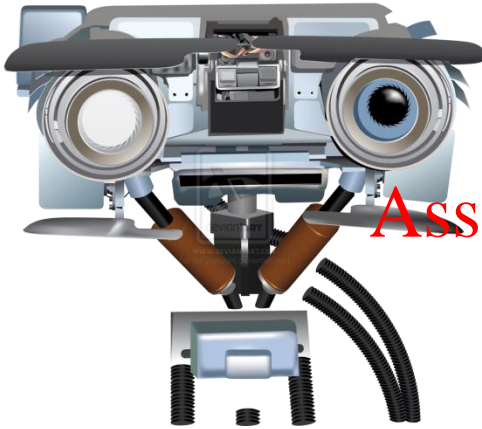


Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help

Supervised Learning III

<https://powcoder.com>

Add WeChat powcoder

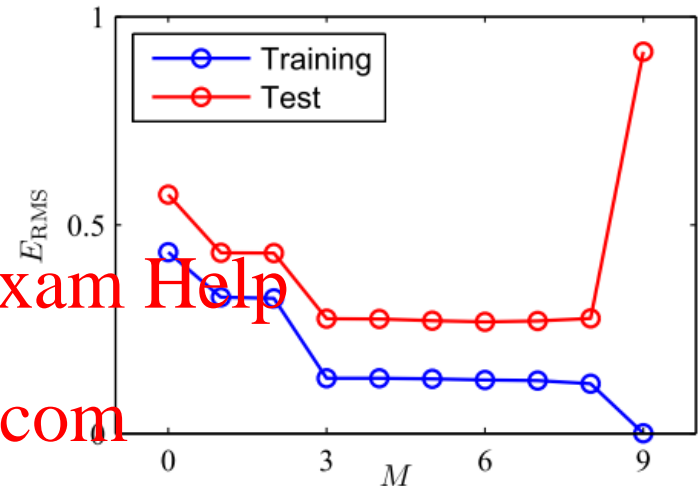
Classification, Regularization

Assignment Project Exam Help Detecting overfitting

Add WeChat powcoder

Plot model complexity versus
objective function on test/train data

As model becomes more complex,
performance on training keeps
improving while on test data it increases



<https://powcoder.com>

Horizontal axis: measure of model complexity
In this example, we use the maximum order of the polynomial basis
functions.

Vertical axis: For regression, it would be SSE or mean SE (MSE)
For classification, the vertical axis would be classification error rate or
cross-entropy error function

Assignment Project Exam Help

Overcoming overfitting

Add WeChat powcoder

- Basic ideas

- Use more training data.

Assignment Project Exam Help

- Regularization methods

- Cross-validation

<https://powcoder.com>

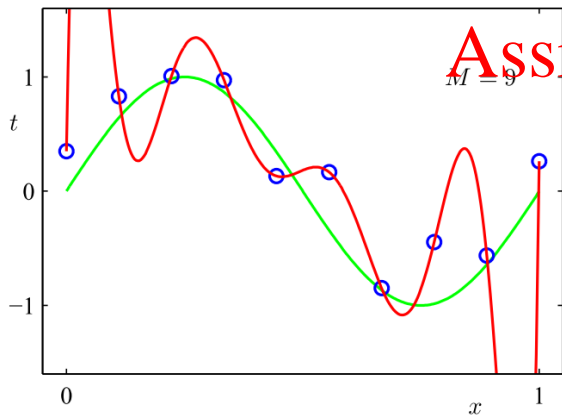
Add WeChat powcoder

Assignment Project Exam Help

Solution: use more data

Add WeChat powcoder

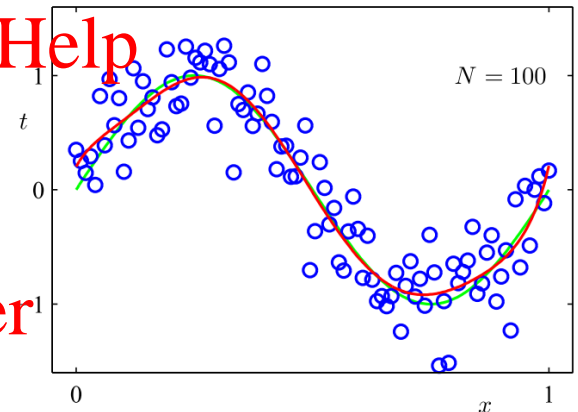
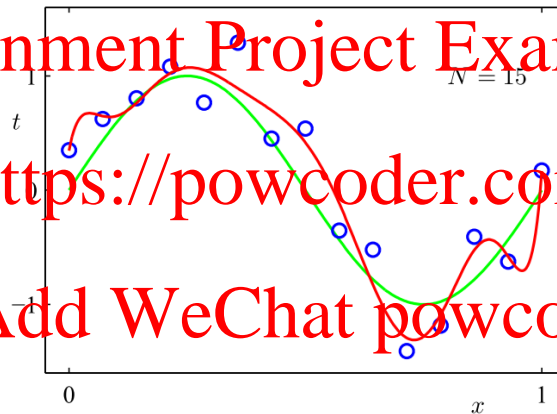
$M=9$, increase N



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



What if we do not have a lot of data?

Assignment Project Exam Help

Overcoming overfitting

Add WeChat powcoder

- Basic ideas

- Use more training data

Assignment Project Exam Help

- Regularization methods

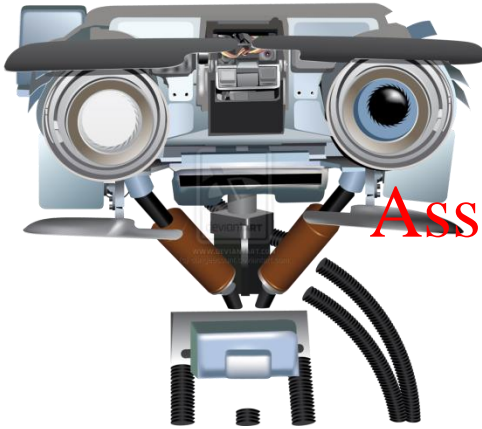
<https://powcoder.com>

- Cross-validation

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help Supervised Learning III

<https://powcoder.com>

Add WeChat powcoder

Regularization

Assignment Project Exam Help

Solution: Regularization

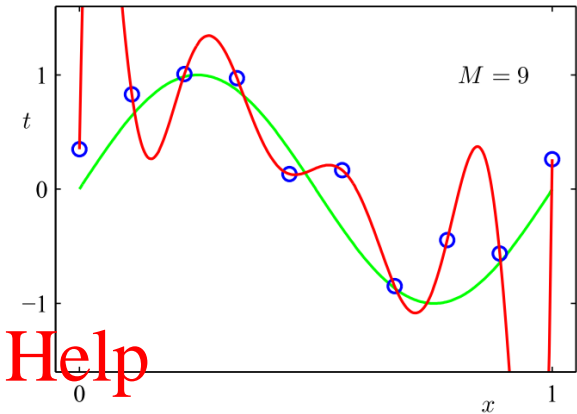
Add WeChat powcoder

- Use regularization:
 - Add $\lambda \|\theta\|_2^2$ term to SSE cost function
 - “L-2” norm squared, ie sum of sq. elements $\sum \theta_j^2$
 - Penalizes large θ
 - λ controls amount of regularization

