ECE 590-10/11
COMP ENG ML & DEEP NEURAL NETS

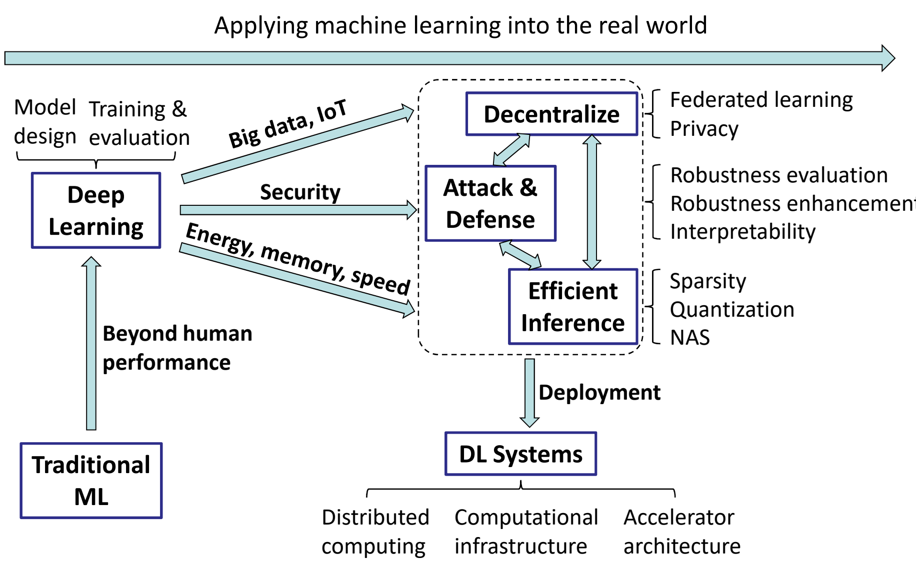
2. LEARNING TYPES

HAI LI & YIRAN CHEN, FALL 2019

1

1

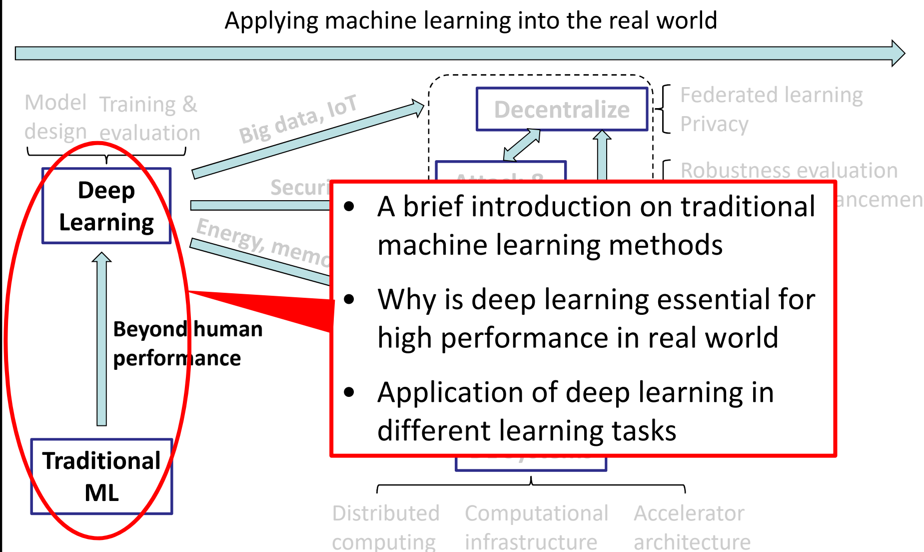
Roadmap of the course



2

2

This lecture

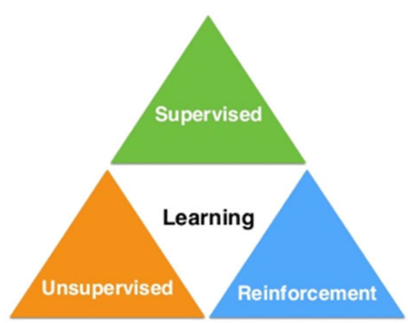


3

3

The overview of learning

- People consider learning as finding a function.
- Supervised learning:
  - f(input) → target (continuous/discrete)
- Unsupervised learning:
  - f(input) → likelihood
- Reinforcement learning:
  - f(state, action) → value



4

4

Supervised learning

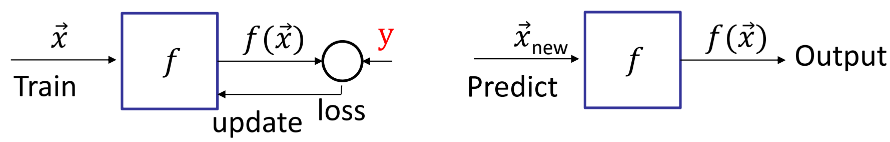
- Function: f(input) → target
  - Continuous value: Regression
  - Discrete value: Classification

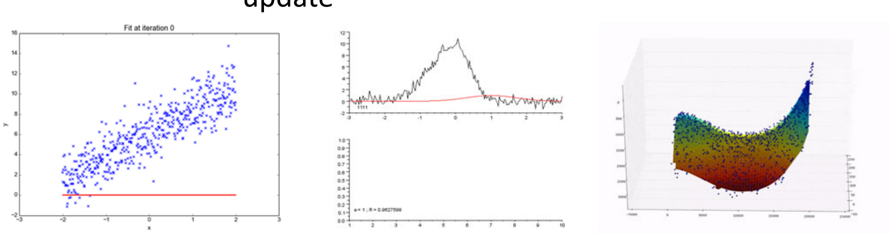
5

5

Regression

- Use data to identify relationships among variables and then apply these relationships to predict.
- Continuous output.





