

Assignment Project Exam Help

Announcements

Add WeChat powcoder

Reminder: pset5 self-grading form and pset6 out, due Today 11/19 11:59pm Boston Time

Assignment Project Exam Help

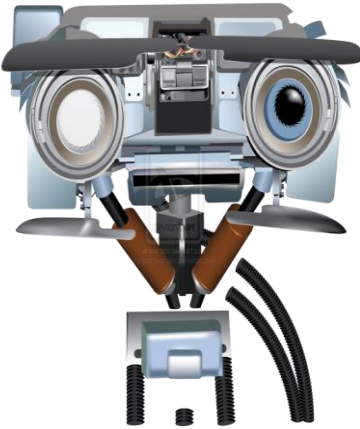
- Class challenge out Today (will discuss in class)

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help
Semi-Supervised
Learning
<https://powcoder.com>

Add WeChat powcoder

Slides credit: Jerry Zhu, Aarti Singh

Assignment Project Exam Help

Supervised Learning

Add WeChat powcoder

Feature Space \mathcal{X}

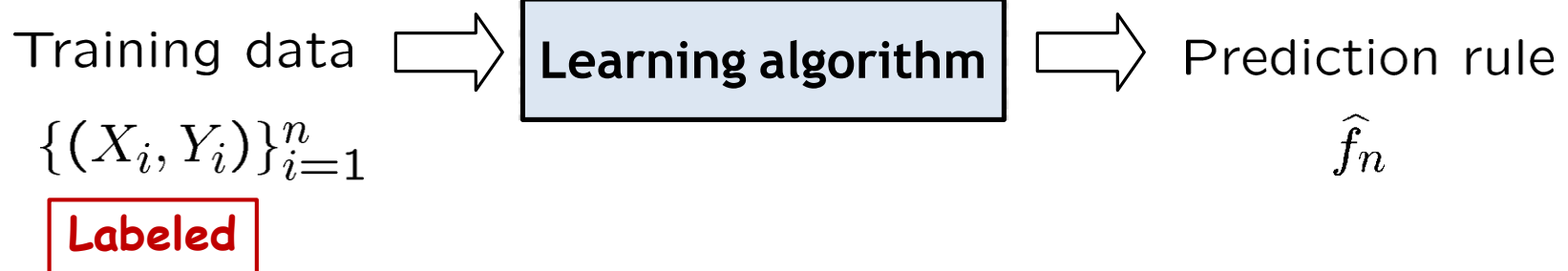
Label Space \mathcal{Y}

Goal: Construct a **predictor** $f : \mathcal{X} \rightarrow \mathcal{Y}$ to minimize

$$R(f) \equiv \mathbb{E}_{XY} [\text{loss}(Y, f(X))]$$

<https://powcoder.com>

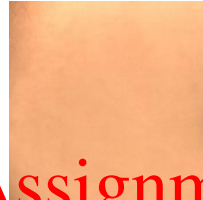
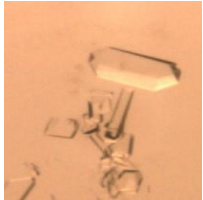
Optimal predictor (Bayes Rule) depends on unknown P_{XY} , so instead
learn a good prediction rule from training data $\{(X_i, Y_i)\}_{i=1}^n \stackrel{\text{iid}}{\sim} P_{XY}(\text{unknown})$



Assignment Project Exam Help

Labeled and Unlabeled data

Add WeChat powcoder



“Crystal” “Needle” “Empty”

Assignment Project Exam Help

0 1 2 3 4 5 6 7 8
8 9 0 1 2 3 4 5 6 7

<https://powcoder.com>



“0” “1” “2” ...

Add WeChat powcoder

Human expert/
Special equipment/
Experiment

“Sports”
“News”
“Science”
...



Unlabeled data, X_i

Cheap and abundant !

Labeled data, Y_i

Expensive and scarce !

Assignment Project Exam Help

Free-of-cost labels?

