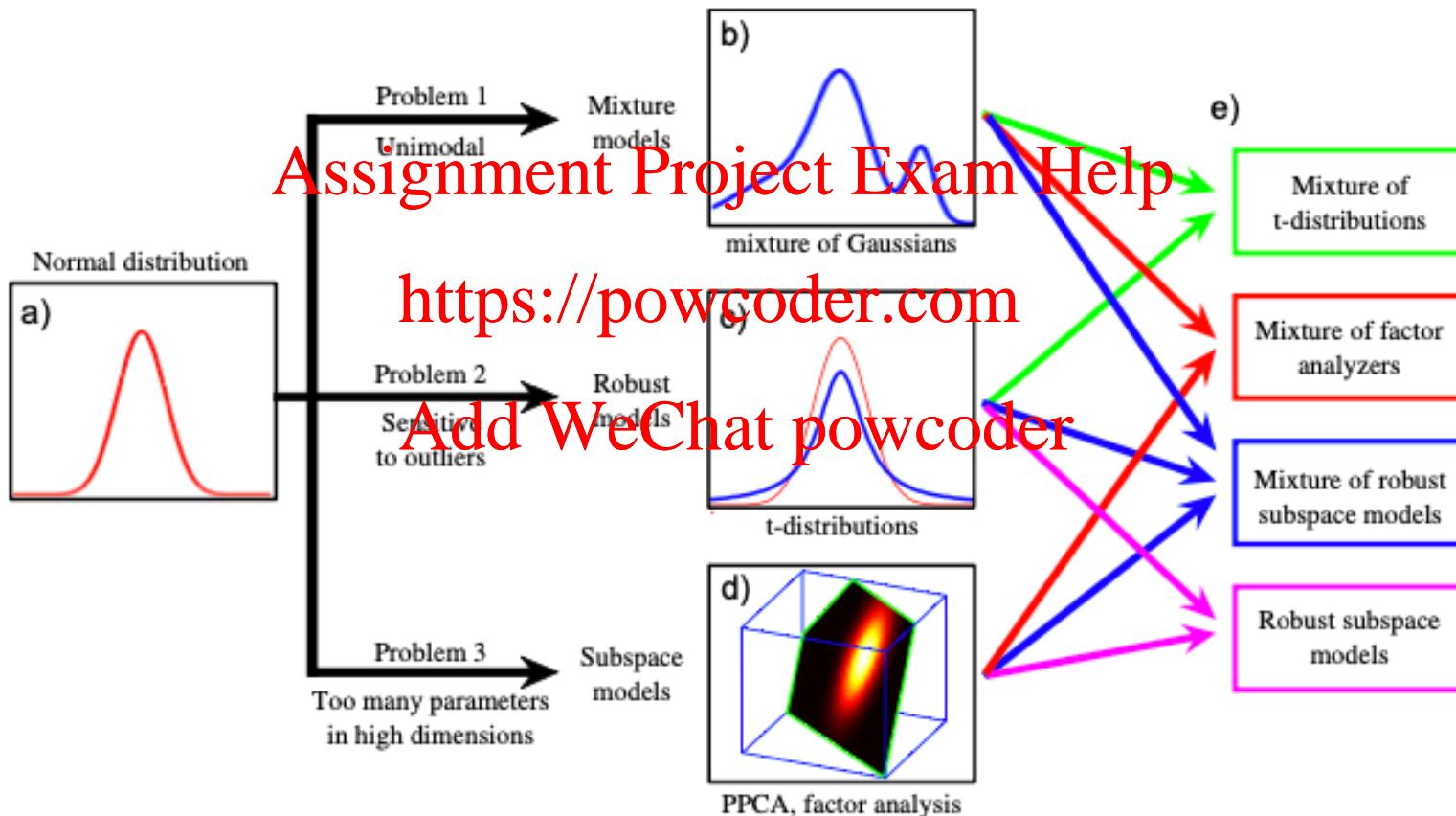
Assignment Project Exam Help
<https://powcoder.com>