Source of image: https://towardsdatascience.com/linear-regression-the-easier-way-6f941aa471ea; https://commons.wikimedia.org/wiki/File:Regression_pic_assymetrique.gif; http://www.semspirit.com/artificial-intelligence/machine-learning/regression/support-vector-regression/support-vector-regression-in-python/

6

6

Traditional regression methods

- Linear regression

$$\vec{x} = [x_1 \quad x_2 \quad \dots \quad x_n]$$

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \sum_{for\ all\ data} \|\vec{x}\boldsymbol{\omega} - y\|_2^2$$

- Polynomial regression

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ {x_1}^2 & {x_2}^2 & \dots & {x_n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ {x_1}^{p-1} & {x_2}^{p-1} & \dots & {x_n}^{p-1} \end{bmatrix}$$

$$\boldsymbol{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{for\ all\ data} \|\boldsymbol{X}^T\boldsymbol{\beta} - \boldsymbol{y}\|_2^2, (\boldsymbol{\beta} \in \mathbb{R}^p)$$

7

7

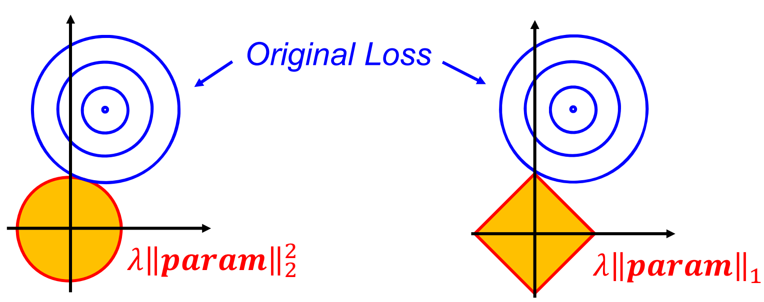
Traditional regression methods

- Ridge regression

Minimize error, while push *parameters* smaller

- Lasso regression

Minimize error, while push *parameters* more sparse



$New\ Loss = Loss + \lambda\|\boldsymbol{param}\|_2^2$ $New\ Loss = Loss + \lambda\|\boldsymbol{param}\|_1$

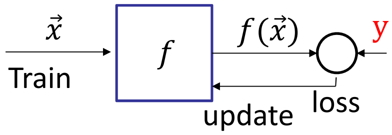
Source of image: http://alex.smola.org/teaching/cmu2013-10-701/slides/13_recitation_lasso.pdf

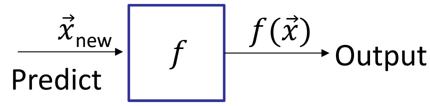
8

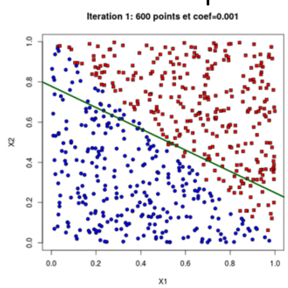
8

Classification

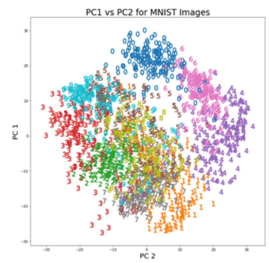
- Similar with regression.
- The only different: Label y & output are discrete.







Linear



Non-Linear

Source of Image: http://www-lisic.univ-littoral.fr/~teytaud/files/Cours/Apprentissage/perceptron/; https://www.mathworks.com/matlabcentral/fileexchange/52003-viewboundary

9

9

Traditional classification methods

- Logistic regression

$$\Pr(Y_i = y | X_i) = p_i^y (1 - p_i)^{1-y} = \frac{e^{\beta \cdot X_i \cdot y}}{1 + e^{\beta \cdot X_i}}$$

- Naive Bayes classifier
- K nearest neighbor (KNN)
  - An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small)
- Learn more in https://www.deeplearningbook.org/

10

10

Unsupervised learning

- Function: f(input) -> likelihood
  - Clustering
  - Generative model

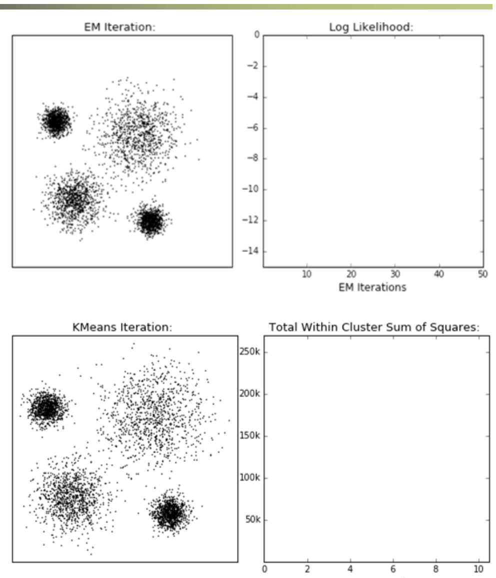
11

11

Clustering

Gaussian mixture model

K-Means



Source of image: https://dashee87.github.io/data%20science/general/Clustering-with-Scikit-with-GIFs/

12

12

Generative Model

- Given training data, generate new samples from same distribution.
- Traditional Method: Naïve Bayes
- Naïve Bayes assumption: features are independent.

$$p(X) = p(x_1, \dots, x_k) = \prod_{k=1}^{K} p(x_k)$$

- The problem is reduced to estimating the parameters' $p(x_k)$ for each feature separately and multiplying these to find the probability for any possible combination. It cannot generate reasonable new images with high-dimensional information.