Add WeChat powcoder

Luis von Ahn: Games with a purpose (ReCaptcha)

Assignment Project Exam Help

Email address

Password

<https://powcoder.com>

Add WeChat powcoder

STELLA UDON

Type the two words:

reCAPTCHA™ stop spam. read books.

Log In

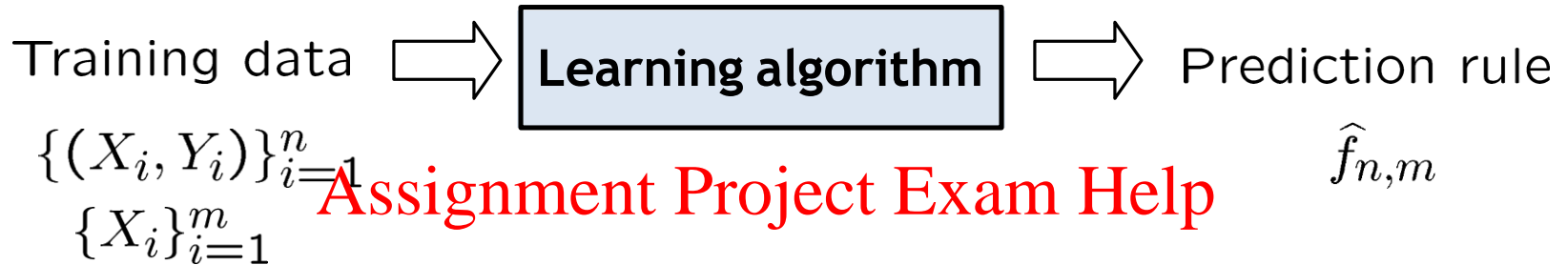
Word challenging to OCR
(Optical Character Recognition)

You provide a free label!

Assignment Project Exam Help

Semi-Supervised learning

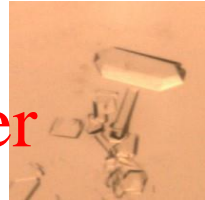
Add WeChat powcoder



Supervised learning (SL)

<https://powcoder.com>

Labeled data $\{X_i, Y_i\}_{i=1}^n$



X_i Y_i

“Crystal”

Semi-Supervised learning (SSL)

Labeled data $\{X_i, Y_i\}_{i=1}^n$ **and** Unlabeled data $\{X_i\}_{i=1}^m$

$$m \gg n$$

Goal: Learn a better prediction rule than based on labeled data alone.

Assignment Project Exam Help

Semi-Supervised learning in Humans

Cognitive science

Computational model of how humans learn from labeled and unlabeled data.

<https://powcoder.com>

- concept learning in children: x =animal, y =concept (e.g., dog)
- Daddy points to a brown animal and says "dog!"
- Children also observe animals by themselves

Assignment Project Exam Help

Can unlabeled data help?

Add WeChat powcoder

- Positive labeled data
- Negative labeled data
- Unlabeled data

Assignment Project Exam Help

Supervised Decision Boundary

<https://powcoder.com>

Semi-Supervised Decision Boundary

Add WeChat powcoder



Assume each class is a coherent group (e.g. Gaussian)

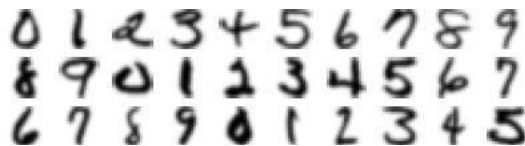
Then unlabeled data can help identify the boundary more accurately.

Assignment Project Exam Help

Can unlabeled data help?