Add WeChat powcoder

eps=0.20, min_samples=5



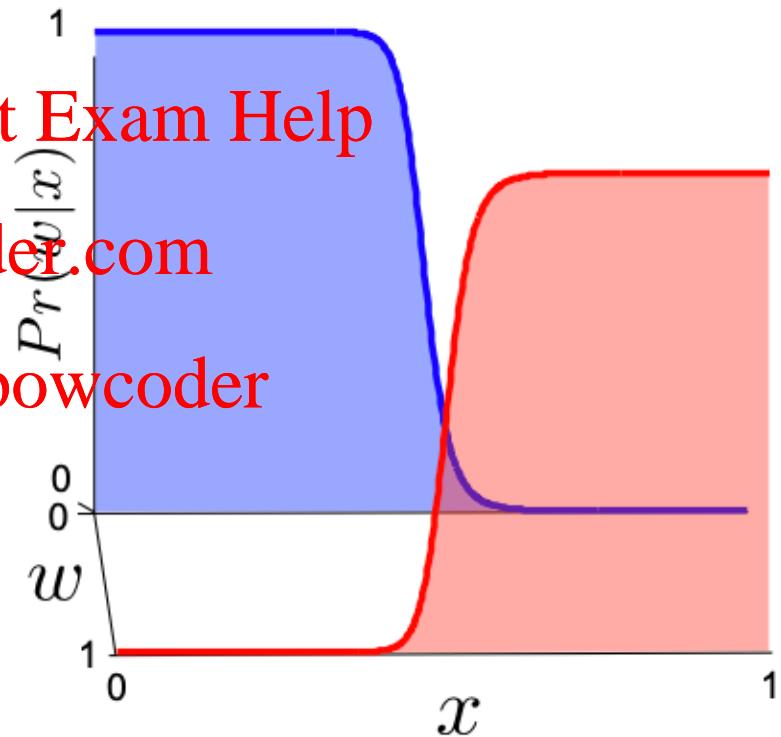
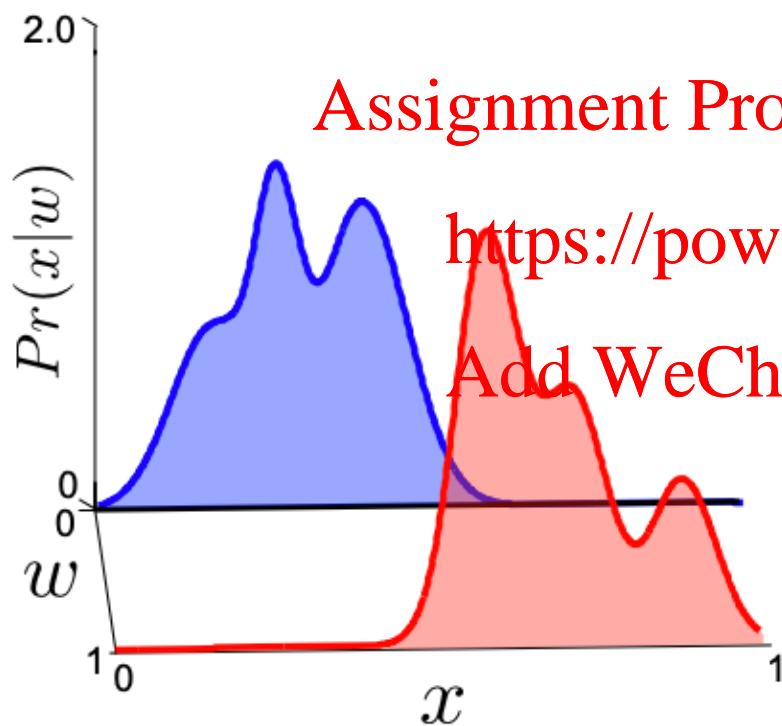
PROBLEMS WITH MULTI-VARIATE NORMAL DENSITY