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



M = 9

0.35

232.37

-5321.83

48568.31

-231639.30

640042.26

-1061800.52

1042400.18

-557682.99

125201.43

Assignment Project Exam Help

Regularized Linear Regression

Add WeChat powcoder

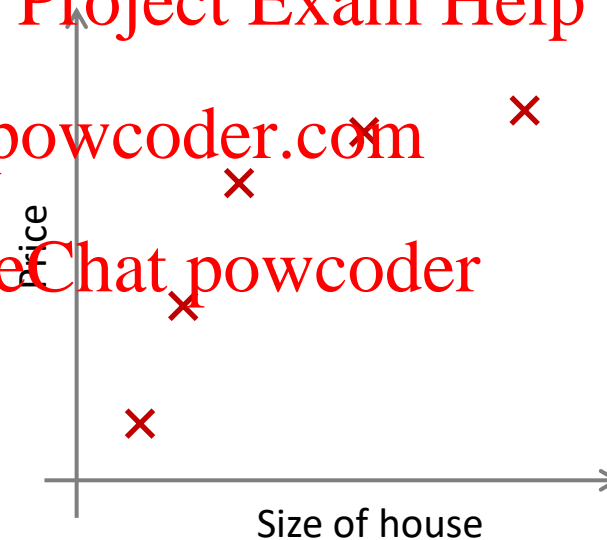
$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

$$\min_{\theta} J(\theta)$$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Gradient descent for Linear Regression

Add WeChat powcoder

Repeat {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

replace with

<https://powcoder.com>

}

$$\theta_j := \theta_j (1 - \alpha \frac{\lambda}{m}) - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Assignment Project Exam Help

Regularized Normal Equation

Add WeChat powcoder

Suppose $m \leq n$,
(#examples) (#features)

$$\theta = (X^T X)^{-1} X^T y \quad \text{Non-invertible/singular}$$

Assignment Project Exam Help

If $\lambda > 0$,

<https://powcoder.com>

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \right)^{-1} X^T y$$

Add WeChat powcoder

Assignment Project Exam Help

Regularized Logistic Regression

Add WeChat powcoder

Hypothesis:

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Assignment Project Exam Help

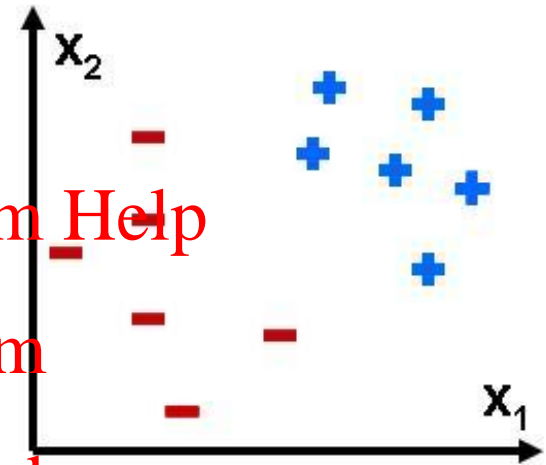
<https://powcoder.com>

Add WeChat powcoder

Cost Function:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right]$$

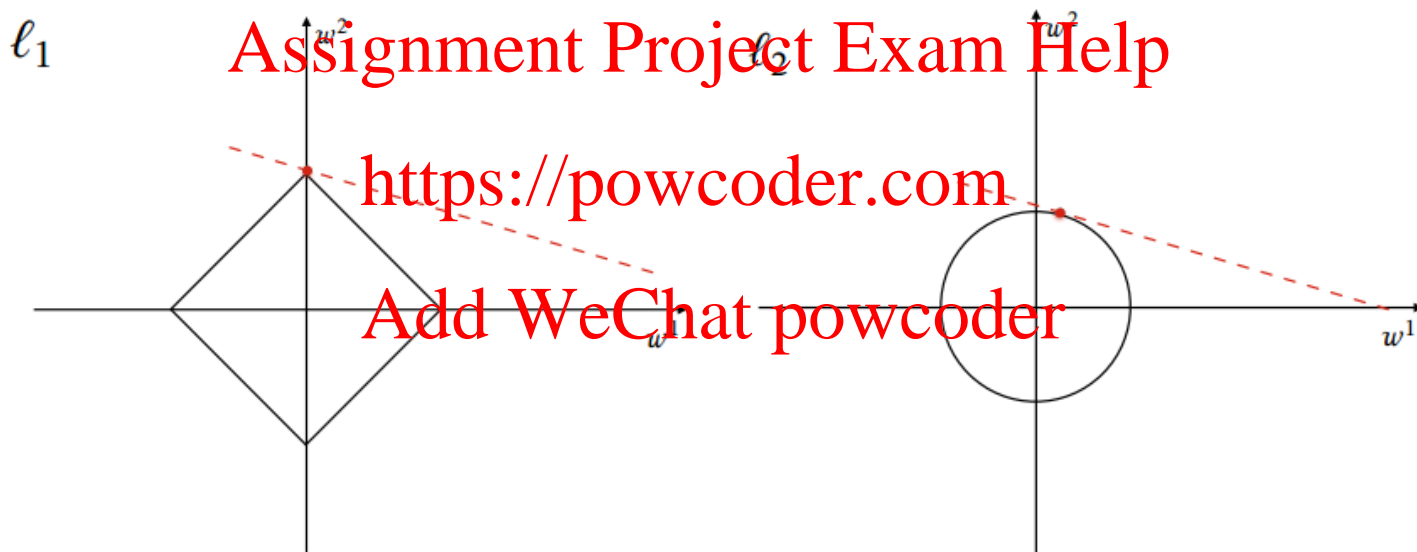
Goal: minimize cost $\min_{\theta} J(\theta)$



Assignment Project Exam Help

Many types of Regularization

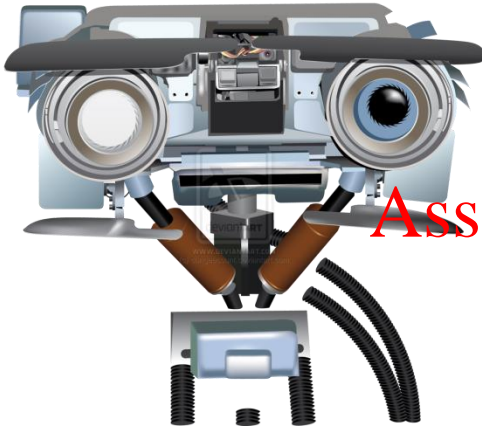
- Most common are ℓ_1 and ℓ_2



ℓ_1 often used to create sparsity

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help Supervised Learning III

