

Announcements

Reminder: pset5 self-grading form and pset6 out, due Thursday 11/19 11:59pm A Boston Time
Assignment Project Exam Help

<https://powcoder.com>

- Class challenge out Thursday (will discuss in class)
- Lab this week: Q-Learning and GANs

Add WeChat at powcoder



Generative Adversarial Assignment Project Exam Help Networks (GANs)

<https://powcoder.com>

Add WeChat powcoder

Unsupervised Learning IV

Today

- Supervised vs Unsupervised Learning (recap)
- Density Models
- Generative Adversarial Networks (GANs)
- Cycle-GANs

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Supervised vs Assignment Project Exam Help Unsupervised Learning

<https://powcoder.com>

Add WeChat powcoder
Generative Adversarial Networks (GANs)

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Assignment Project Exam Help

<https://powcoder.com>

Goal: Learn a *function* to map $x \rightarrow y$

Add WeChat powcoder

Examples: Classification,
regression, object detection,
semantic segmentation, image
captioning, etc.

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map $x \rightarrow y$

Examples: Classification,
regression, object detection,
semantic segmentation, image
captioning, etc.



→ Cat

Classification

[This image is CC0 public domain](#)

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map $x \rightarrow y$

Examples: Classification,
regression, object detection,
semantic segmentation, image
captioning, etc.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



DOG, DOG, CAT

Object Detection

This image is CC0 public domain

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map $x \rightarrow y$

Examples: Classification, regression, object detection, semantic segmentation, image captioning, etc.



GRASS, CAT,
TREE, SKY

Semantic Segmentation

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Supervised vs Unsupervised Learning

Unsupervised Learning

Data: x

Just data, no labels!

Assignment Project Exam Help

<https://powcoder.com>

Goal: Learn some underlying hidden *structure* of the data

Add WeChat powcoder

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.

Supervised vs Unsupervised Learning

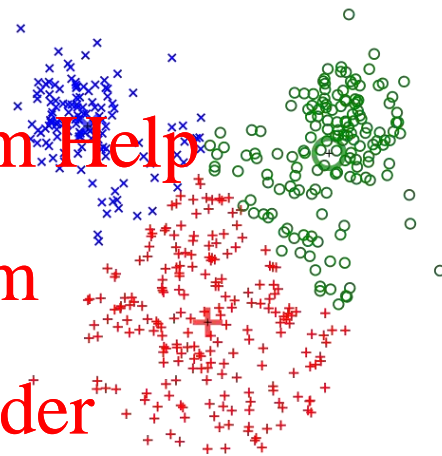
Unsupervised Learning

Data: x

Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.



K-means clustering

This image is CC0 public domain

Supervised vs Unsupervised Learning

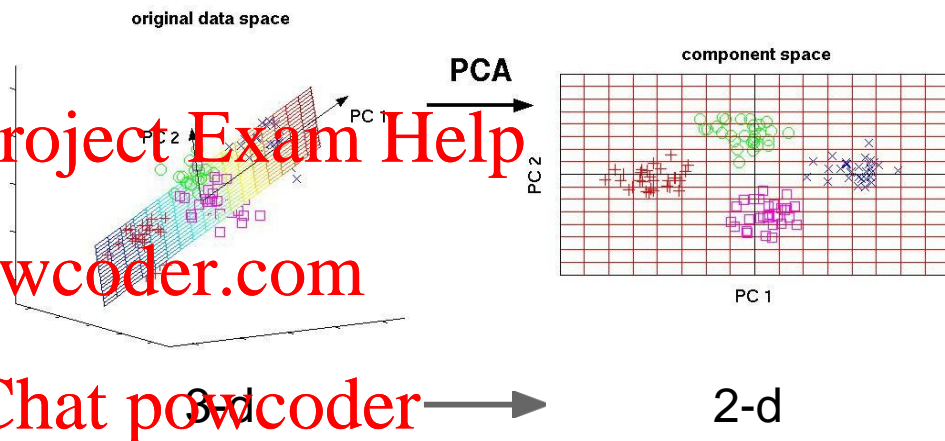
Unsupervised Learning

Data: x

Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data.

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.



Principal Component Analysis
(Dimensionality reduction)

This image from Matthias Scholz
is CC0 public domain

Supervised vs Unsupervised Learning

Unsupervised Learning

Data: x

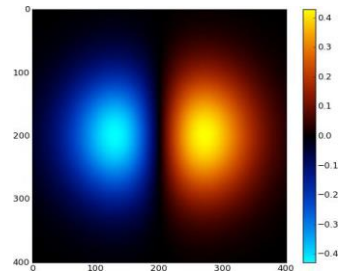
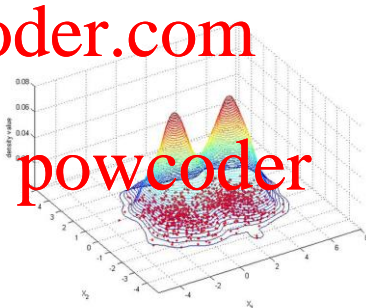
Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data.

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.



1-d density estimation



2-d density estimation

2-d density images [left](#) and [right](#) are [CC0 public domain](#)

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map $x \rightarrow y$

Examples: Classification, regression, object detection, semantic segmentation, image captioning, etc.

Unsupervised Learning

Data: x

Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Supervised vs Unsupervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map $x \rightarrow y$

Examples: Classification, regression, object detection, semantic segmentation, image captioning, etc.

Unsupervised Learning

Training data is cheap

Data: x

Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.