Assignment Project Exam Help

<https://powcoder.com>
Types of models
Add WeChat powcoder

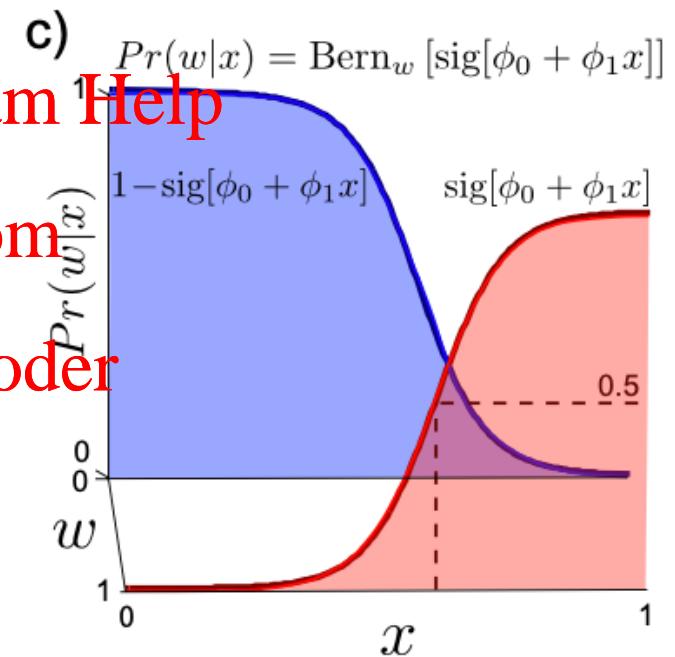
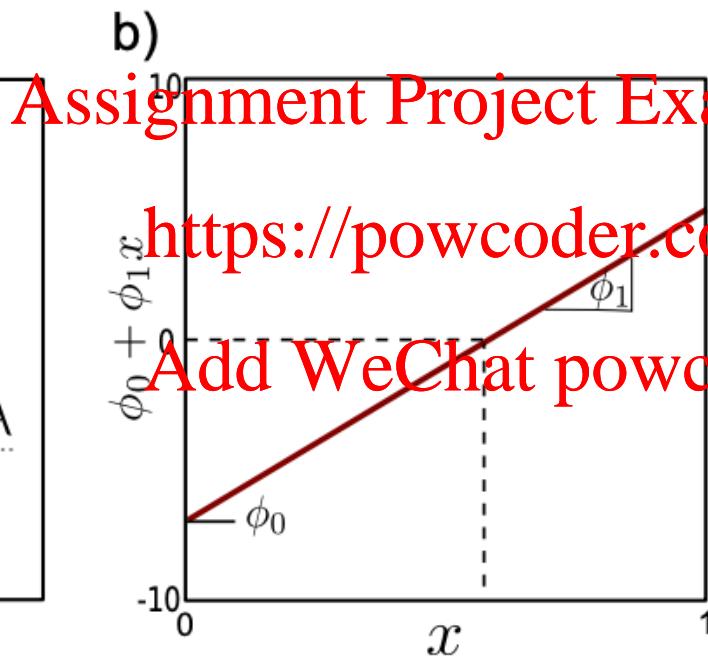
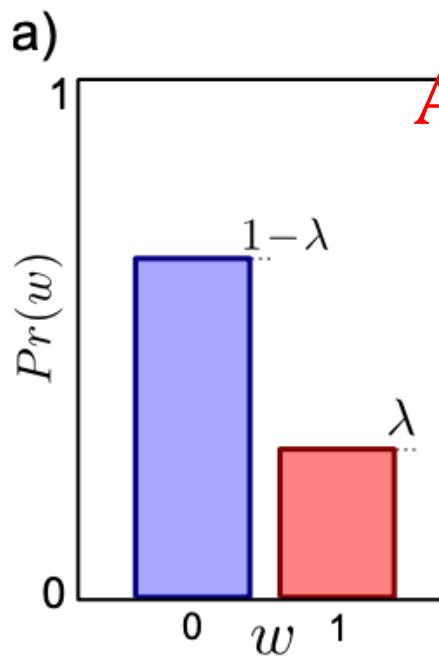
GENERATIVE VS DISCRIMINATIVE



CLASSIFICATION (DISCRIMINATIVE)

LOGISTIC REGRESSION REVISITED

MODEL CONTINGENCY OF THE WORLD ON DATA



World state:
Bernoulli distribution

Linear model

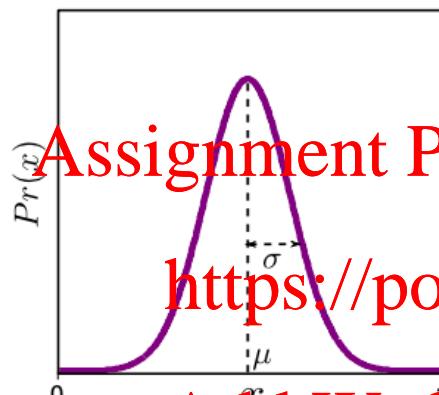
Probability /
Decision surface

CLASSIFICATION (GENERATIVE)

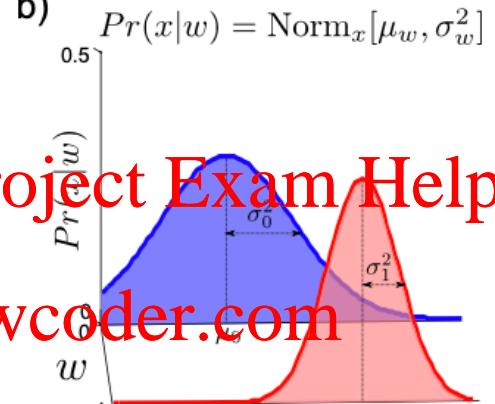
GAUSSIAN MIXTURE

MODEL CONTINGENCY OF DATA ON THE WORLD

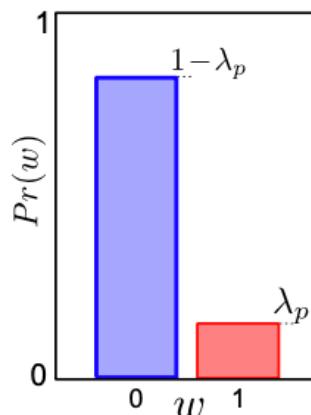
a)