<https://powcoder.com>

Add WeChat powcoder

Bias-Variance

Assignment Project Exam Help

Bias vs Variance

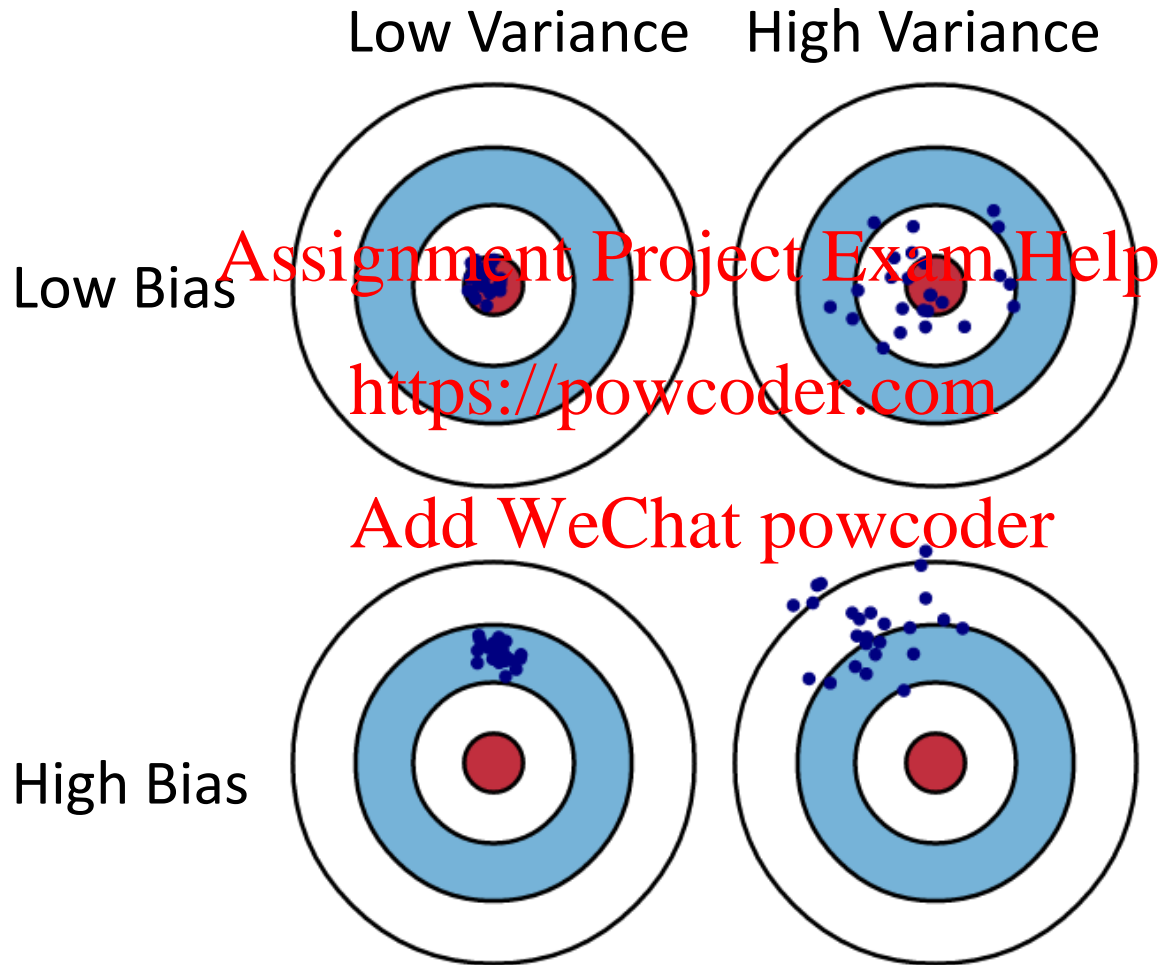
- Understanding how different sources of error lead to bias and variance helps us improve model fitting

Assignment Project Exam Help

- **Error due to Bias:** The error due to bias is taken as the difference between the expected (or average) prediction of our model and the correct value which we are trying to predict (imagine you could repeat the whole model fitting process on many datasets)
- **Error due to Variance:** The variance is how much the predictions for a given point vary between different realizations of the model.

Assignment Project Exam Help

Graphical Illustration



Assignment Project Exam Help

The Bias-Variance Trade-off

There is a trade-off between bias and variance:

- **Less complex** models (fewer parameters) have high bias and hence low variance
- **More complex** models (more parameters) have low bias and hence high variance
- **Optimal** model will have a balance

Assignment Project Exam Help

Which is worse?

Add WeChat powcoder

- A gut feeling many people have is that they should minimize bias even at the expense of variance

Assignment Project Exam Help

- This is mistaken logic. It is true that a high variance and low bias model can perform well in some sort of long-run average sense. However, in practice modelers are always dealing with a single realization of the data set
- In these cases, long run averages are irrelevant, **bias and variance are equally important**, and one should not be improved at an excessive expense to the other.

<https://powcoder.com>

Add WeChat powcoder

How to deal with bias/variance

Add WeChat powcoder

- Can deal with variance by
 - Bagging, e.g. Random Forest
 - Bagging trains multiple models on random subsamples of the data, averages their prediction
- Can deal with high bias by
 - Decreasing regularization/increasing complexity of model
 - Also known as *model selection*

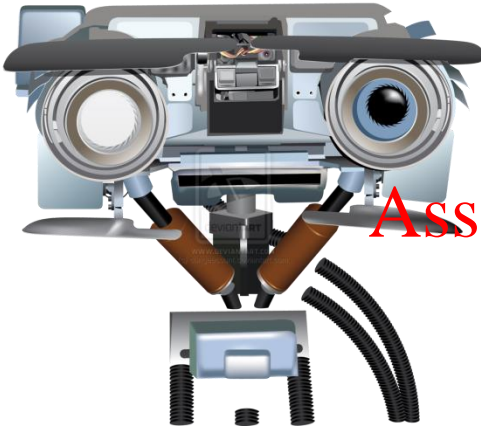
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help Supervised Learning III

<https://powcoder.com>

Add WeChat powcoder

Model selection and
training/validation/test sets

Assignment Project Exam Help

Add WeChat powcoder



*Not performing well
on training data*

(underfit)

Add WeChat powcoder

*Not generalizing well from
training to unseen data*

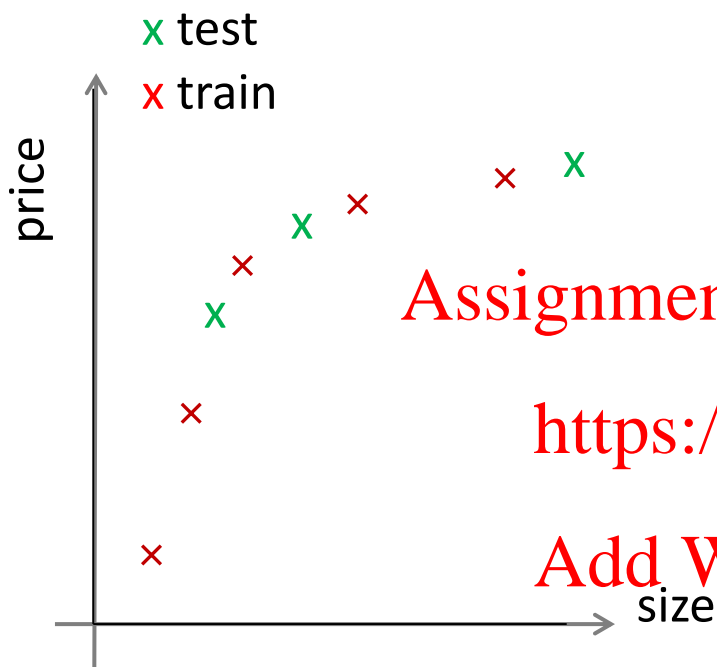
(overfit)

Model selection

Assignment Project Exam Help

Add WeChat powcoder

Hyperparameters (e.g., degree of polynomial, regularization weight, learning rate) must be selected prior to training.



Assignment Project Exam Help

How to choose them?

<https://powcoder.com>

Add WeChat powcoder

Try several values, choose one with the lowest test error?

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

Problem: test error is likely an overly optimistic estimate of generalization error because we “cheat” by fitting the hyperparameter to the actual test examples.

Train/Validation/Test Sets

Assignment Project Exam Help
Add WeChat powcoder

	Size	Price
train	2104	400
	1600	330
	2400	369
	1416	232
	3000	540
validation	1985	300
	1534	315
	1427	199
test	1380	212
	1494	243

Solution: split data into three sets.

For each value of a hyperparameter, train on the train set, evaluate learned parameters on the validation set.

Pick the model with the hyperparameter that achieved the lowest validation error.

Report this model's test set error.