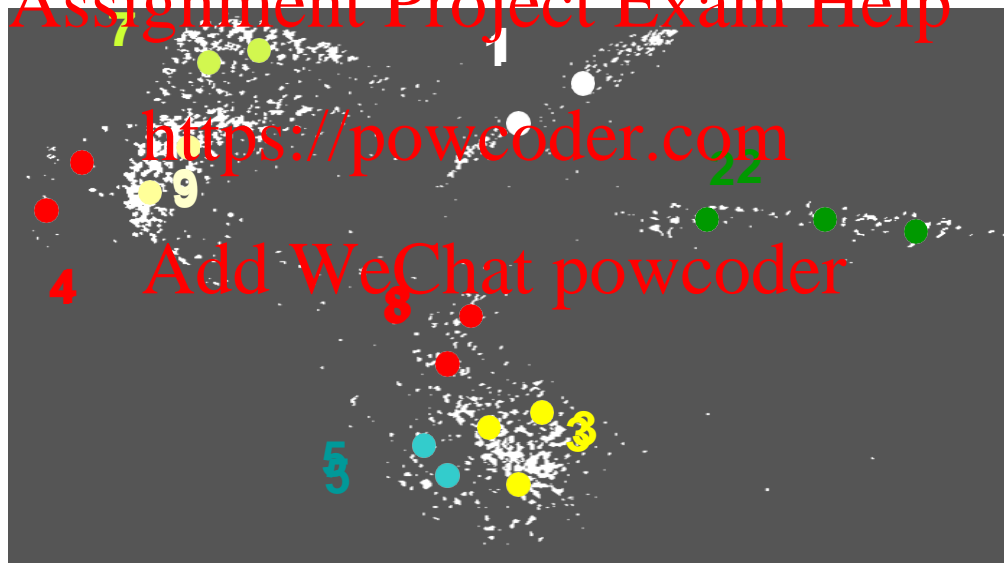
Add WeChat powcoder

Unlabeled Images



Labels "0" "1" "2" ...

Assignment Project Exam Help



<https://powcoder.com>

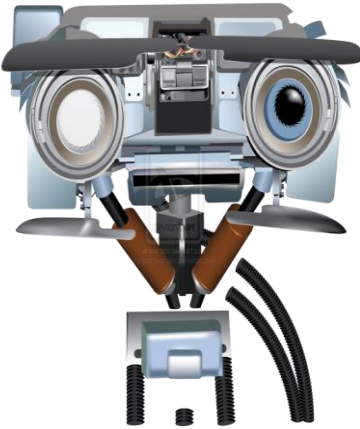
Add WeChat powcoder

This embedding can be done by manifold learning algorithms, e.g. tSNE

“Similar” data points have “similar” labels

Assignment Project Exam Help

Add WeChat powcoder



Assignment Project Exam Help

Algorithms

<https://powcoder.com>

Add WeChat powcoder

Semi-Supervised Learning

Slides credit: Jerry Zhu, Aarti Singh

Assignment Project Exam Help

Some SSL Algorithms

Add WeChat powcoder

- Self-Training
 - Generative methods, mixture models
 - Graph-based methods
 - Co-Training
 - Semi-supervised SVM
 - Many others
- Assignment Project Exam Help
- <https://powcoder.com>
- Add WeChat powcoder

Assignment Project Exam Help

Notation

Add WeChat powcoder

- instance \mathbf{x} , label y
- learner $f : \mathcal{X} \mapsto \mathcal{Y}$
- labeled data $(X_l, Y_l) = \{(x_{1:l}, y_{1:l})\}$
- unlabeled data $X_u = \{\mathbf{x}_{l+1:l+u}\}$, available during training. Usually $l \ll u$. Let $n = l + u$
- test data $\{(x_{n+1...}, y_{n+1...})\}$, not available during training

Assignment Project Exam Help

Self-training

Add WeChat powcoder

Our first SSL algorithm:

Input: labeled data $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$, unlabeled data $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$.

1. Initially, let $L = \{(\mathbf{x}_i, y_i)\}_{i=1}^l$ and $U = \{\mathbf{x}_j\}_{j=l+1}^{l+u}$.
2. Repeat:
 - 3. Train f from L using supervised learning.
 - 4. Apply f to the unlabeled instances in U .
 - 5. Remove a subset S from U ; add $\{(\mathbf{x}, f(\mathbf{x})) | \mathbf{x} \in S\}$ to L .