Holy grail: Solve unsupervised learning
 \Rightarrow understand structure of the data



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Generative Models

Given training data, generate new samples from same distribution



Training data $\sim p_{\text{data}}(x)$

Generated samples $\sim p_{\text{model}}(x)$

Want to learn $p_{\text{model}}(x)$ similar to $p_{\text{data}}(x)$

Add WeChat powcoder

Generative Models

Given training data, generate new samples from same distribution



Training data $\sim p_{\text{data}}(x)$

Generated samples $\sim p_{\text{model}}(x)$

Want to learn $p_{\text{model}}(x)$ similar to $p_{\text{data}}(x)$

Add WeChat powcoder

Addresses density estimation, a core problem in unsupervised learning

Several flavors:

- Explicit density estimation: explicitly define and solve for $p_{\text{model}}(x)$
- Implicit density estimation: learn model that can sample from $p_{\text{model}}(x)$ w/o explicitly defining it

Why Generative Models?

- Realistic samples for artwork, super-resolution, colorization, etc.



- Generative models of time-series data can be used for simulation and planning (reinforcement learning applications!)
- Training generative models can also enable inference of latent representations that can be useful as general features

Figures from L-R are copyright: (1) [Alec Radford et al. 2016](#); (2) [David Berthelot et al. 2017](#); [Phillip Isola et al. 2017](#). Reproduced with authors' permission.

Taxonomy of Generative Models

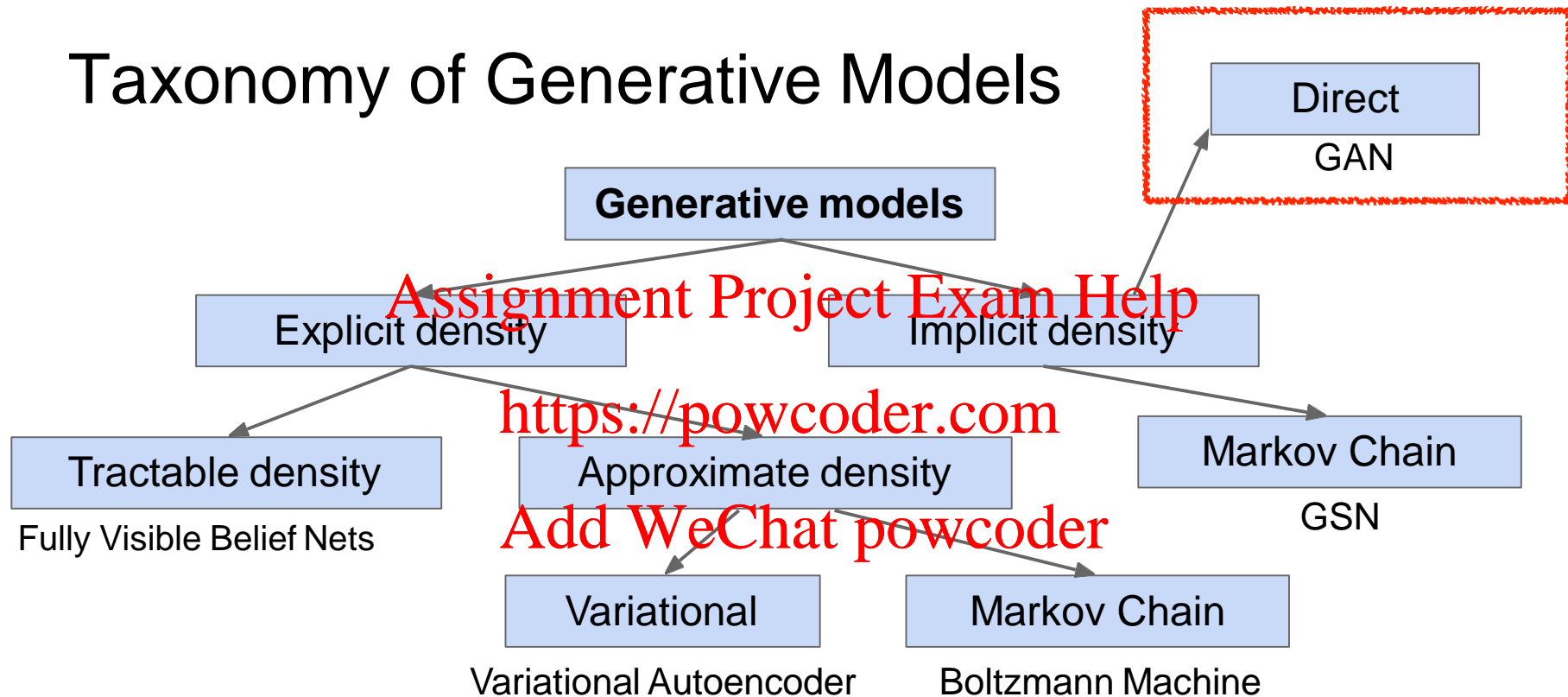
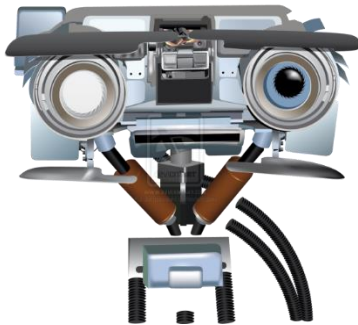


Figure copyright and adapted from Ian Goodfellow, Tutorial on Generative Adversarial Networks, 2017.



Assignment Project Exam Help Generative Adversarial Networks

<https://powcoder.com>

Add WeChat powcoder (GANs)

Overview

Four modern approaches to generative modeling.

- Generative adversarial networks (today)
- Reversible architectures
- Autoregressive models
- Variational autoencoders

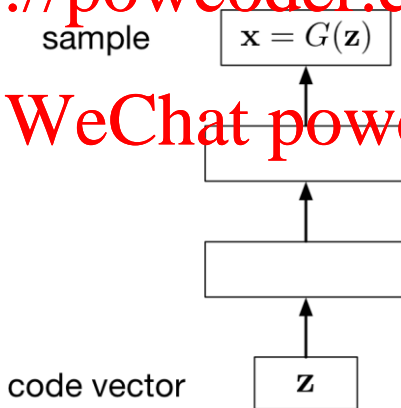
All four approaches have different pros and cons.