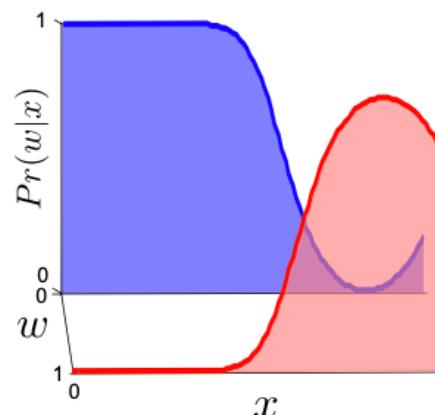
b)



c)



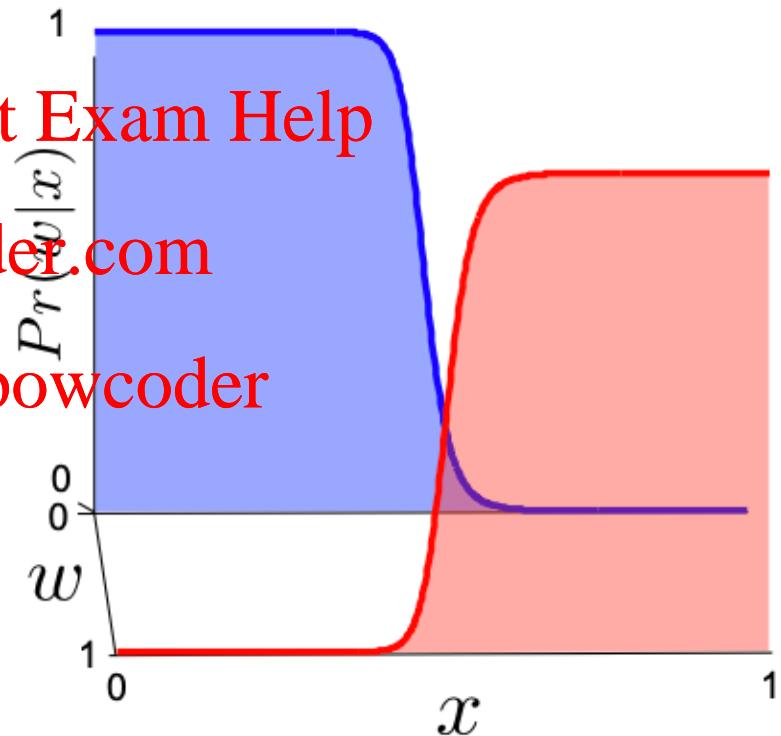
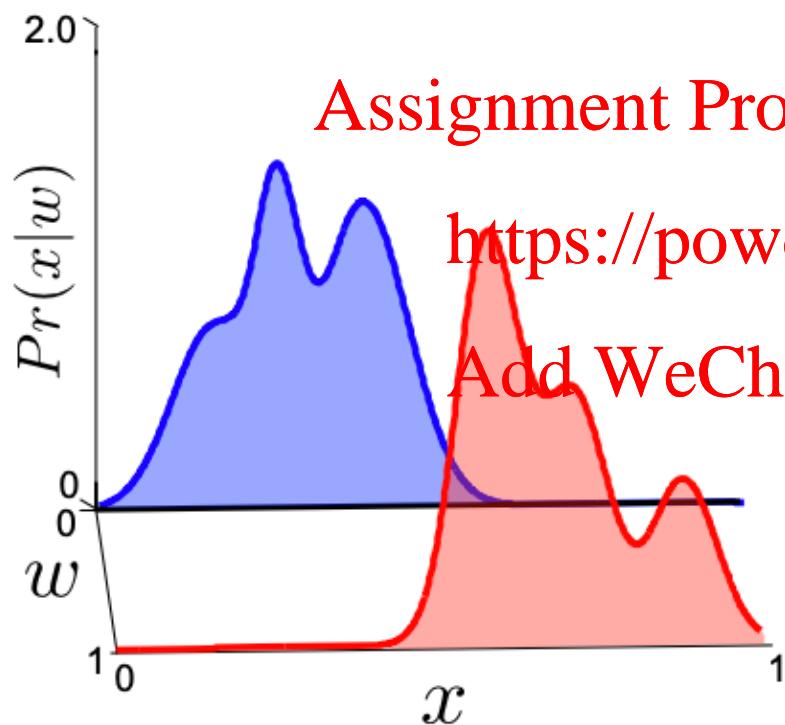
d)



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

Add WeChat powcoder

WHAT SORT OF MODEL SHOULD WE USE?



WHAT SORT OF MODEL SHOULD WE USE?