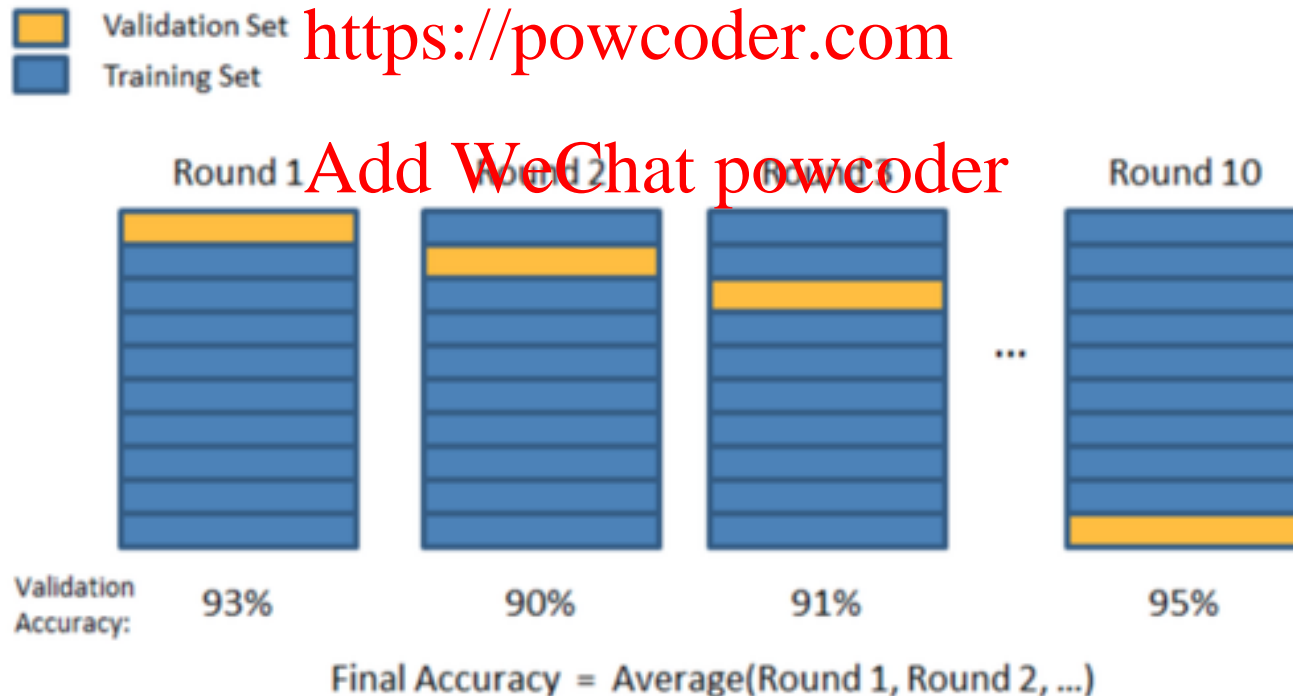
N-Fold Cross Validation

Add WeChat powcoder

- What is we don't have enough data for train/test/validation sets?
- Solution: use N-fold cross validation.
- Split training set into train/validation sets N times
- Report average predictions over N val sets, e.g. N=10:

Assignment Project Exam Help

<https://powcoder.com>



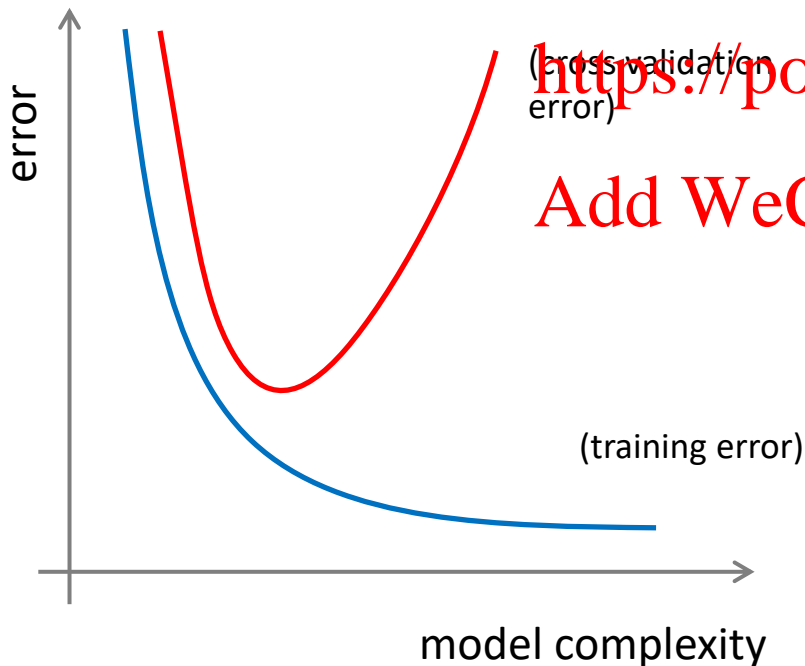
Add WeChat powcoder

Diagnosing bias vs. variance

Add WeChat powcoder

Suppose your learning algorithm is performing less well than you were hoping. ($J_{cv}(\theta)$ or $J_{test}(\theta)$ is high.) Is it a bias problem or a variance problem?

Assignment Project Exam Help Bias (underfit):



<https://powcoder.com> $J_{train}(\theta)$ will be high,

Add WeChat powcoder $J_{cv}(\theta) \approx J_{train}(\theta)$

Variance (overfit):