Implicit Generative Models

- Implicit generative models implicitly define a probability distribution
- Start by sampling the code vector \mathbf{z} from a fixed, simple distribution (e.g. spherical Gaussian)
- The generator network computes a differentiable function G mapping \mathbf{z} to an \mathbf{x} in data space

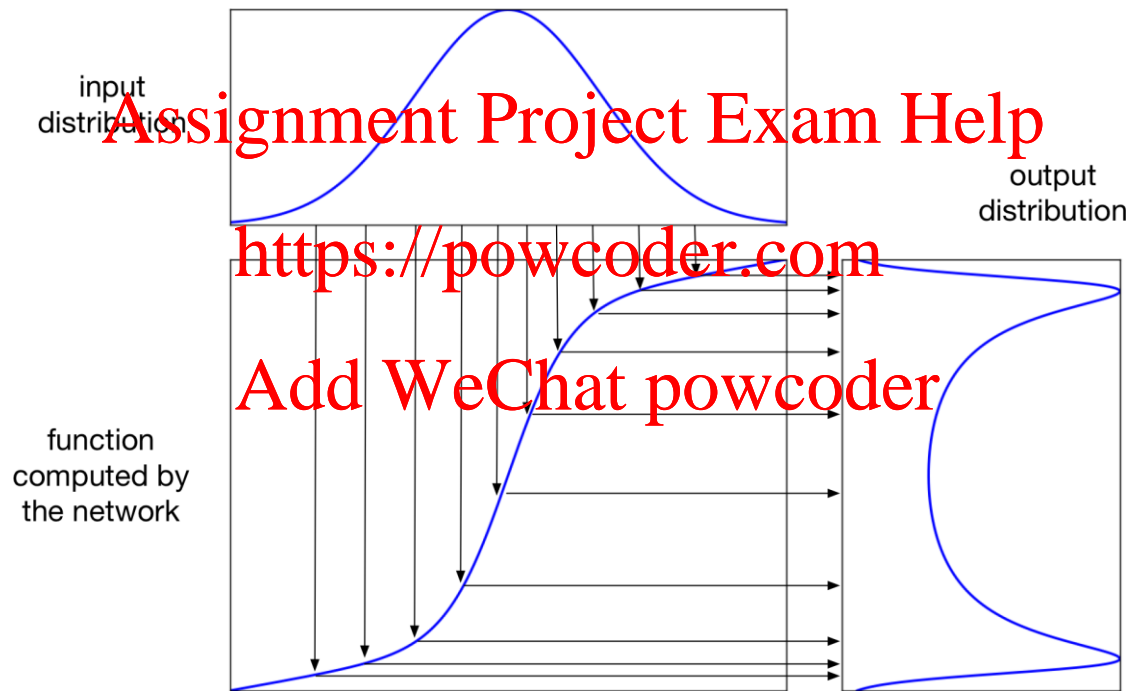
<https://powcoder.com>

Add WeChat powcoder

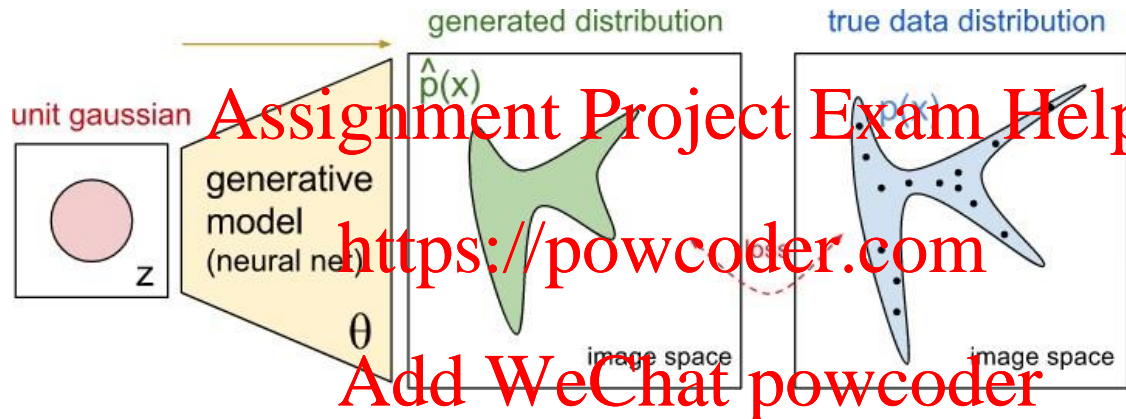


Implicit Generative Models

A 1-dimensional example:

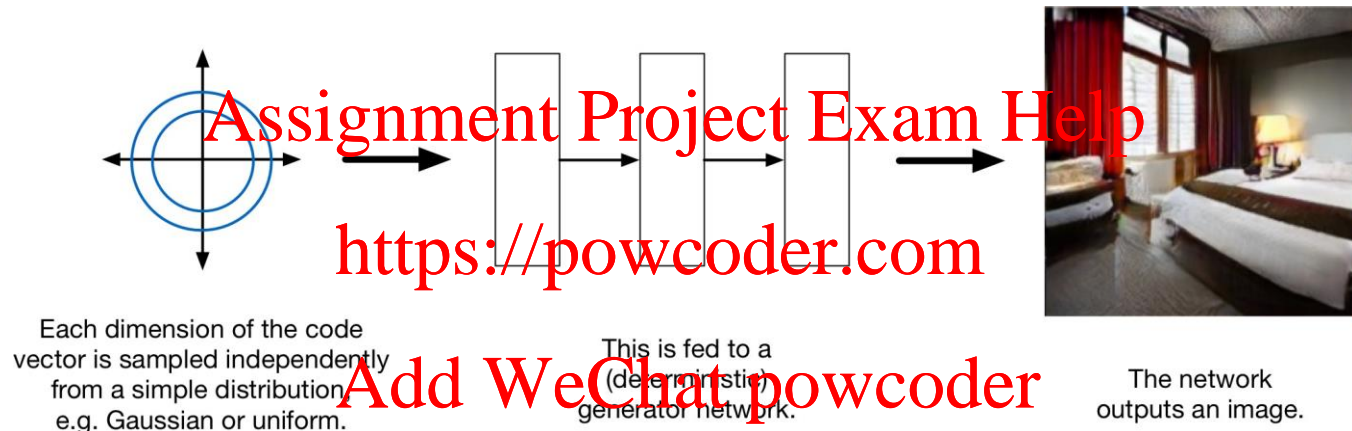


Implicit Generative Models



<https://blog.openai.com/generative-models/>

Implicit Generative Models

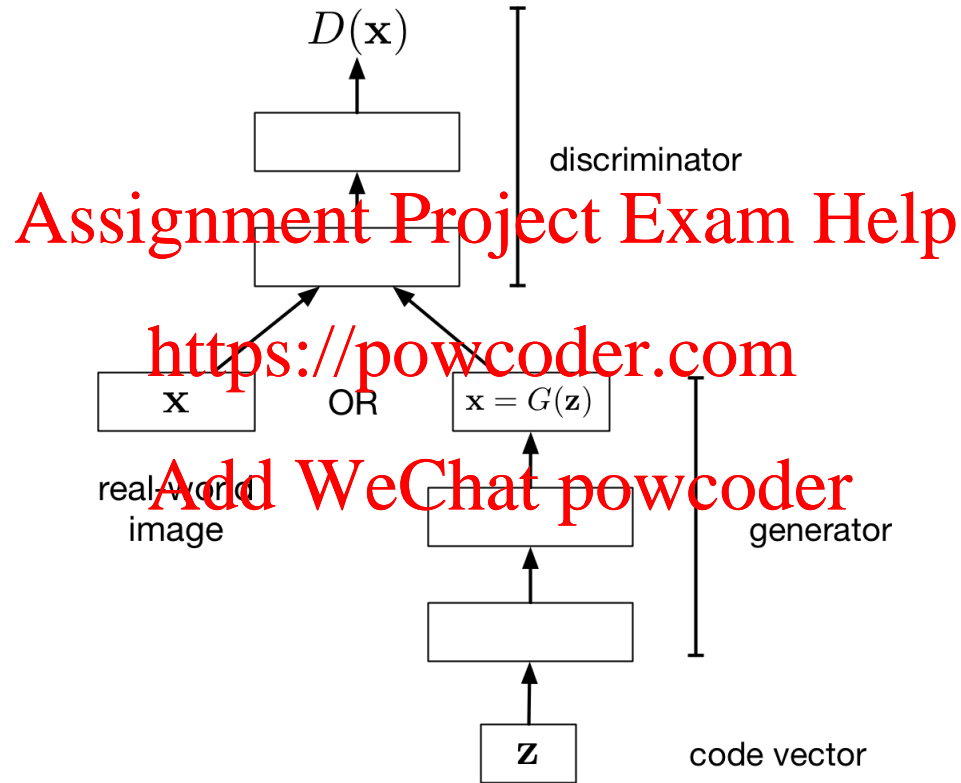


This sort of architecture sounded preposterous to many of us, but amazingly, it works.

Generative Adversarial Networks

- The advantage of implicit generative models: if you have some criterion for evaluating the quality of samples, then you can compute its gradient with respect to the network parameters, and update the network's parameters to make the sample a little better
- The idea behind Generative Adversarial Networks (GANs): train two different networks
 - The generator network tries to produce realistic-looking samples
 - The discriminator network tries to figure out whether an image came from the training set or the generator network
- The generator network tries to fool the discriminator network

Generative Adversarial Networks



Generative Adversarial Networks

- Let D denote the discriminator's predicted probability of being data
- Discriminator's cost function: cross-entropy loss for task of classifying real vs. fake images

Assignment Project Exam Help

$$J_D = \mathbb{E}_{x \sim D}[-\log D(x)] + \mathbb{E}_z[-\log(1 - D(G(z)))]$$

- One possible cost function for the generator: the opposite of the discriminator's

Add WeChat powcoder

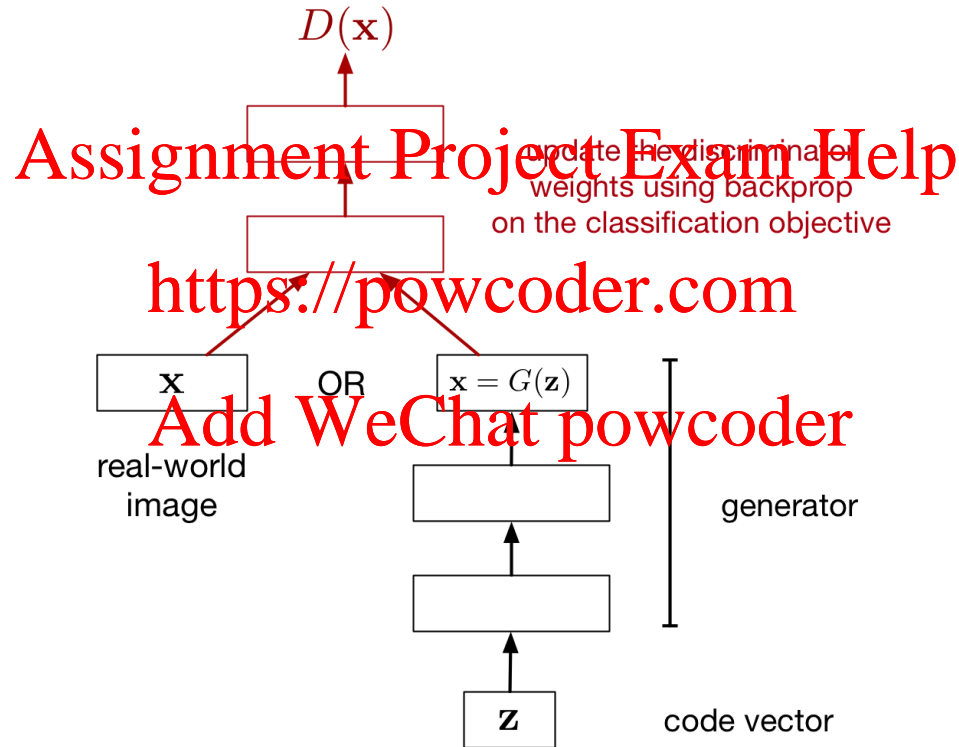
$$J_G = -J_D$$
$$= \text{const} + \mathbb{E}_z[\log(1 - D(G(z)))]$$