TL;DR NO DEFINITIVE ANSWER

- Inference is generally simpler with discriminative models.
 - Generative models calculate this probability via Bayes' rule, and sometimes this requires a computationally expensive algorithm.
 - Generative models might waste modelling power.
The data are generally of much higher dimension than the world, and modelling it is costly. Moreover, there may be many aspects of the data which do not influence the state;
- Assignment Project Exam Help**
<https://powcoder.com>
Add WeChat powcoder
- Using discriminative approaches, it is harder to exploit this knowledge: essentially we have to re-learn these phenomena from the data.
 - Sometimes parts of the training or test data vector \mathbf{x} may be missing. Here, generative models are preferred.
 - It is harder to impose prior knowledge in a principled way in discriminative models.

SUMMARY OF APPROACHES

Assignment Project Exam Help

Model $Pr(w|x)$

Model $Pr(x|w)$

<https://powcoder.com>

Regression $x \in [-\infty, \infty], w \in [-\infty, \infty]$	Linear regression	Linear regression
Classification $x \in [-\infty, \infty], w \in \{0, 1\}$	Logistic regression	Probability density function

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>
Best practice...

Add WeChat powcoder

BEST PRACTICE...



Julius Adebayo
@julius_adebayo

Reading through open source code for a nice paper,
found this gem :)

```
# This controls the behavior of BatchNormalizer in UMAP
# For some reason I don't fully understand, even though we are
# not training these models where this flag is used,
# it must be set to True or training does not converge.
TRAINING_KWARG_FOR_SECOND_MODEL = True
```

8:54 AM · Feb 8, 2020 · Twitter Web App

7 Retweets 61 Likes



Sean Augenstein @seanAugenstein · 23h
Replying to @julius_adebayo

Hi, I'm one of the authors of this paper! (and the #tensorflowfederated #tff GAN code). Thanks for appreciating this and not finding it infuriating :-) Credit to my esteemed colleague Brendan McMahan for the exact wording of the comment. We have some scars from debugging this.

1 4

Assignment Project Exam Help

Add WeChat powcoder



Sean Augenstein @seanAugenstein · 23h

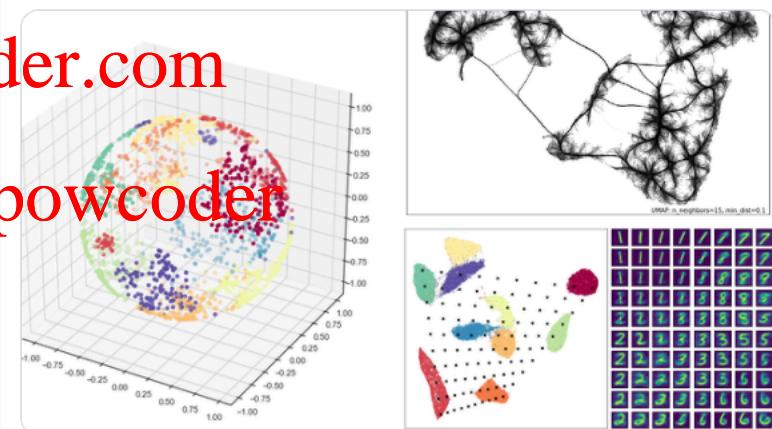
I'll be @ICLR2020 presenting this work, if you're attending I'd love to chat about using generative models with #differentialprivacy and #federatedlearning to gain insights into decentralized data in a privacy-preserving way (for debugging ML models and other use cases).

1 2



Leland McInnes
@leland_mcinnes

The first release candidate for UMAP 0.4 is out providing lots of new features, including performance improvements, embedding to different manifolds, inverse transform, and plotting tools.

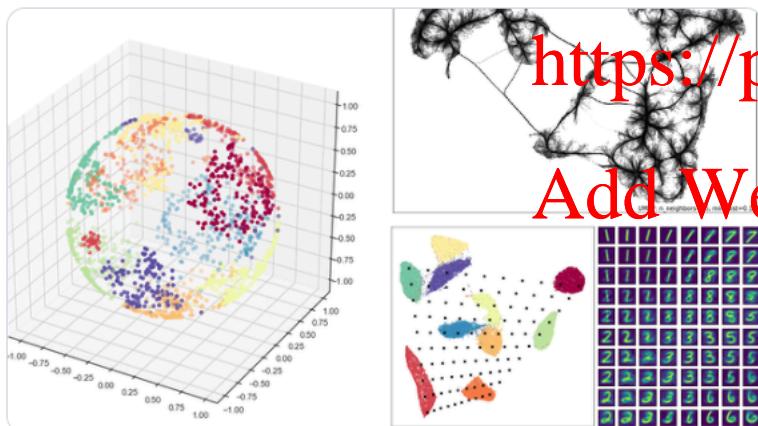


5:35 PM · Feb 9, 2020 · Twitter Web App

BEST PRACTICE...