Fail on pixel cartoon image generation using naïve Bayes

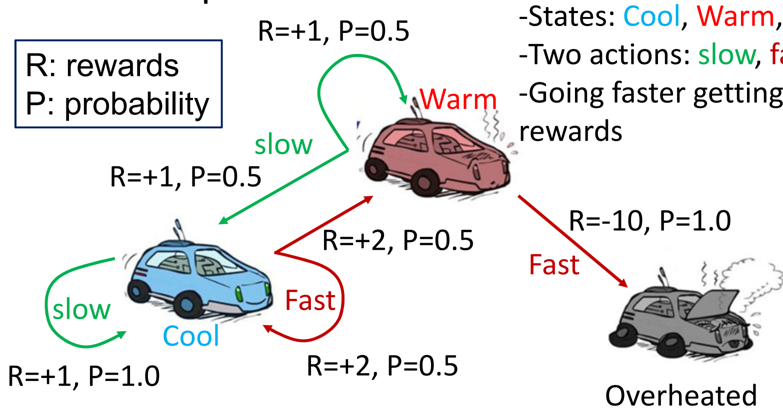
17

17

Reinforcement learning

- Function: f(state, action) -> reward value
- Basic reinforcement is modeled as a Markov decision process.

-Example: car racing
-States: Cool, Warm, Overheated
-Two actions: slow, fast
-Going faster getting double rewards



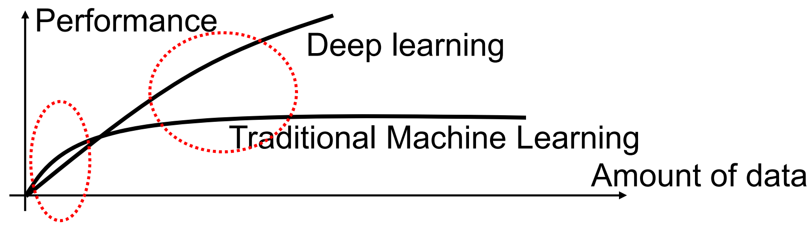
Source of image: https://medium.com/@sanchittanwar75/markov-chains-and-markov-decision-process-e91cda7fa8f2

18

18

Why do we need Deep Learning

- In the real world.
  - Input data is high-dimensional
  - Mapping function is complicated
  - Traditional methods usually fail
- Deep Neural Network is capable of fitting functions with high complexity.



Source of image: https://becominghuman.ai/what-is-deep-learning-and-why-you-need-it-9e2fc0f0e61b

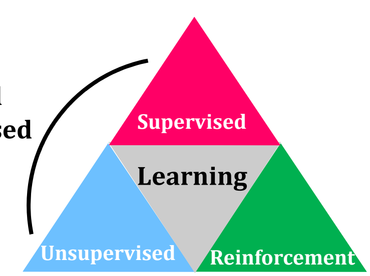
20

20

Learning Paradigms

Labeled data
Direct feedback
Predict outcome/future
Apps: Classification

Semi-supervised
Weakly-supervised



No labels
No feedback
Find hidden representations
Apps: Reconstruction

Decision process
Reward system
Learn series of actions
Apps: Decision-making

http://www.jiqizhixin.com/article/1883

21

21

Supervised Learning

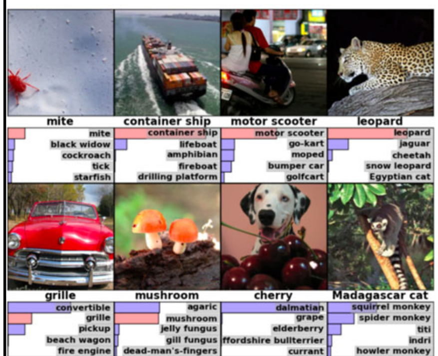
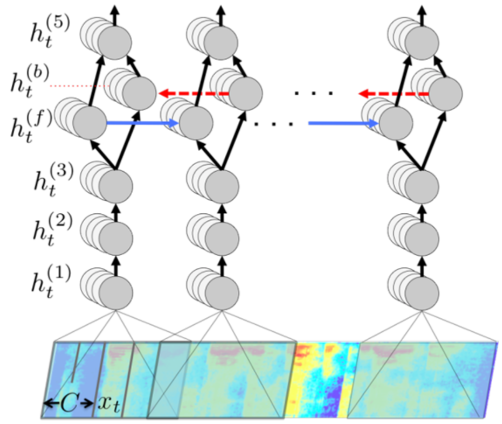


Image Classification



Speech Recognition

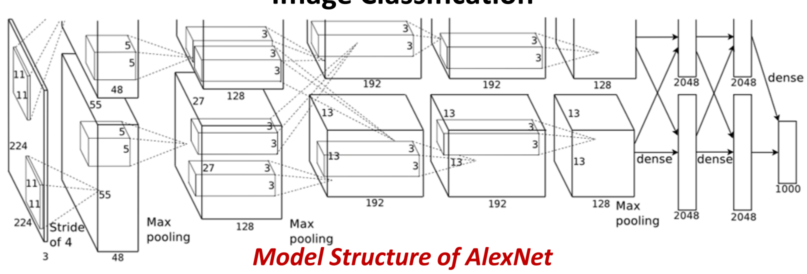
A. Krizhevsky, I. Sutskever and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012.
https://devblogs.nvidia.com/parallelforall/deep-speech-accurate-speech-recognition-gpu-accelerated-deep-learning/

22

22

Convolutional neural network

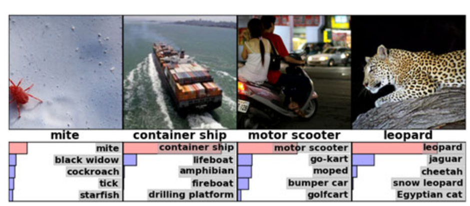
Image Classification



Model Structure of AlexNet

Model	Top-1 Error	Top-5 Error
Sparse coding	47.10%	28.20%
SIFT + FVs	45.70%	25.70%
AlexNet	37.50%	17.00%

Performance Evaluation



Sample Testing Result from ImageNet

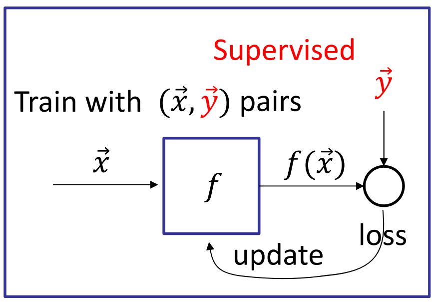
A. Krizhevsky, I. Sutskever and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012.