- This is called the **minimax formulation**, since the generator and discriminator are playing a **zero-sum game** against each other:

$$\max_G \min_D J_D$$

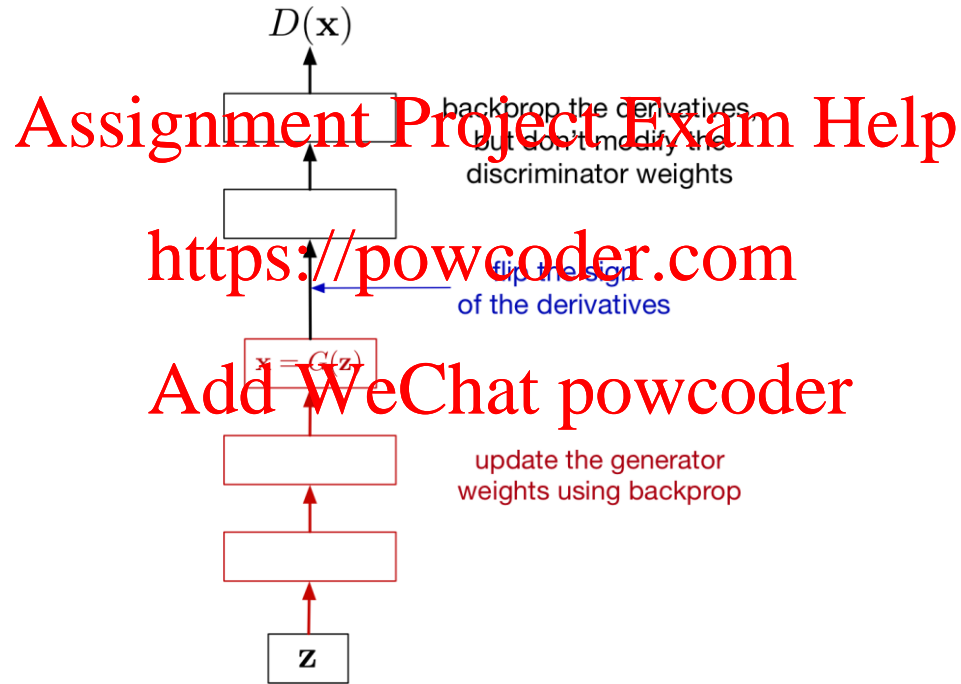
Generative Adversarial Networks

Updating the discriminator:



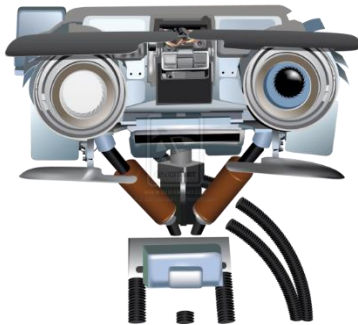
Generative Adversarial Networks

Updating the generator:



Generative Adversarial Networks

- Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder
- Since GANs were introduced in 2014, there have been hundreds of papers introducing various architectures and training methods.
 - Most modern architectures are based on the Deep Convolutional GAN (DC-GAN), where the generator and discriminator are both conv nets.
 - GAN Zoo: <https://github.com/hindupuravinash/the-gan-zoo>
 - Good source of horrible puns (VEEGAN, Checkhov GAN, etc.)



Assignment Project Exam Help
GANs: Application to
Image Generation
<https://powcoder.com>

Add WeChat powcoder
Generative Adversarial Networks (GANs)