The first release candidate for UMAP 0.4 is out providing lots of new features, including performance improvements, embedding to different manifolds, inverse transform, and plotting tools.



5:35 PM · Feb 9, 2020 · Twitter Web App

Assignment Project Exam Help

<https://powcoder.com>

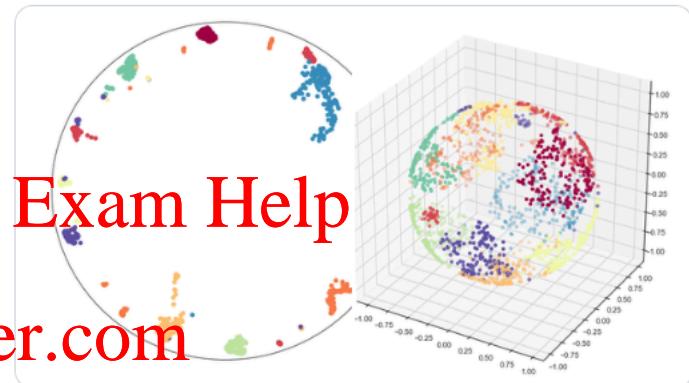
Add WeChat powcoder



Leland McInnes @leland_mcinnes · 3h

Replying to @leland_mcinnes

UMAP 0.4 supports embedding to non-Euclidean manifolds, including spheres, Poincare disks, and more. <umap-learn.readthedocs.io/en/latest/embedding.html>



2

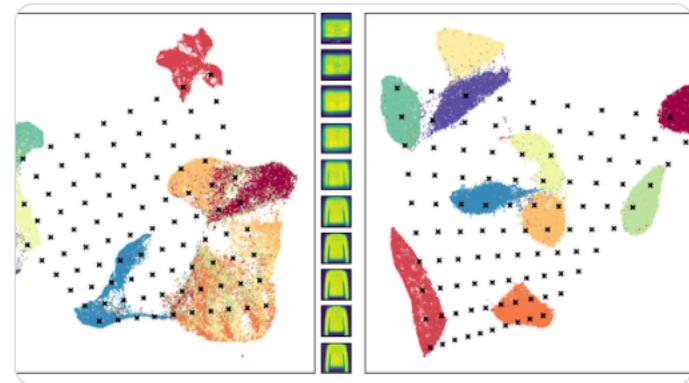
3

26



Leland McInnes @leland_mcinnes · 3h

Support for an "inverse transform" has been added to UMAP 0.4, providing the ability to generate a high dimensional representation of a point in the embedding space. umap-learn.readthedocs.io/en/latest/inverse_transform.html



1

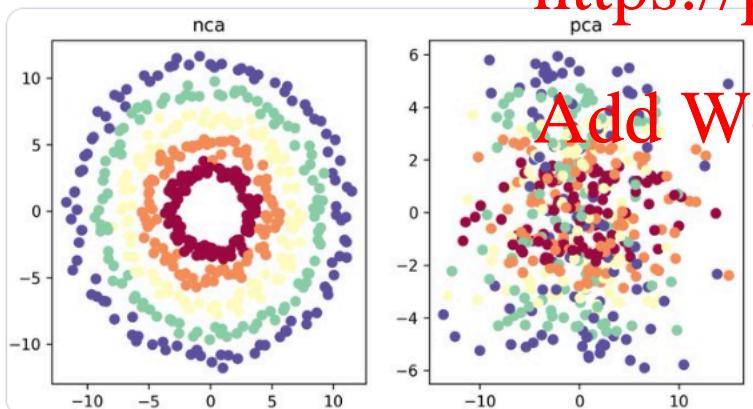
4

19



Kevin Zakka
@kevin_zakka

Implemented a really neat ML algorithm called Neighbourhood Components Analysis in [@PyTorch](#). It allows you to learn a linear transformation of your dataset that maximizes the performance of KNN. For high noise variance datasets, it handily beats PCA.
github.com/kevinzakka/nca



3:07 AM · Feb 8, 2020 · [Twitter Web App](#)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Neighbourhood Components Analysis

Jacob Goldberger, Sam Roweis, Geoff Hinton, Ruslan Salakhutdinov
Department of Computer Science, University of Toronto
{jacob, roweis, hinton, rsalakhu}@cs.toronto.edu