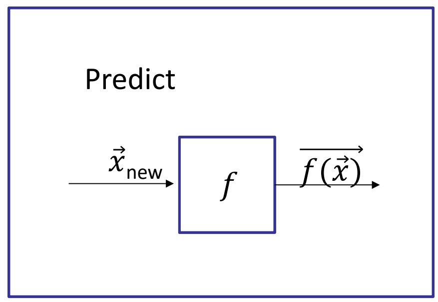
25

25

Structured learning

- Similar with regression and classification.
- The only difference is: Label y & output are structured data.
  - For example: image, sound, text…

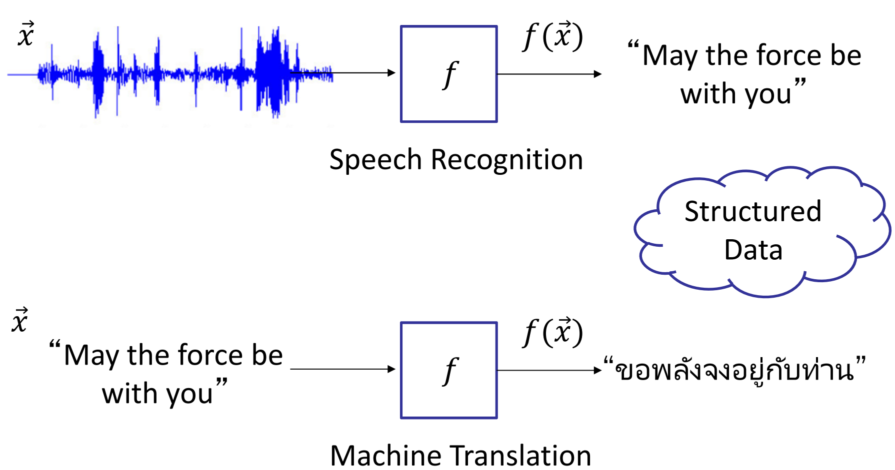




27

27

Structured learning

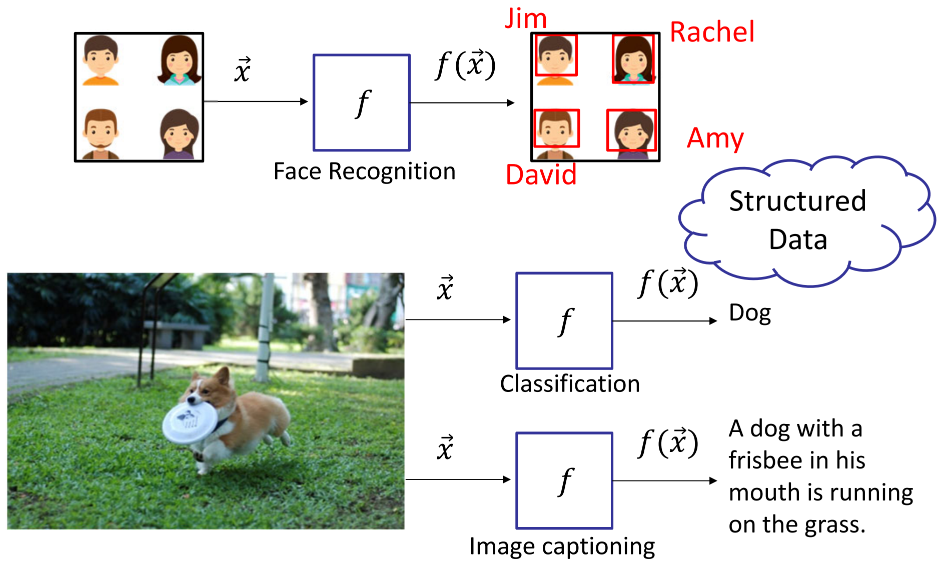


Source of image: http://blog.sina.com.cn/s/blog_4e0987310102v4bn.html

28

28

Structured learning



Source of image: https://wang970727.pixnet.net/blog/post/29313243-%5B%E7%8E%A9%5D-%E9%A3%9B%E7%9B%A4