$J_{train}(\theta)$ will be low,

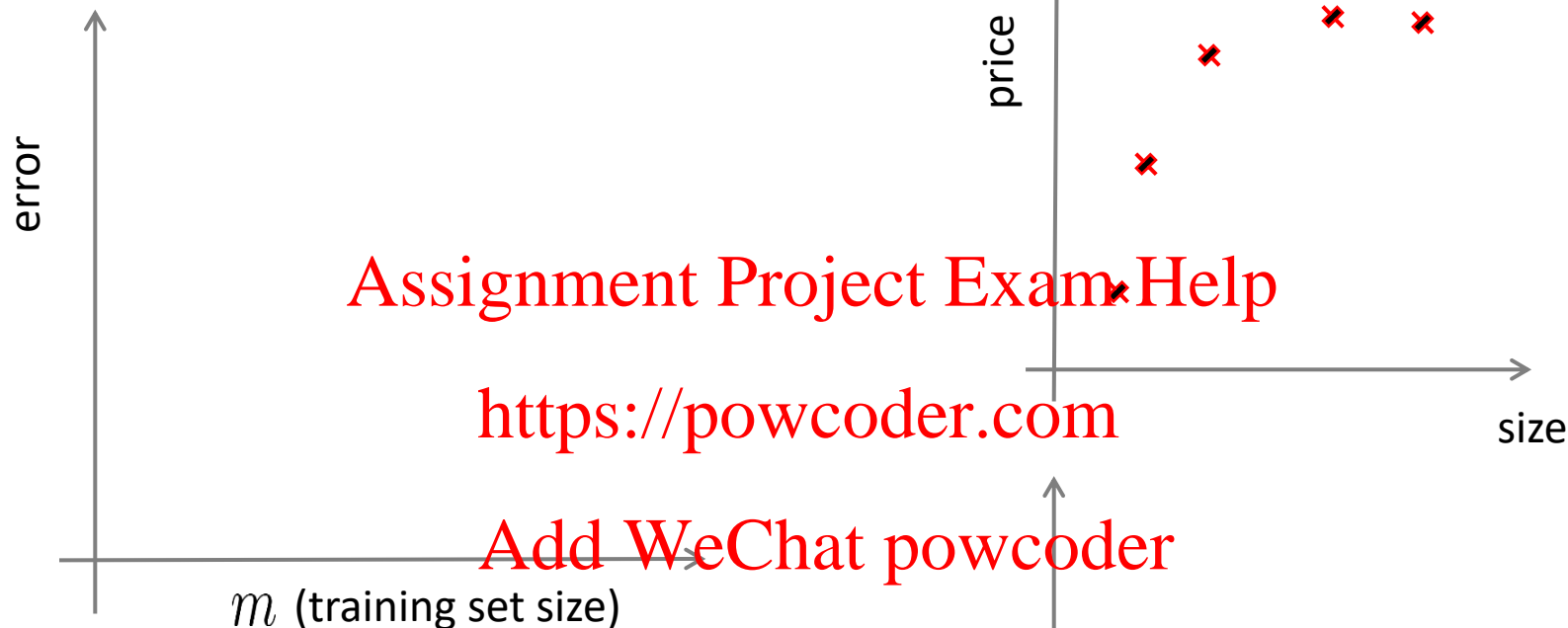
$J_{cv}(\theta) \gg J_{train}(\theta)$

Learning Curves

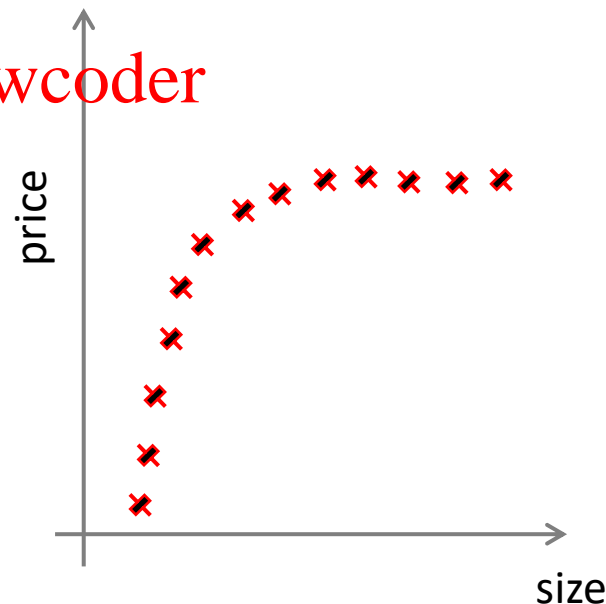
High bias

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Add WeChat powcoder



If a learning algorithm is suffering from high bias, getting more training data will not (by itself) help much.

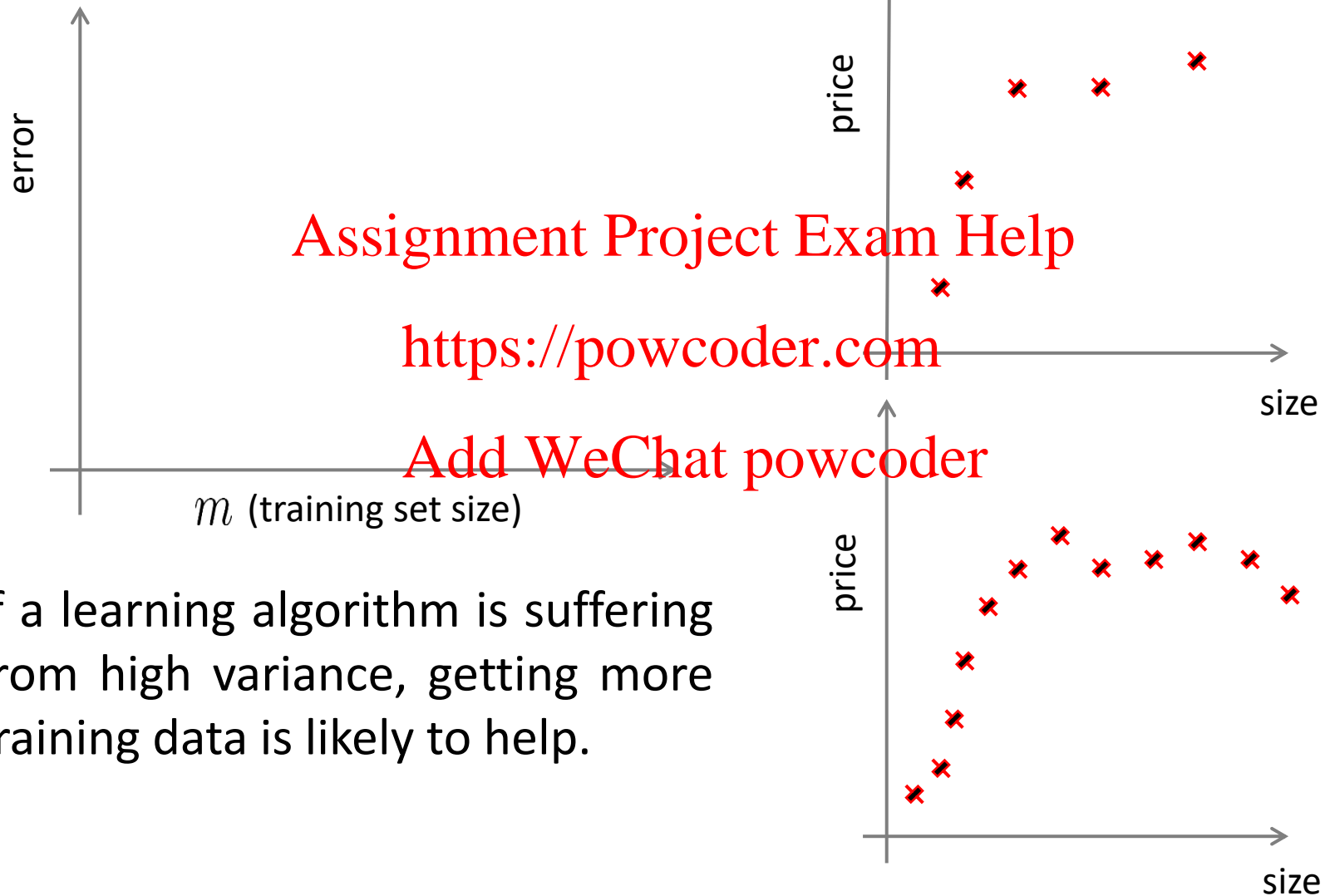


Learning Curves: Assignment Project Exam Help

High variance

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small λ)



If a learning algorithm is suffering from high variance, getting more training data is likely to help.

Assignment Project Exam Help

Debugging a learning algorithm

Suppose you have implemented regularized linear regression to predict housing prices. However, when you test your hypothesis in a new set of houses, you find that it makes unacceptably large errors in its prediction. What should you try next?

To fix high variance

- Get more training examples
- Try smaller sets of features
- Try increasing λ

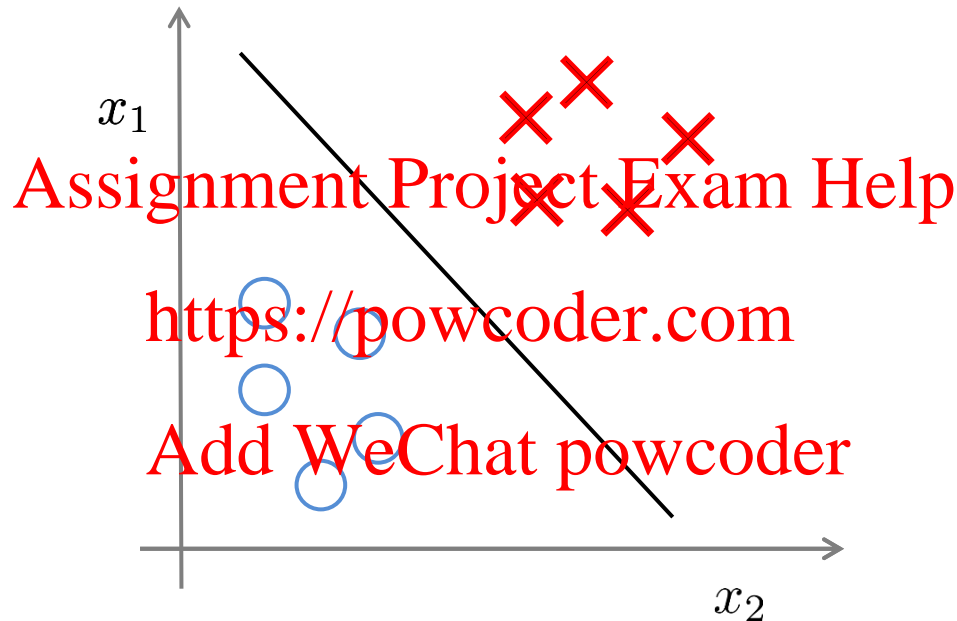
To fix high bias

- Try getting additional features
- Try adding polynomial features
- Try decreasing λ