Self-training is a *wrapper* method

- the choice of learner for f in step 3 is left completely open
- good for many real world tasks like natural language processing
- but mistake by f can reinforce itself

Assignment Project Exam Help

Self-training Example

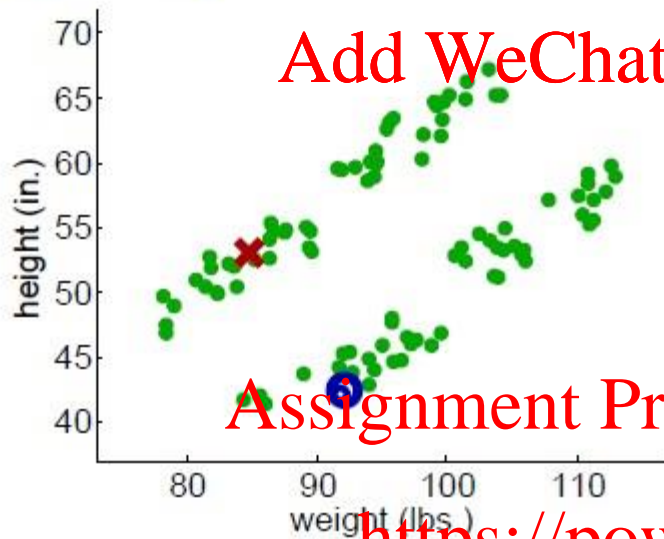
Add WeChat powcoder

Propagating 1-NN

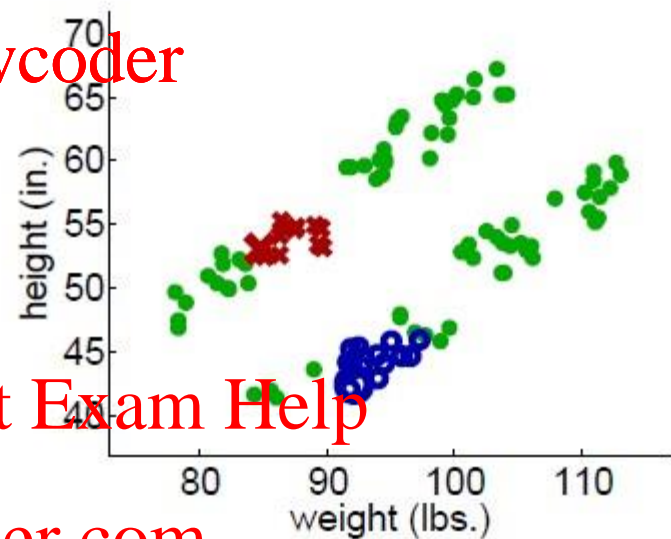
Input: labeled data $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$, unlabeled data $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$,
distance function $d()$.

1. Initially, let $L = \{(\mathbf{x}_i, y_i)\}_{i=1}^l$ and $U = \{\mathbf{x}_j\}_{j=l+1}^{l+u}$.
2. Repeat until U is empty:
3. Select $\mathbf{x} = \operatorname{argmin}_{\mathbf{x} \in U} \min_{\mathbf{x}' \in L} d(\mathbf{x}, \mathbf{x}')$.
4. Set $f(\mathbf{x})$ to the label of \mathbf{x} 's nearest instance in L .
 Break ties randomly.
5. Remove \mathbf{x} from U ; add $(\mathbf{x}, f(\mathbf{x}))$ to L .

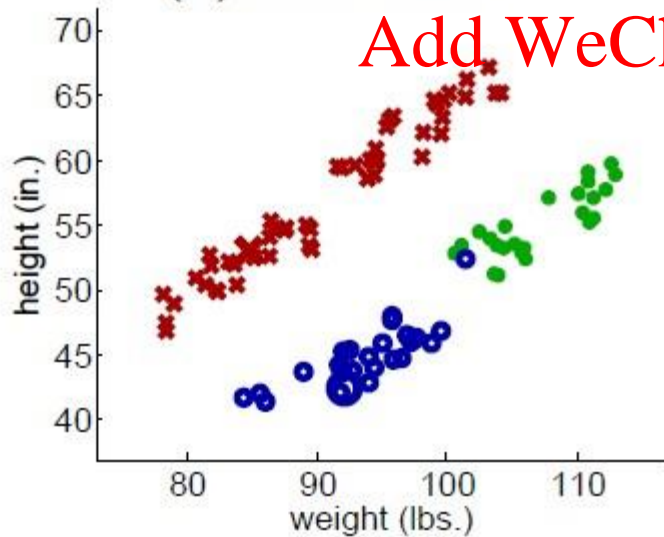
Propagating 1-Nearest-Neighbor: now it works