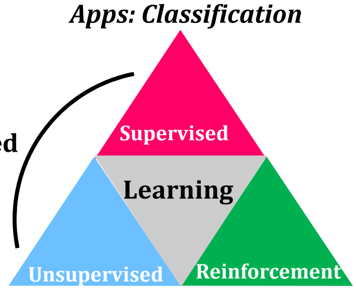
29

29

Learning Paradigms

Labeled data
Direct feedback
Predict outcome/future
Apps: Classification

Semi-supervised
Weakly-supervised



No labels
No feedback
Find hidden representations
Apps: Reconstruction

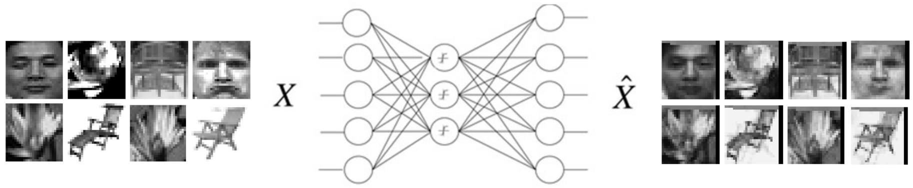
Decision process
Reward system
Learn series of actions
Apps: Decision-making

http://www.jiqizhixin.com/article/1883

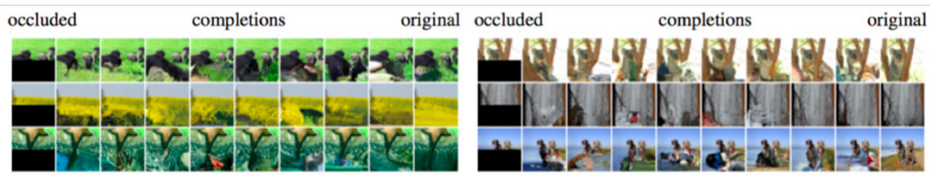
30

30

Unsupervised Learning



Generative Model



Pixel RNN

A.Radford, M. Metz and S. Chintala:Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ICLR*, 2016
A. Oord, N. Kalchbrenner, K. Kavukcuoglu: Pixel Recurrent Neural Networks. arXiv preprint arXiv:1601.06759.

31

31

DL for unsupervised learning

- Generative model: Learning by creating
  - Autoencoder (AE)
  - Variational autoencoder (VAE)
  - Generative adversarial network (GAN)
- Autoregressive: Creating by predicting
  - OpenAI's GPT-2, a huge language model

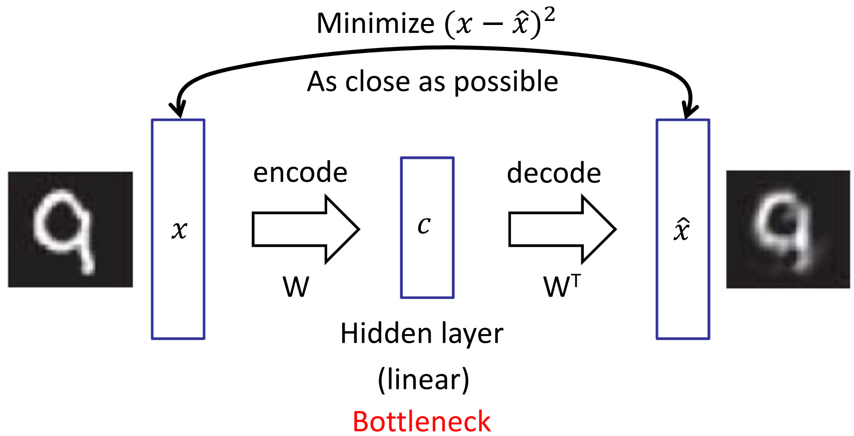
> *what I cannot create, I do not understand.*
> *--- Richard Feynman*

32

32

Autoencoder

- Autoencoder
  - Use Deep Neural Network to Encode and Decode.