Assignment Project Exam Help

Add WeChat powcoder

Supervised learning

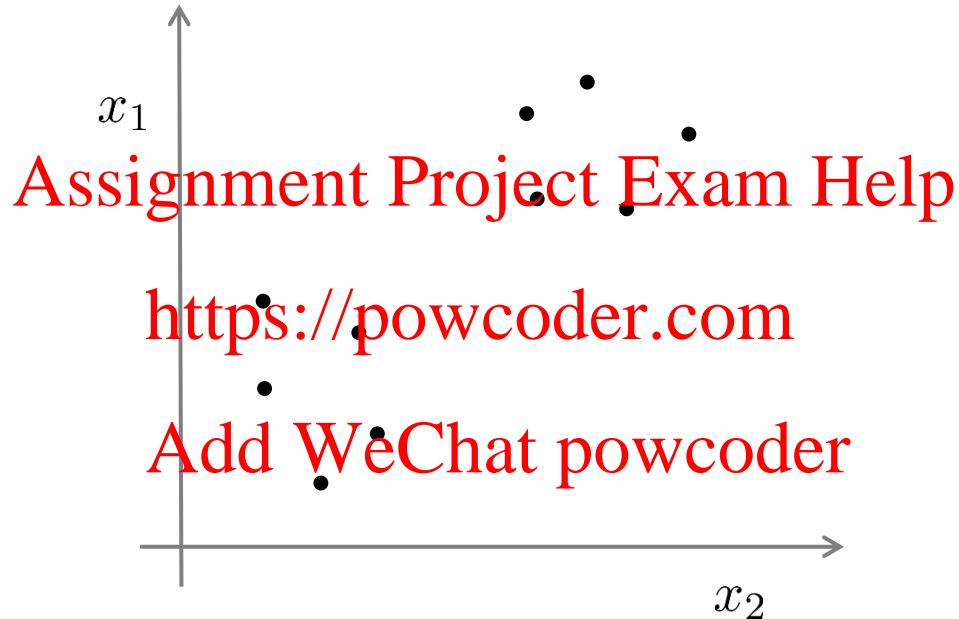


Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \dots, (x^{(m)}, y^{(m)})\}$