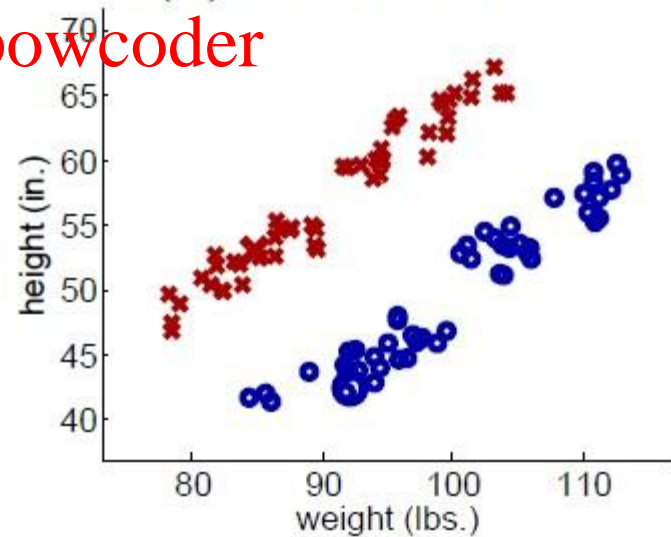
(a) Iteration 1



(b) Iteration 25



(c) Iteration 74

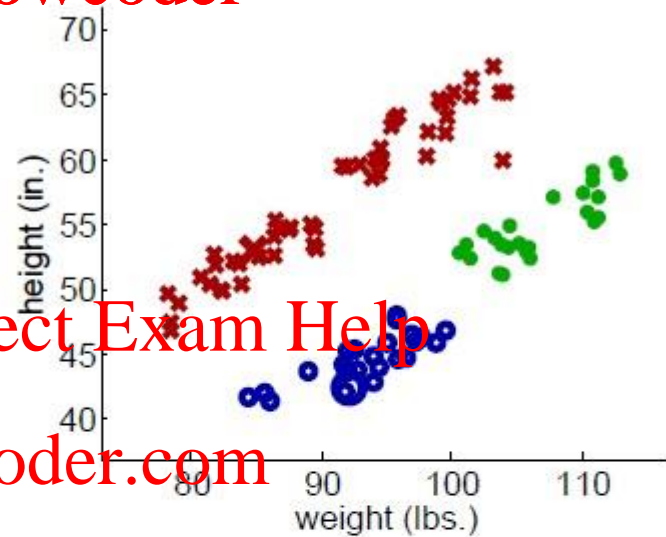
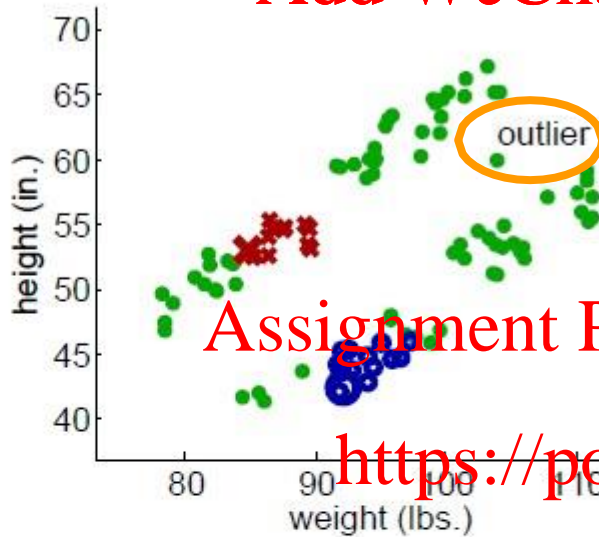


(d) Final labeling of all instances

Propagating 1-Nearest-Neighbor: now it doesn't

But with a single outlier.