Abstract

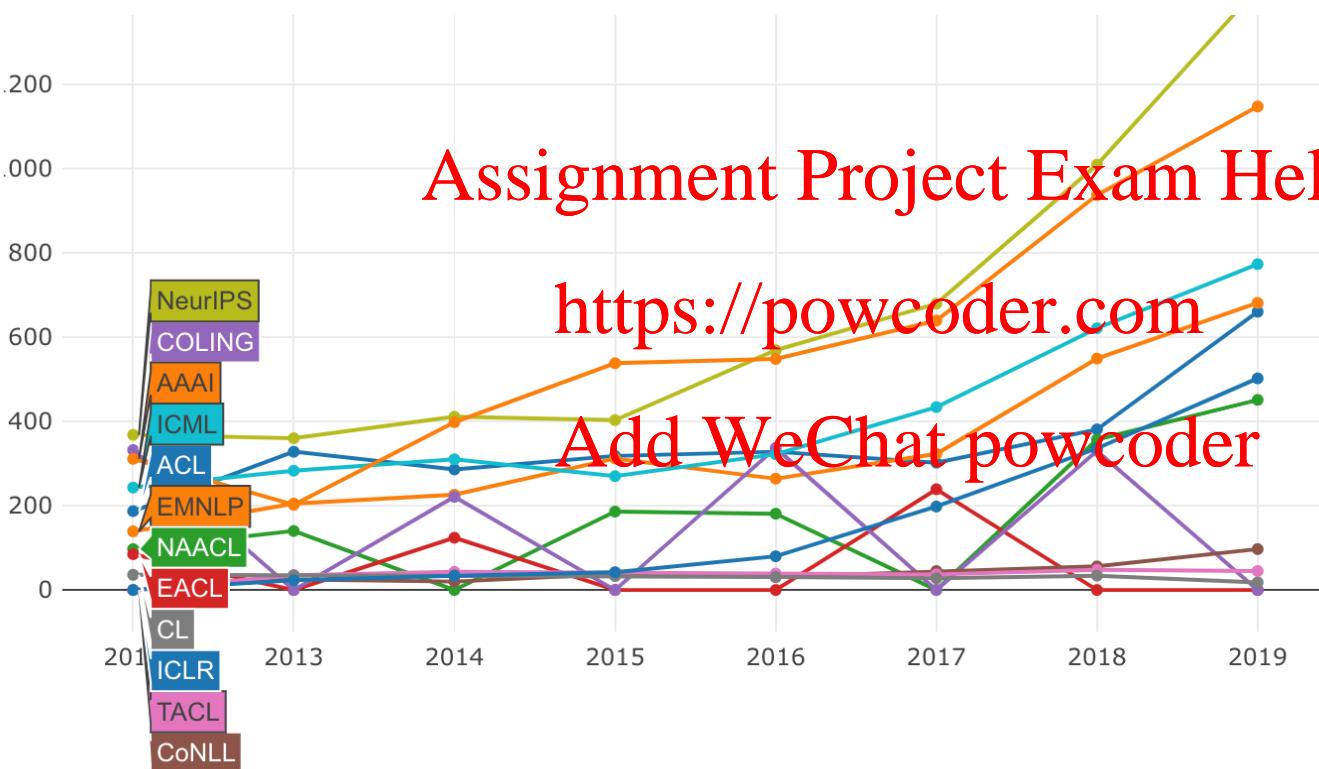
In this paper we propose a novel method for learning a Mahalanobis distance measure to be used in the KNN classification algorithm. The algorithm directly maximizes a stochastic variant of the leave-one-out KNN score on the training set. It can also learn a low-dimensional linear embedding of labeled data that can be used for data visualization and classification. Unlike other methods, our classification model is non-parametric, making no assumptions about the shape of the class distributions or the boundaries between them. The performance of the method is demonstrated on several data sets, both for metric learning and linear dimensionality reduction.

1 Introduction

Nearest neighbor (KNN) is an extremely simple yet surprisingly effective method for classification. Its appeal stems from the fact that its decision surfaces are nonlinear, there is only a single integer parameter (which is easily tuned with cross-validation), and the expected quality of predictions improves automatically as the amount of training data increases. These advantages, shared by many non-parametric methods, reflect the fact that although the final classification machine has quite high capacity (since it accesses the entire reservoir of training data at test time), the trivial learning procedure rarely causes overfitting itself.

However, KNN suffers from two very serious drawbacks. The first is computational, since it must store and search through the entire training set in order to classify a single test point. (Storage can potentially be reduced by “editing” or “thinning” the training data; and in low dimensional input spaces, the search problem can be mitigated by employing data structures such as KD-trees or ball-trees[4].) The second is a modeling issue: how should the distance metric used to define the “nearest” neighbours of a test point be defined? In this paper, we attack both of these difficulties by learning a quadratic distance metric which optimizes the

BEST PRACTICE...



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

Percentage of papers mentioning GitHub (indicating that the code is made available):
ACL 70%,
EMNLP 69%,
NAACL 68%
ICLR 56%,
NeurIPS 46%,
ICML 45%,
AAAI 31%.
It seems the NLP papers are releasing their code much more freely.