Assignment Project Exam Help

Add WeChat powcoder

Unsupervised learning



Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$

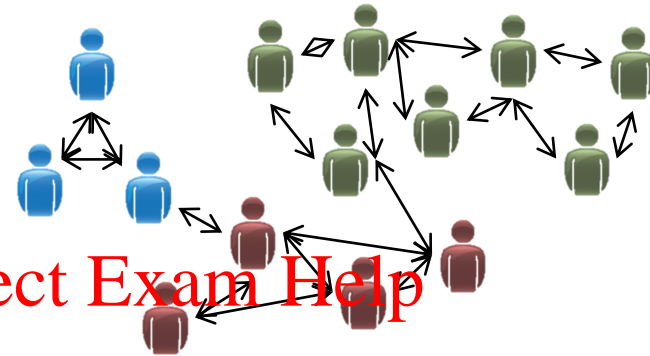
Clustering

Assignment Project Exam Help

Add WeChat powcoder



Gene analysis



Social network analysis

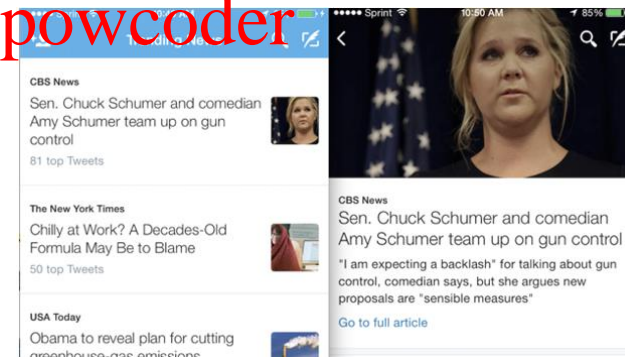
Assignment Project Exam Help

<https://powcoder.com>



Types of voters

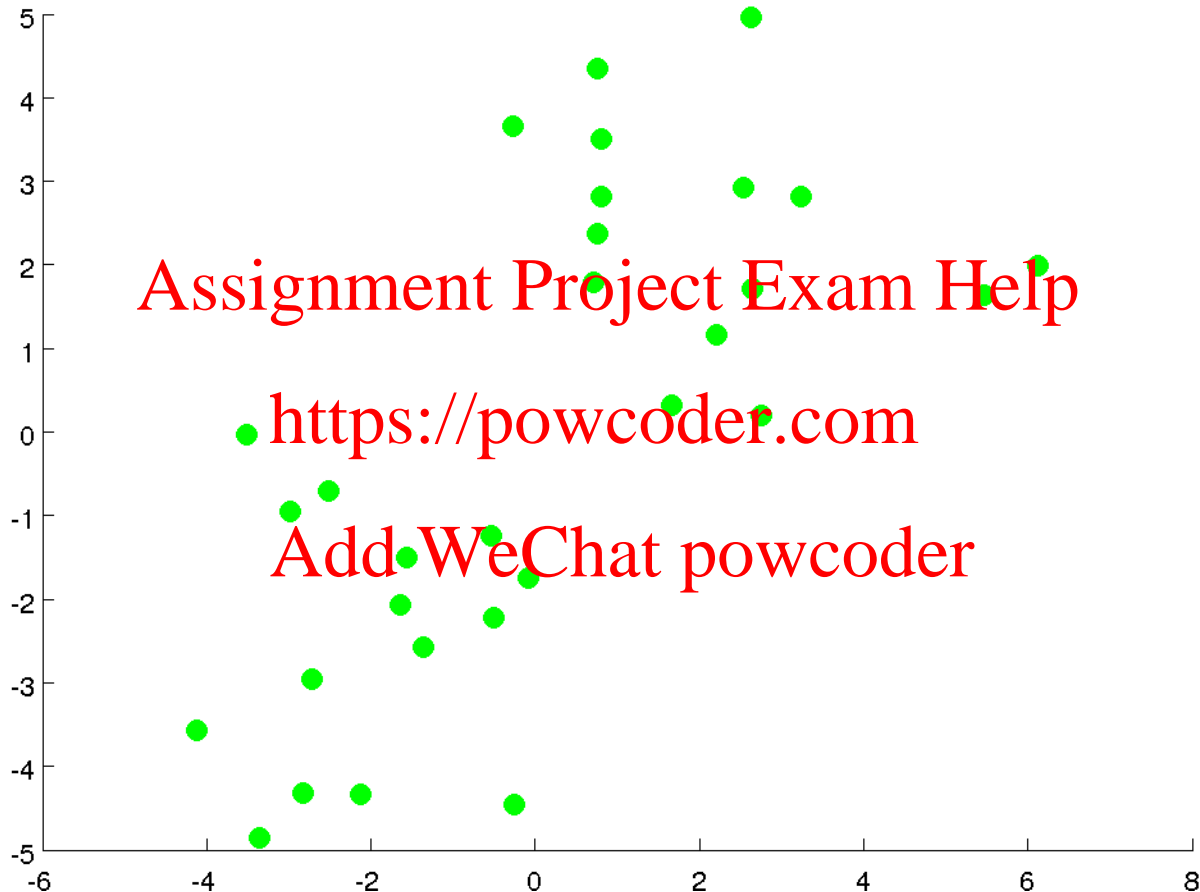
Add WeChat powcoder



Trending news

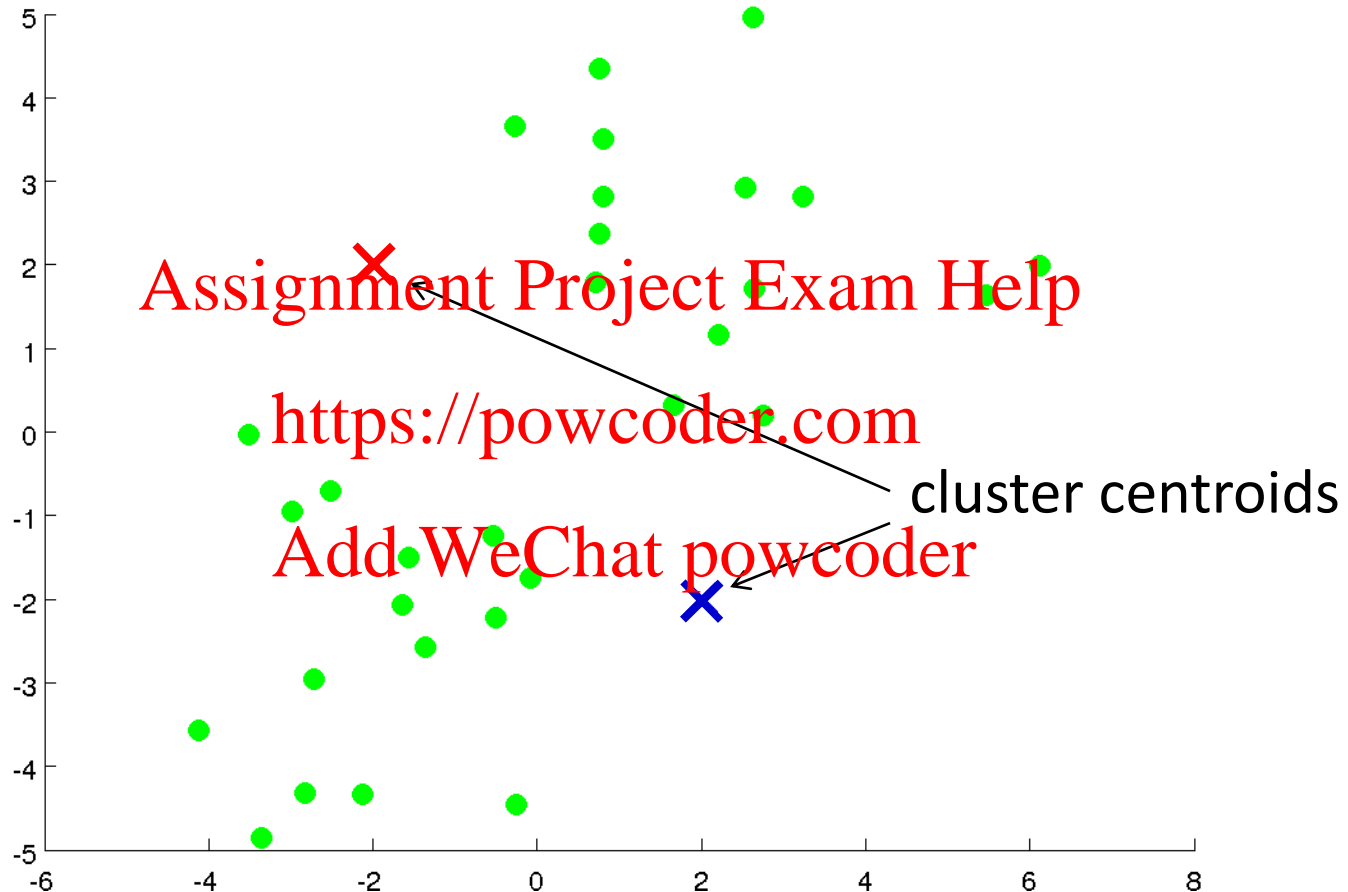
Assignment Project Exam Help

Add WeChat powcoder



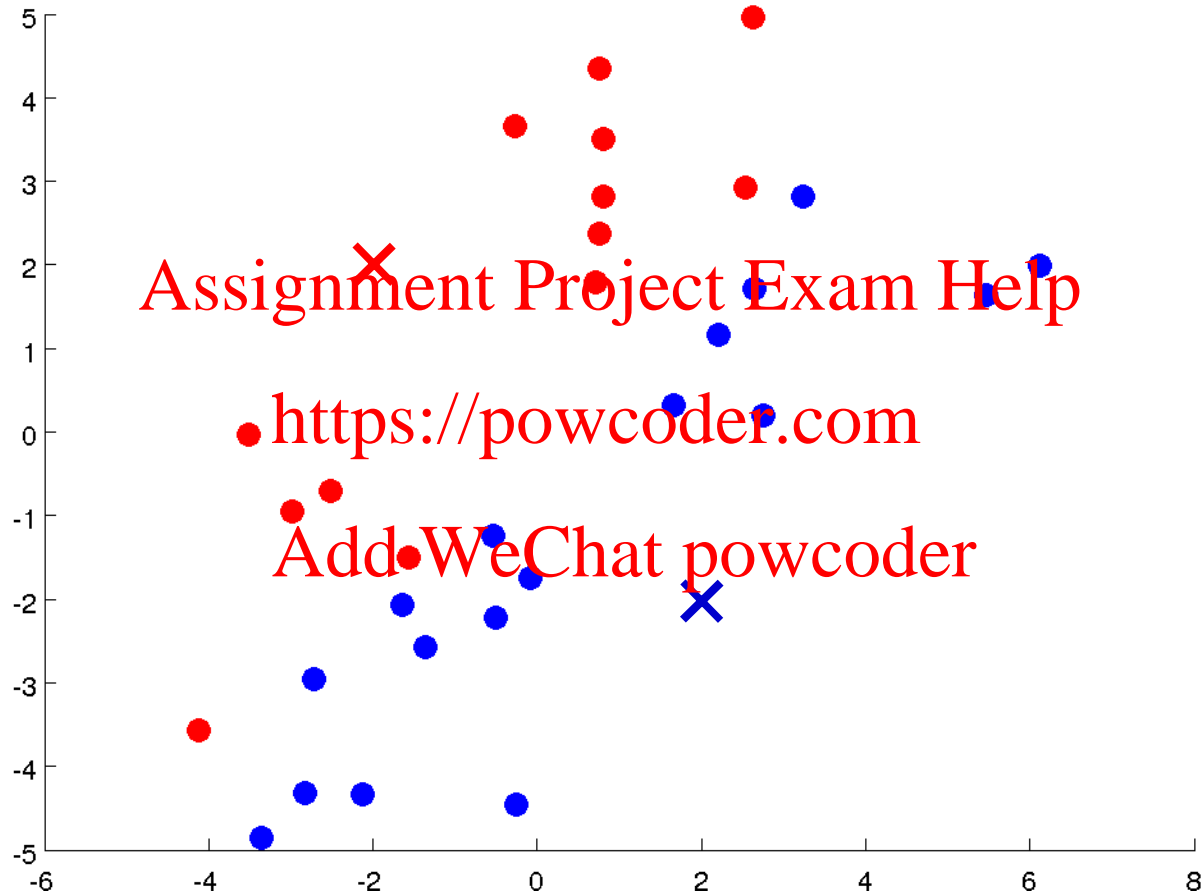
Assignment Project Exam Help