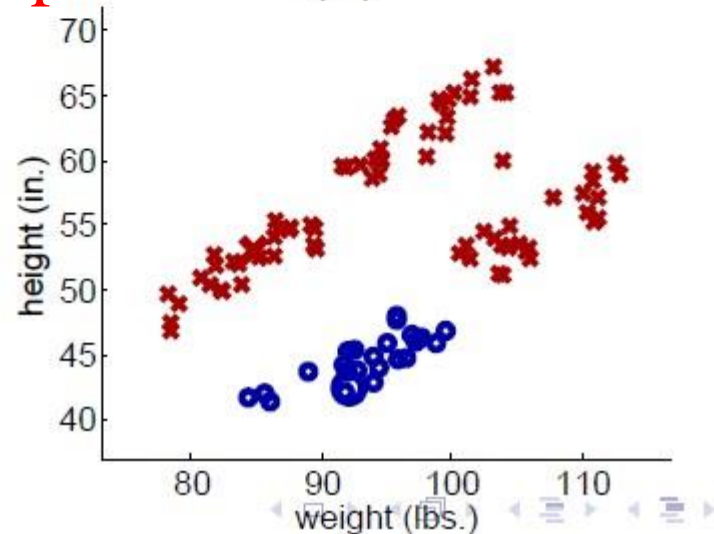
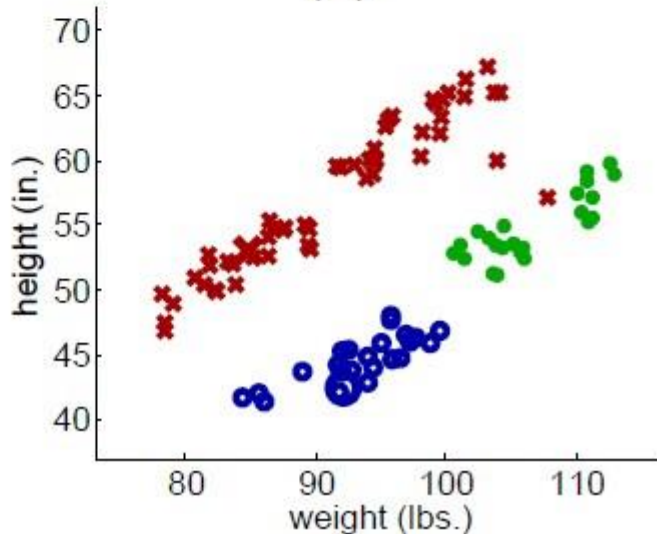
Add WeChat powcoder



Assignment Project Exam Help

<https://powcoder.com>

(a) Add WeChat powcoder (b)



Assignment Project Exam Help

Related: Cluster and Label

Add WeChat powcoder

Input: $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l), \mathbf{x}_{l+1}, \dots, \mathbf{x}_{l+u}$,
a clustering algorithm \mathcal{A} , a supervised learning algorithm \mathcal{L}

1. Cluster $\mathbf{x}_1, \dots, \mathbf{x}_{l+u}$ using \mathcal{A} .
2. For each cluster, let S be the labeled instances in it:
3. Learn a supervised predictor from S : $f_S = \mathcal{L}(S)$.
4. Apply f_S to all unlabeled instances in this cluster.

Output: labels on unlabeled data y_{l+1}, \dots, y_{l+u} .

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

But again: **SSL sensitive to assumptions**—in this case, that the clusters coincide with decision boundaries. If this assumption is incorrect, the results can be poor.