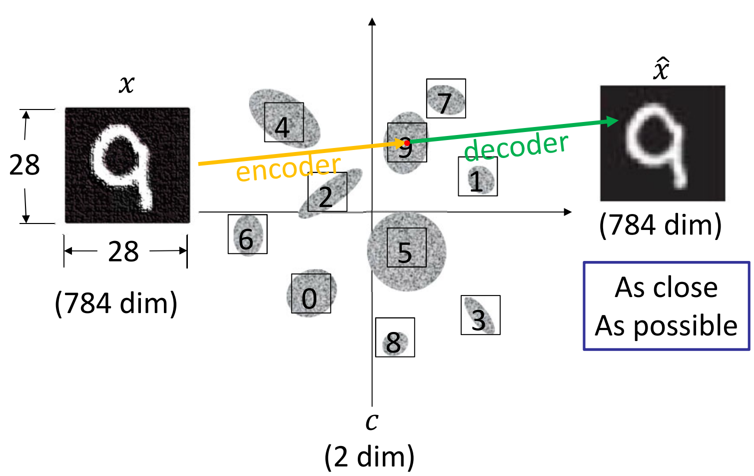


Source of images: https://zhuanlan.zhihu.com/p/25939348

33

33

What Autoencoder can do

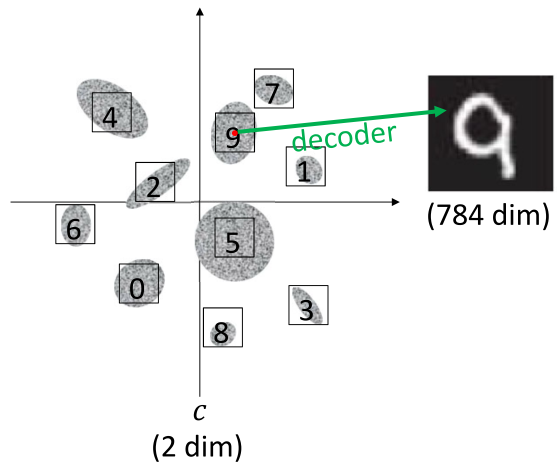


Source of images: https://zhuanlan.zhihu.com/p/25939348

34

34

What Autoencoder can do

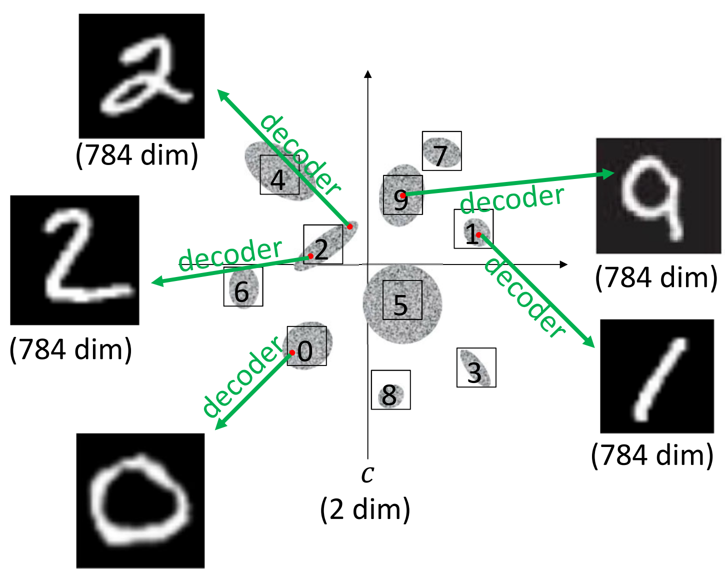


Source of images: https://zhuanlan.zhihu.com/p/25939348

35

35

What Autoencoder can do

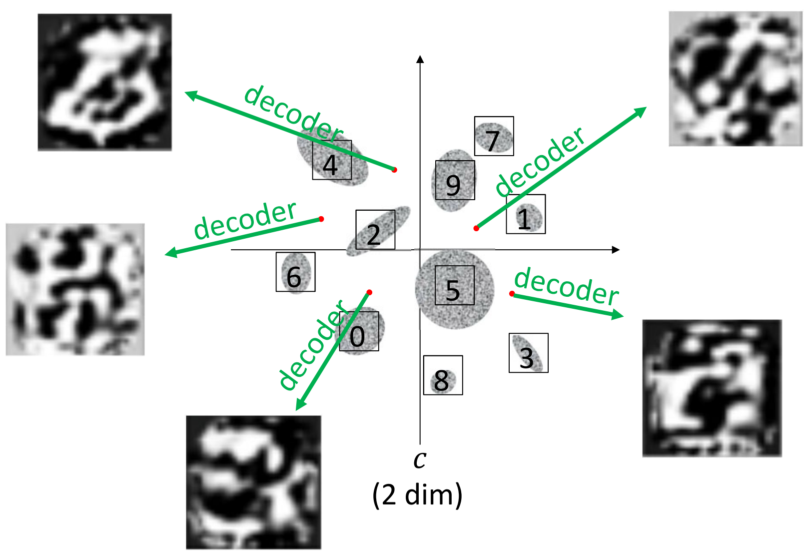


Source of images: https://zhuanlan.zhihu.com/p/25939348

36

36

However, Autoencoder cannot do



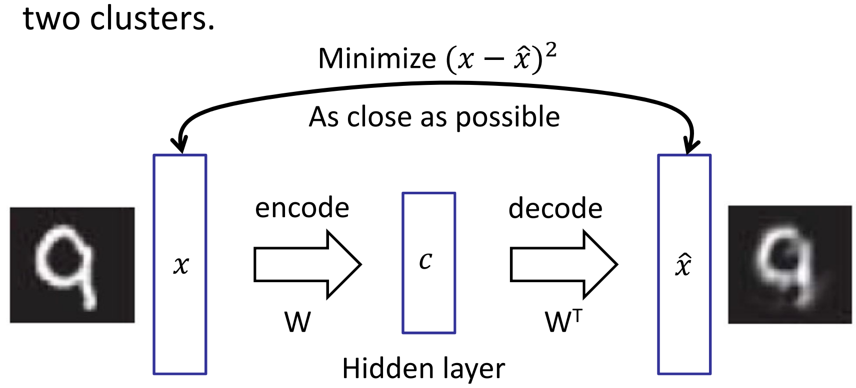
Source of images: https://zhuanlan.zhihu.com/p/25939348

37

37

Autoencoder is not good at generating

- We only ask AE to keep the original data closer to the reconstructed data.
- We never tell the algorithm what to do in the middle of two clusters.



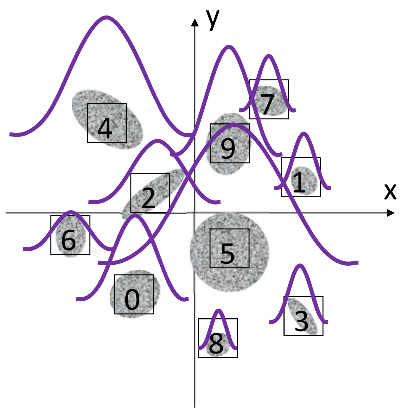
Source of images: https://zhuanlan.zhihu.com/p/25939348

38

38

So, what can we do?

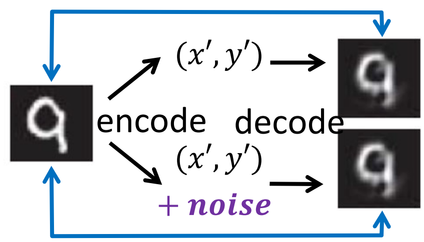
—VAE - Variational Autoencoder



(Gaussian kernel is always a good choice.)

- We add each original data points with some "influence" to other space near it and hope to fill the whole space with probability.

Not only keeps these two close



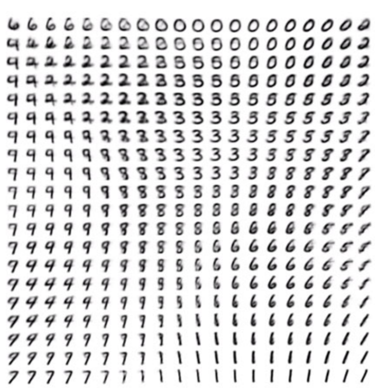
But also can keeps these two close

Source of images: https://zhuanlan.zhihu.com/p/25939348

39

39

VAE's results



Now:
Data points between two digits looks like digits ☺
Face generating looks good. But not realistic enough.