Add WeChat powcoder



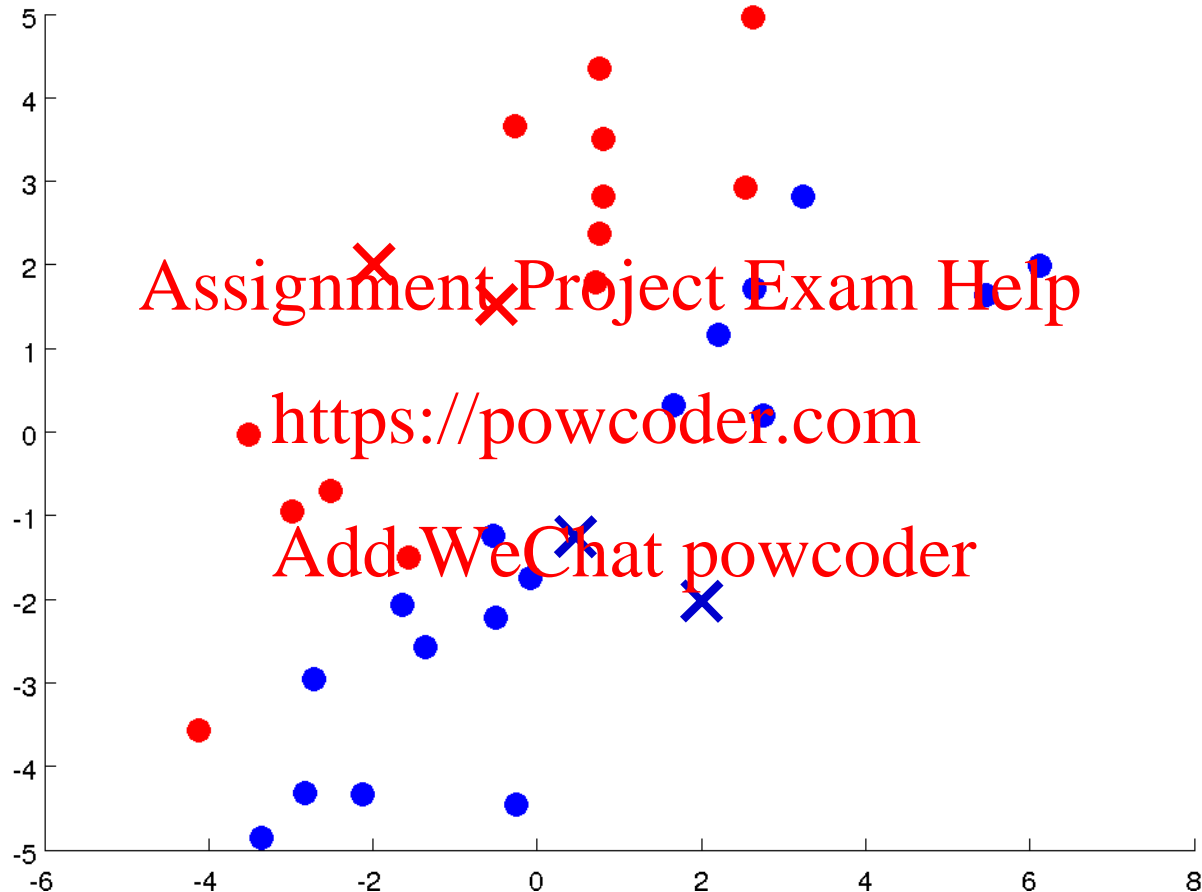
Assignment Project Exam Help

Add WeChat powcoder



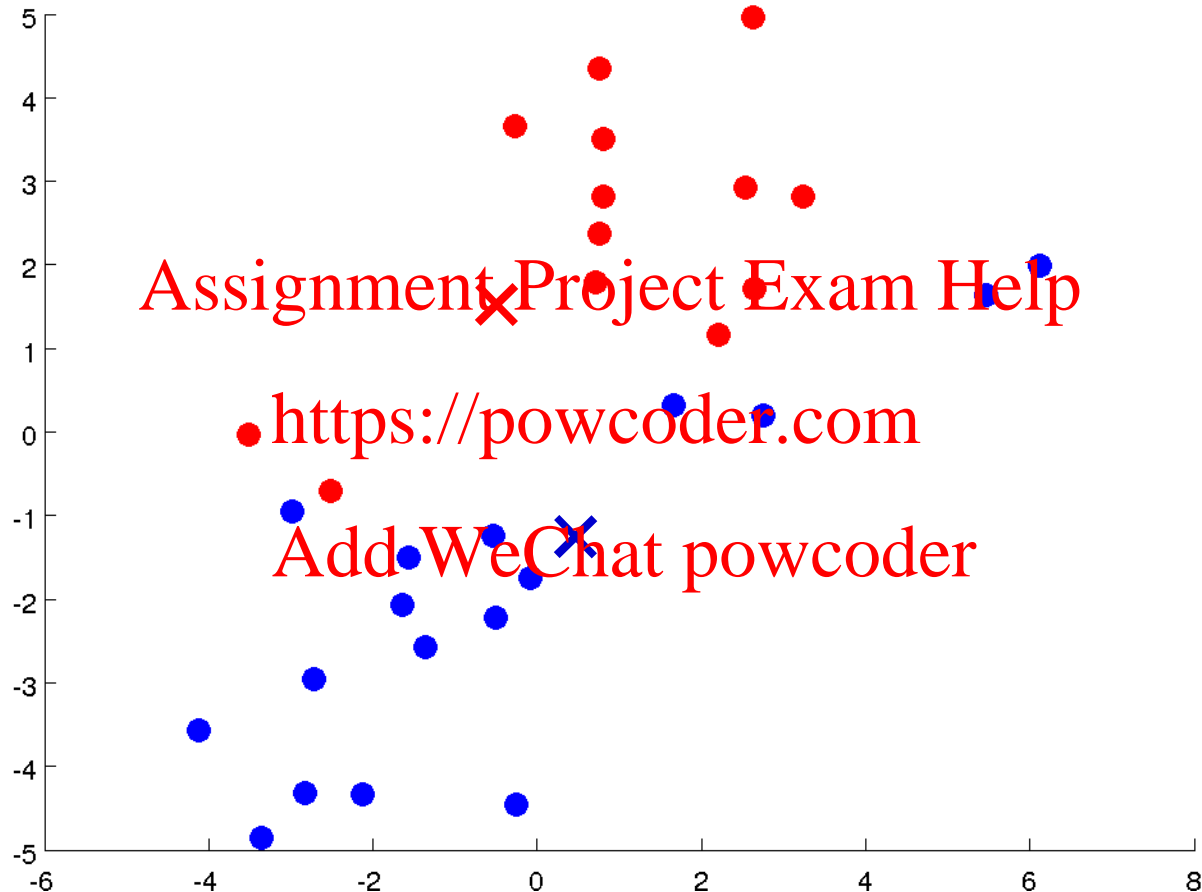
Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help

Next Class

Add WeChat powcoder

Unsupervised Learning I: Clustering:

clustering, k-means, Gaussian mixtures.

Assignment Project Exam Help

<https://powcoder.com>

Reading: Bishop 9.1-9.2

Add WeChat powcoder

Assignment Project Exam Help

PSet 2 Out

Add WeChat powcoder

- Due in 1 week: 9/24 11:59pm GMT -5 (Boston Time)

Assignment Project Exam Help

- Regression, <https://powcoder.com> gradient descent

Add WeChat powcoder