Assignment Project Exam Help

Co-training

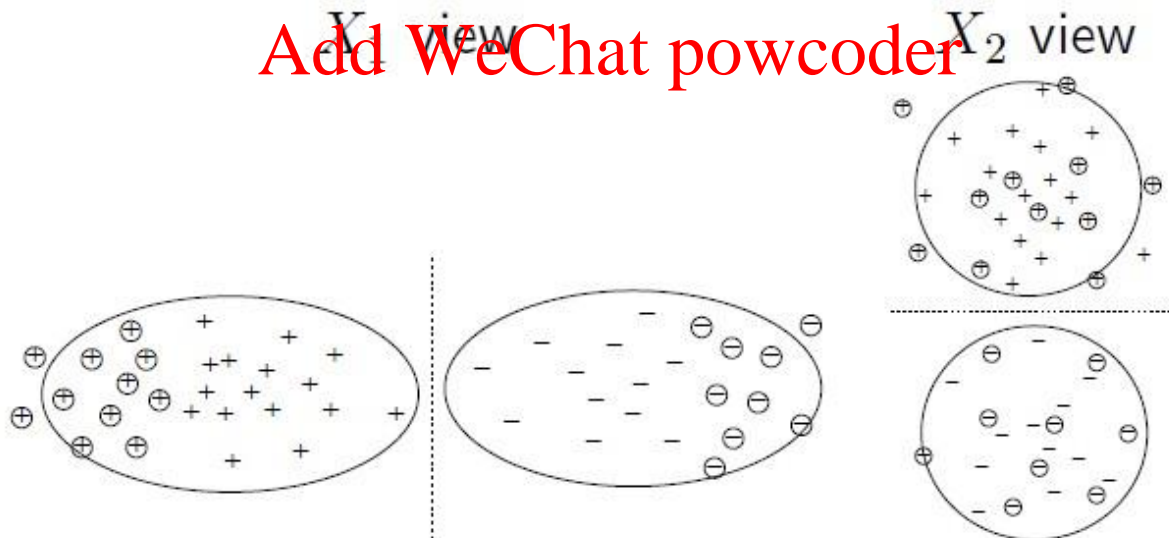
Add WeChat powcoder

Assumptions

- feature split $x = [x^{(1)}; x^{(2)}]$ exists
- $x^{(1)}$ or $x^{(2)}$ alone is sufficient to train a good classifier

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Co-training Algorithm

Co-training (Blum & Mitchell, 1998) (Mitchell, 1999) assumes that

- (i) features can be split into two sets;
- (ii) each sub-feature set is sufficient to train a good classifier.

Assignment Project Exam Help

- Initially two separate classifiers are trained with the labeled data, on the two sub-feature sets respectively.
- Each classifier then classifies the unlabeled data, and 'teaches' the other classifier with the few unlabeled examples (and the predicted labels) they feel most confident.
- Each classifier is retrained with the additional training examples given by the other classifier, and the process repeats.

Assignment Project Exam Help

Co-training Algorithm

Add WeChat powcoder

Blum & Mitchell'98

Input: labeled data $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$, unlabeled data $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$
each instance has two views $\mathbf{x}_i = [\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}]$,
and a learning speed k .

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

1. let $L_1 = L_2 = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$.
2. Repeat until unlabeled data is used up:
3. Train view-1 $f^{(1)}$ from L_1 , view-2 $f^{(2)}$ from L_2 .
4. Classify unlabeled data with $f^{(1)}$ and $f^{(2)}$ separately.
5. Add $f^{(1)}$'s top k most-confident predictions $(\mathbf{x}, f^{(1)}(\mathbf{x}))$ to L_2 .
 Add $f^{(2)}$'s top k most-confident predictions $(\mathbf{x}, f^{(2)}(\mathbf{x}))$ to L_1 .
 Remove these from the unlabeled data.

Assignment Project Exam Help

Semi-Supervised Learning

Add WeChat powcoder

- Generative methods

- Graph-based methods

Assignment Project Exam Help

- Co-Training

<https://powcoder.com>

- Semi-Supervised SVMs

Add WeChat powcoder

- Many other methods

SSL algorithms can use unlabeled data to help improve prediction accuracy if data satisfies appropriate assumptions

Assignment Project Exam Help

Next Class

Add WeChat powcoder

Practical Advice for Applying ML

Machine learning system design; feature engineering; feature pre-processing, learning with large datasets; SGD and mini-batch GD

<https://powcoder.com>

Add WeChat powcoder