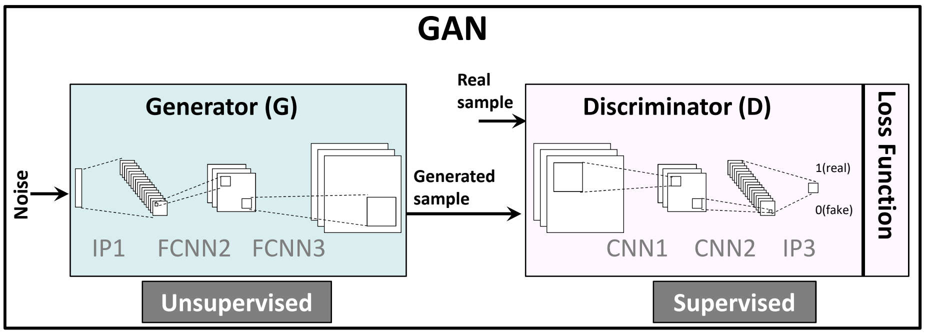
Source of images: Stanford cs231n_2018_lecture12

41

41

Generative Adversarial Networks (GAN)

- Two DNNs - a **generator** and a **discriminator** are **co-trained**
- More complex **training procedure** and **data dependency**

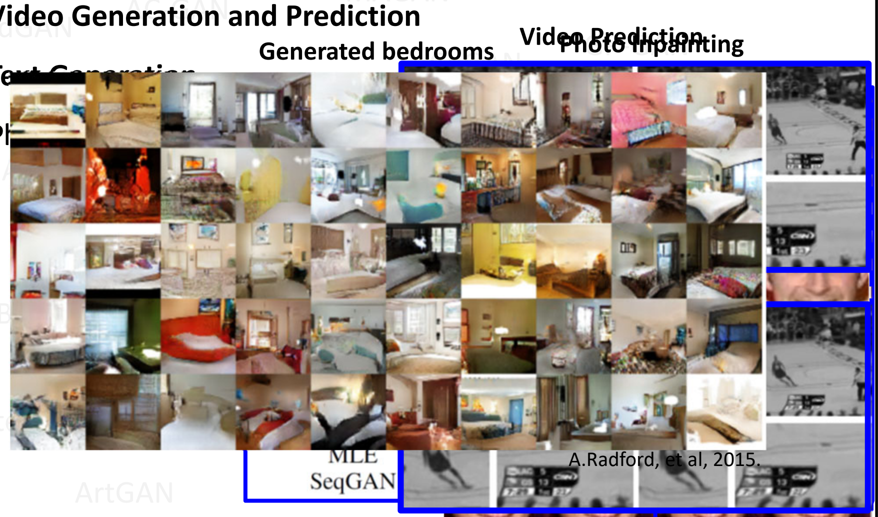


42

42

Generative Adversarial Networks (GAN)

- **Image Generation and Manipulation**
- **Video Generation and Prediction**
- **Te**
- **Pl**



43

GAN’s results

Can you tell which one is a fake image?

Source of images: https://blog.insightdatascience.com/generating-custom-photo-realistic-faces-using-ai-d170b1b59255

46

46

Autoregressive, Creating by predicting.

- The data is split into a sequence of small pieces, each of which is predicted in turn.
- Such models can be used to generate data by successively guessing what will come next, feeding in a guess as input and guessing again.
- OpenAI’s GPT-2
  - Give it a human written start, and it will go on and on. People can use it to write novel story, make fake news, …

Which could be dangerous

48

48

GPT-2's example

Human written:

A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

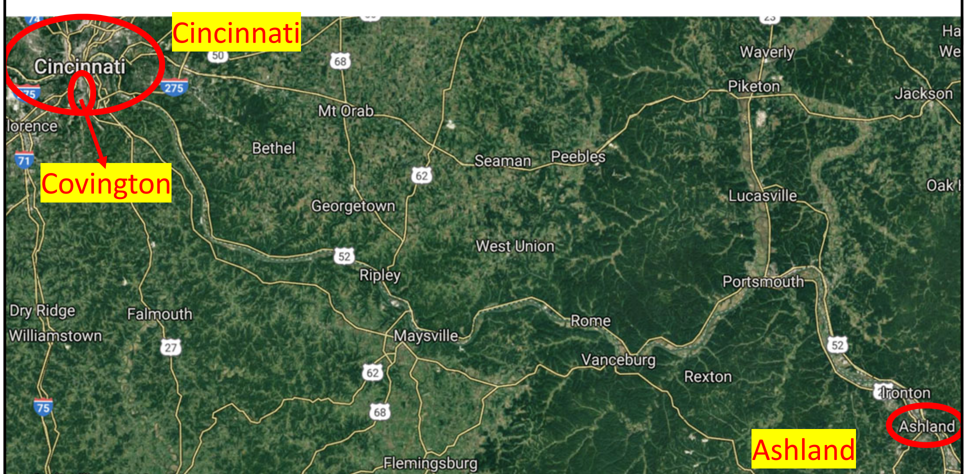
Machine written:

The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief. "The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation," said Tom Hicks, the U.S. Energy Secretary, in a statement. "Our top priority is to secure the theft and ensure it doesn't happen again." The stolen material was taken from the University of Cincinnati's Research Triangle Park nuclear research site, according to a news release from Department officials. The Nuclear Regulatory Commission …

49

49

GPT-2's example



The news generated automatically seems very convincible!