Generative Adversarial Nets

Generated samples

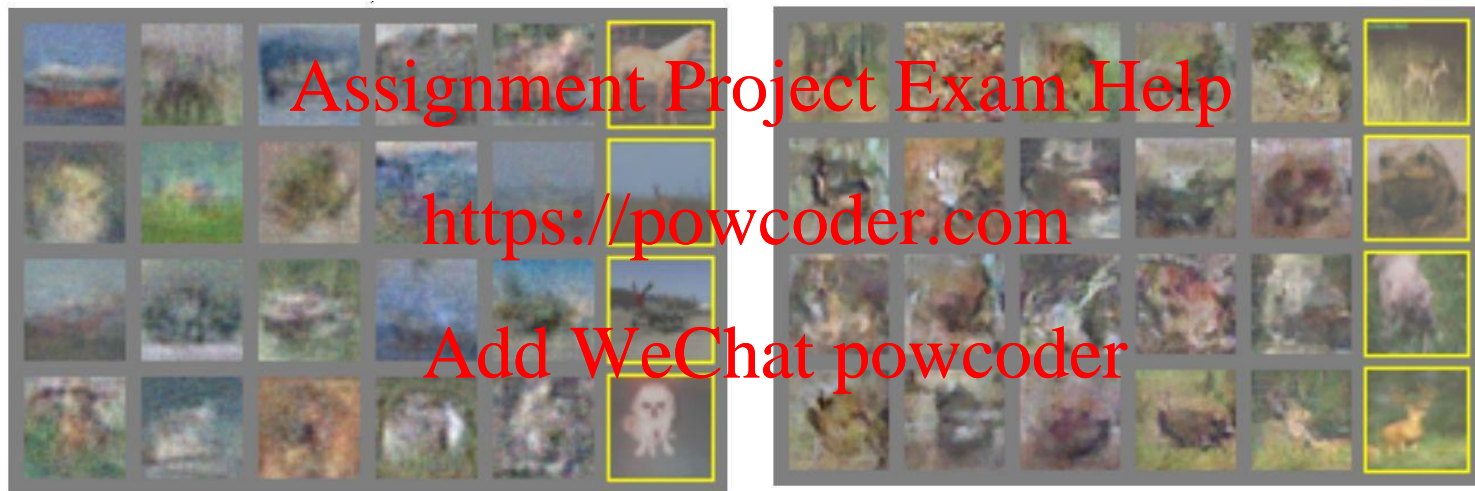


Nearest neighbor from training set

Figures copyright Ian Goodfellow et al., 2014. Reproduced with permission.

Generative Adversarial Nets

Generated samples (CIFAR-10)



Nearest neighbor from training set

Figures copyright Ian Goodfellow et al., 2014. Reproduced with permission.

GAN Samples

Celebrities:



Karras et al., 2017. Progressive growing of GANs for improved quality, stability, and variation

GAN Samples

Bedrooms:



Karras et al., 2017. Progressive growing of GANs for improved quality, stability, and variation

GAN Samples

Objects:



Karras et al., 2017. Progressive growing of GANs for improved quality, stability, and variation

GAN Samples

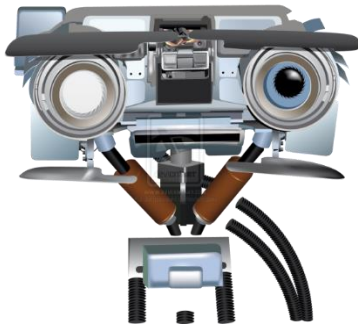
GANs revolutionized generative modeling by producing crisp, high-resolution images.

The catch: we don't know how well they're modeling the distribution.

Can't measure the log-likelihood they assign to held-out data. Could they be memorizing training examples? (E.g., maybe they sometimes produce photos of real celebrities?)

We have no way to tell if they are dropping important modes from the distribution.

See Wu et al., "On the quantitative analysis of decoder-based generative models" for partial answers to these questions.



Assignment Project Exam Help

Cycle GANs

<https://powcoder.com>

Add WeChat powcoder

Generative Adversarial Networks (GANs)

CycleGAN

Style transfer problem: change the style of an image while preserving the content.

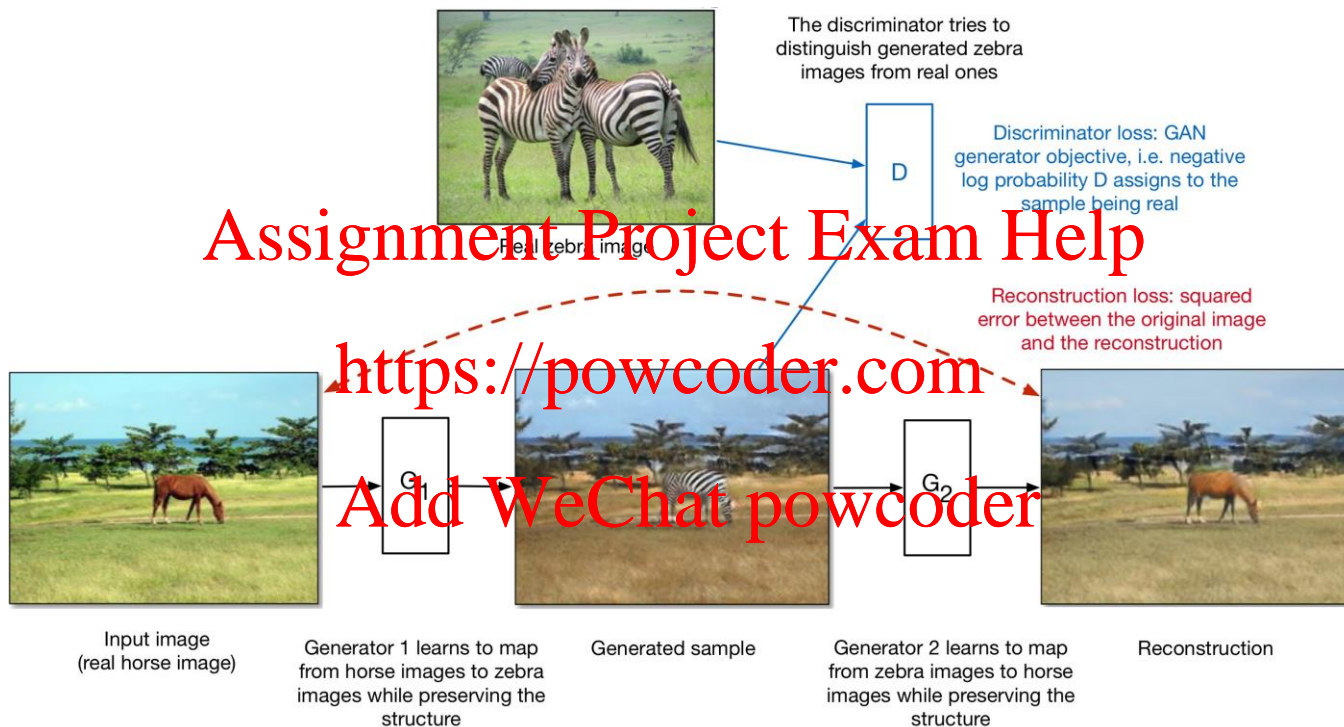


Data: Two unrelated collections of images, one for each style

CycleGAN

- If we had paired data (same content in both styles), this would be a supervised learning problem. But this is hard to find.
- The CycleGAN architecture learns to do it from unpaired data.
 - Train two different generator nets to go from style 1 to style 2, and vice versa.
 - Make sure the generated samples of style 2 are indistinguishable from real images by a discriminator net.
 - Make sure the generators are **cycle-consistent**: mapping from style 1 to style 2 and back again should give you almost the original image.