Trending Research

Trending

Latest

Greatest

Subscribe

Assignment Project Exam Help

1 Jan 2021 • lab-mmlu • PyTorch

★ 2,641
5.33 stars / hour

Paper

Code

<https://powcoder.com>

Add WeChat powcoder

Dynamically Routing Between Capsules

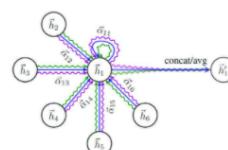
★ 2,036
4.85 stars / hour

We use the length of the activity vector to represent the probability that the entity exists and its orientation to represent the instantiation parameters.

Ranked #1 on [Image Classification on MultiMNIST](#)

Paper

Code



Fake News Detection on Social Media using Geometric Deep Learning

10 Feb 2019 • gordicaleksa/pytorch-GAT • PyTorch

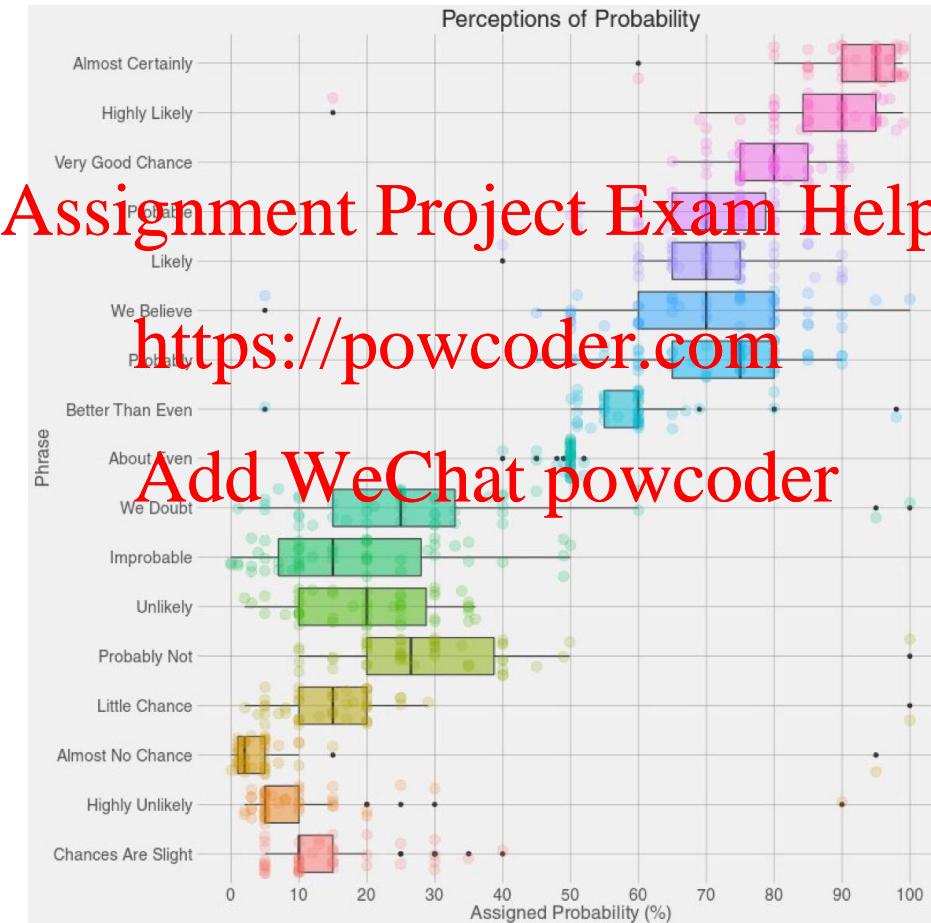
★ 634
3.91 stars / hour

One of the main reasons is that often the interpretation of the news requires the knowledge of political or social context or 'common sense', which current NLP algorithms are still missing.

Paper

Code

PERCEPTIONS OF PROBABILITY





Santiago
@svpino

...

You can blow everyone's mind on your machine learning interview.

Assignment Project Exam Help

- Pick a problem
- Build a model that solves it
- Build an API around it
- Host it
- Set up monitoring
- Build a retraining pipeline
- Write about it

Add WeChat powcoder

This will put you on a whole new level.

5:33 am · 5 Feb 2021 · Twitter Web App

@chipro

@random_forests

@zacharylipton

@yudapearl

@svpino

@jackclarkSF

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

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**
Individual Coursework workshop
to **Thursday 11th Feb 2021 at 12:00 am**

LECTURE 3 TERM 2:

MSIN0097

UCL
SCHOOL OF
MANAGEMENT

PREDICTIVE ANALYTICS Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

A P MOORE