Source of images: google map

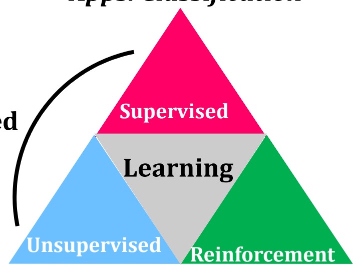
50

50

Learning Paradigms

Labeled data
Direct feedback
Predict outcome/future
Apps: Classification

Semi-supervised
Weakly-supervised



No labels
No feedback
Find hidden representations
Apps: Reconstruction

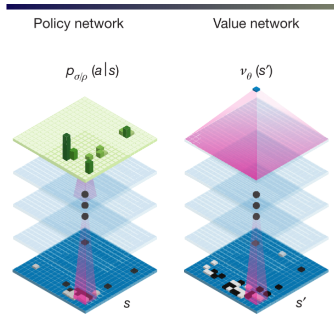
Decision process
Reward system
Learn series of actions
Apps: Decision-making

http://www.jiqizhixin.com/article/1883

51

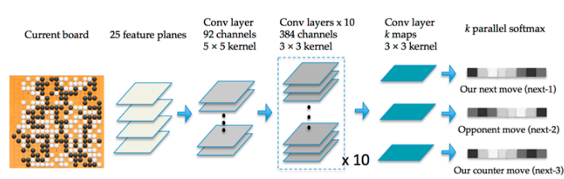
51

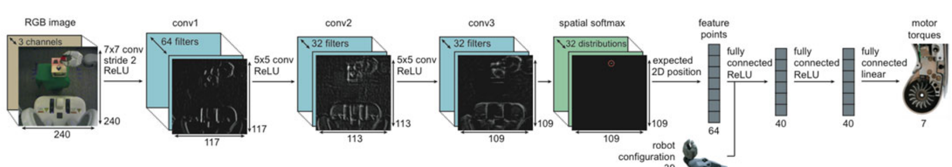
Reinforcement Learning



Alpha Go

DarkForest





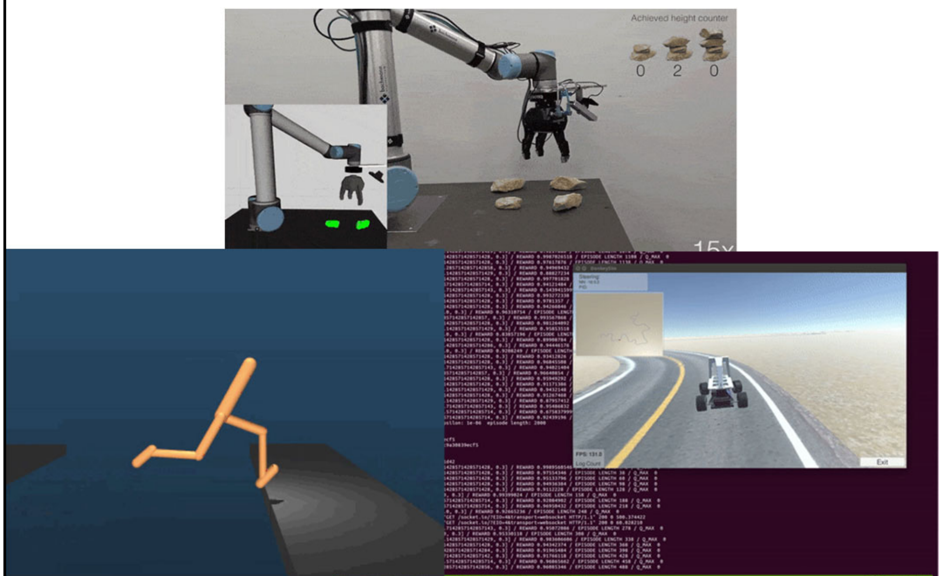
DNN in Robotics

D. Silver et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 2016.
T. Yuandong, Z. Yan. Better Computer Go Player with Neural Network and Long-term Prediction. arXiv preprint arXiv:1511.06410.
Levine, Sergey, et al. "End-to-end training of deep visuomotor policies." arXiv preprint arXiv:1504.00702.

54

54

Reinforcement learning examples



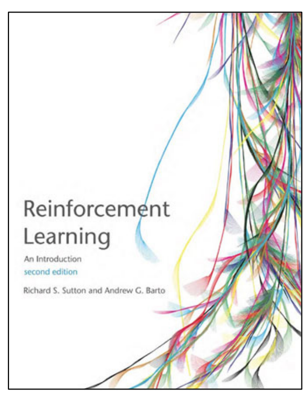
Source of images: https://www.theverge.com/tldr/2017/7/10/15946542/deepmind-parkour-agent-reinforcement-learning; https://flyyufelix.github.io/2018/09/11/donkey-rl-simulation.html; https://www.freecodecamp.org/news/a-brief-introduction-to-reinforcement-learning-7799af5840db/

55

55

Reinforcement learning reference

- Deep mind RL online course
- https://www.youtube.com/playlist?list=PL7-jPKtc4r78-wCZcQn5IqyuWhBZ8fOxT



Source of images: https://www.datasciencecentral.com/profiles/blogs/a-semi-supervised-classification-algorithm-using-markov-chain-and

56

56

In this lecture, we learned:

- A brief review of the traditional ML methods.
  - Supervised/Unsupervised/Reinforcement learning
- Why we need deep learning
  - Real world: Dimension is high, function is complex.
  - Deep neural network has the capability of doing so.
- DL examples
  - CNN, Structured learning for supervised
  - AE, VAE, GAN, GPT-2 for unsupervised
  - Deep reinforcement learning

Next Lecture: PyTorch Tutorial.
Please bring your laptop to class.

57

57