CycleGAN



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Total loss = discriminator loss + reconstruction loss

CycleGAN

Style transfer between aerial photos and maps:

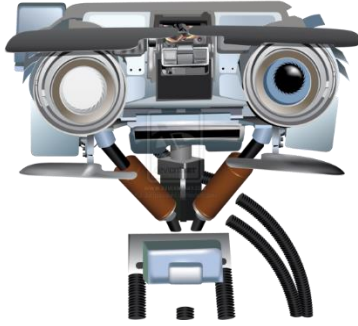


CycleGAN

Style transfer between road scenes and semantic segmentations (labels of every pixel in an image by object category):

Assignment Project Exam Help





Assignment Project Exam Help

GANs: Convolutional Architectures

<https://powcoder.com>

Add WeChat powcoder

Generative Adversarial Networks (GANs)

Generative Adversarial Nets: Convolutional Architectures

Generator is an upsampling network with fractionally-strided convolutions
Discriminator is a convolutional network

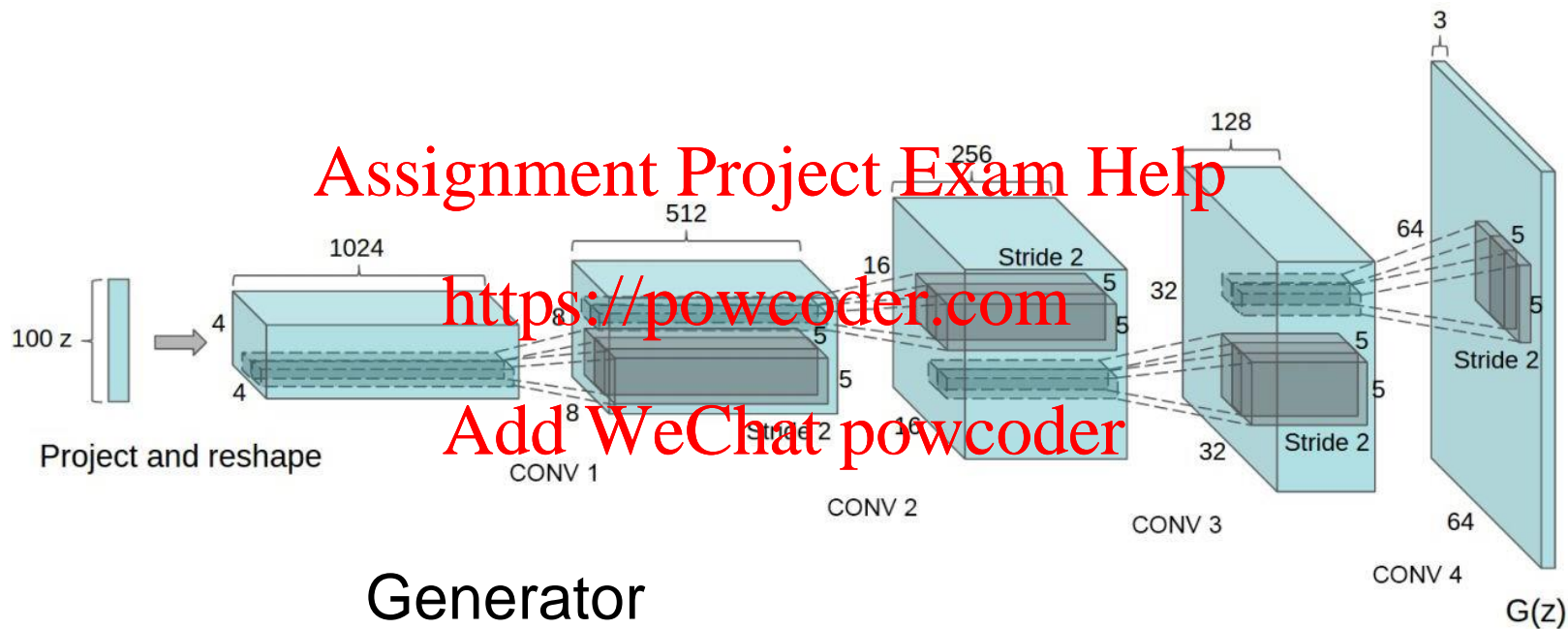
Assignment Project Exam Help

Architecture guidelines for stable Deep Convolutional GANs

- Replace any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator).
- Use batchnorm in both the generator and the discriminator.
- Remove fully connected hidden layers for deeper architectures.
- Use ReLU activation in generator for all layers except for the output, which uses Tanh.
- Use LeakyReLU activation in the discriminator for all layers.

Radford et al, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", ICLR 2016

Generative Adversarial Nets: Convolutional Architectures



Radford et al, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", ICLR 2016

Generative Adversarial Nets: Convolutional Architectures

Samples
from the
model look
amazing!



Radford et al,
ICLR 2016

Generative Adversarial Nets: Convolutional Architectures

Interpolating
between
random
points in latent
space



Radford et al,
ICLR 2016

Generative Adversarial Nets: Interpretable Vector Math

Radford et al, ICLR 2016

Smiling woman Neutral woman Neutral man

Samples
from the
model



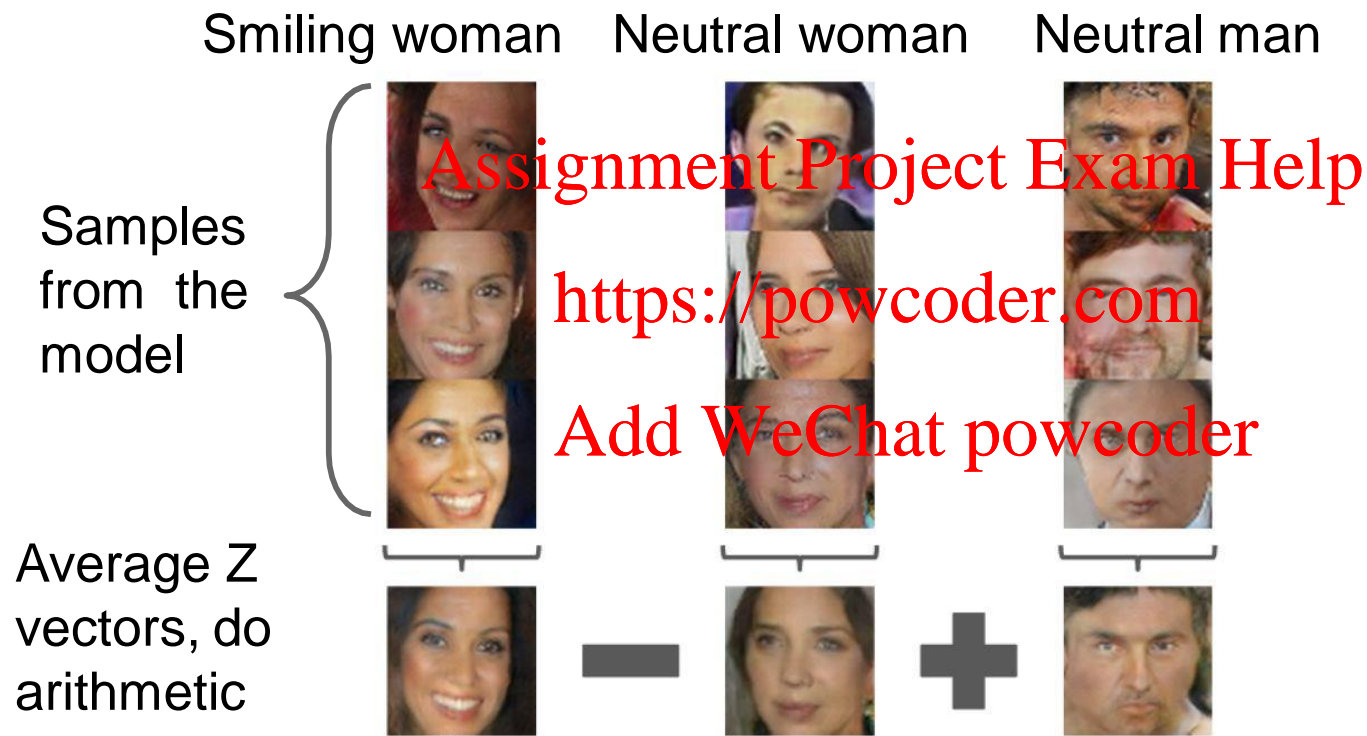
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Generative Adversarial Nets: Interpretable Vector Math

Radford et al, ICLR 2016



Generative Adversarial Nets: Interpretable Vector Math

Radford et al, ICLR 2016

Smiling woman

Neutral woman

Neutral man

Samples from the model

Average Z vectors, do arithmetic

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Smiling Man



Generative Adversarial Nets: Interpretable Vector Math

Glasses man

No glasses man

No glasses woman



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Radford et al,
ICLR 2016

Generative Adversarial Nets: Interpretable Vector Math

Glasses man

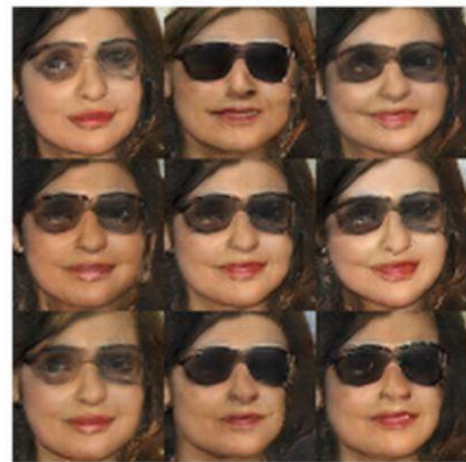
No glasses man

No glasses woman

Radford et al,
ICLR 2016



Woman with glasses



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



-



+



=

2017: Year of the GAN

Better training and generation



(a) Church outdoor.



(b) Dining room.



(c) Kitchen.



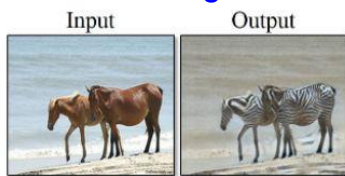
(d) Conference room.

LSGAN. Mao et al. 2017.



BEGAN. Bertholet et al. 2017.

Source->Target domain transfer



horse → zebra



zebra → horse



apple → orange



CycleGAN. Zhu et al. 2017.



→ summer Yosemite



→ winter Yosemite

Text -> Image Synthesis

this small bird has a pink breast and crown, and black primaries and secondaries.

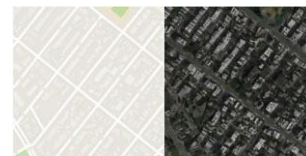


this magnificent fellow is almost all black with a red crest, and white cheek patch.



Reed et al. 2017.

Many GAN applications



Pix2pix. Isola 2017. Many examples at <https://phillipi.github.io/pix2pix/>

Next Class

Unsupervised Learning V: Semi-supervised Learning

Semi-supervised learning (SSL); self-training; co-training;
clustering methods, SSSVM

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder