

机器学习

Machine learning

第十章 神经网络与深度学习 (1)

Neural Network and Deep Learning

授课人：周晓飞

zhouxiaofei@iie.ac.cn

2021-12-17

第十一章 神经网络与深度学习

14.1 概述

14.2 多层感知机

14.3 卷积网络

14.4 Recurrent 网络

14.5 前沿概述

第十一章 神经网络与深度学习

14.1 概述

14.1.1 背景与现状

14.1.2 神经网络模型概述

14.1.3 深度网络模型概述

背景与现状

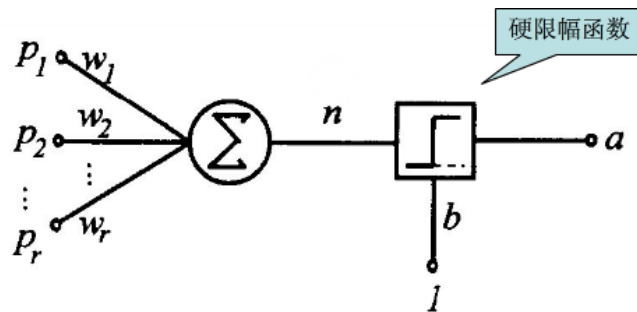
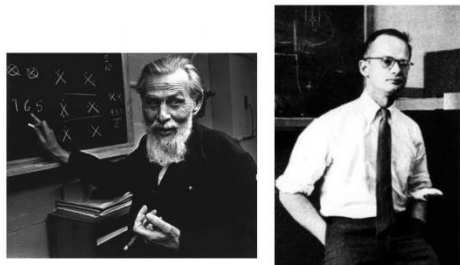
ANN 到 DL 的技术发展

ANN 始于 1890 年

始于开始于美国心理学家 W.James 对于人脑结构与功能的研究。

M-P 模型 (1943 年)

神经科学家麦卡洛克 (W.S.McCulloch) 和数学家皮兹 (W.Pitts) 建立了神经网络和数学模型 (M-P 模型)，人工神经网络的大门由此开启。



背景与现状

ANN 到 DL 的技术发展

Hebb 学习规则 (1949 年)

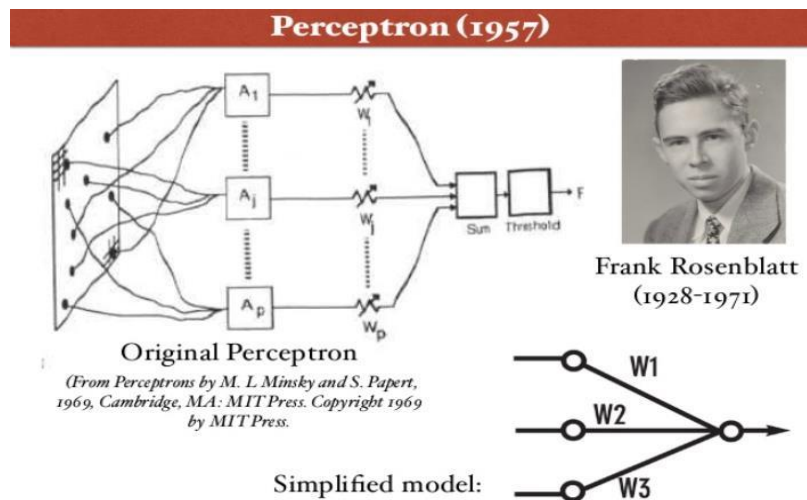
加拿大著名心理学家唐纳德 赫布提出了 Hebb 学习规则，这是一种无监督的学习规则。Hebb 学习规则表明了网络是可以学习的，这启发了后面一系列关于神经网络的研究工作。



ANN 到 DL 的技术发展

感知机 (1958 年)

心理学家 Frank Rosenblatt 受到 Hebb 思想的启发提出了“感知器”。感知器是最早的人工神经网络，也是具有学习功能 M-P 模型。整个 1958 年~1969 年期间，有许多科学家和学者都投入到了“感知器”研究。但是由于当时的计算水平相对落后，计算也显得很吃力。

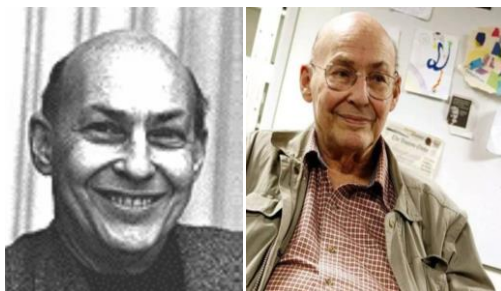


ANN 到 DL 的技术发展

1969 年进入冰河期

马文.明斯基 (Marvin Minsky) 在发表《Perceptrons》[J], 证明了感知器的致命弱点: 不能够解决异或问题。

马文.明斯基 (Marvin Minsky) 在人工智能研究方面的权威性还是相当高的, 他是第一个图灵奖的获得者。



ANN 到 DL 的技术发展

神经网络（1986 年）：BP 算法

Rumelhar 和 Hinton 提出了反向传播算法（BP 算法）是一种监督学习算法，解决了两层神经网络计算的复杂性。

深度学习之父 Geoffrey Hinton，在计算条件不发达以及数据水平也不够的条件下，很多人都放弃了对神经网络的研究，能够坚持对神经网络研究的人很少。反向传播算法的提出，引起了人工神经领域的研究热潮。



ANN 到 DL 的技术发展

卷积神经网络(1989 年)

1989 年，LeCun 发明了卷积神经网络-LeNet，并将其用于数字识别，且取得了较好的成绩，不过当时并没有引起足够的注意。

90 年代 ANN 的滞缓发展

- 统计学习的春天，浅层学习算法蓬勃发展
- 尤其 SVM 算法的诞生，再次将大家的研究方向带离了神经网络。
- BP 算法被指出存在梯度消失问题，该发现对此时的 ANN 发展雪上加霜。

ANN 到 DL 的技术发展

RNN模型 (Jordan(1986), Elman(1990))

递归 (recurrent) 的现代定义由 Jordan (1986 年), 随后 Elman (1990 年) 的 RNN 网络。

LSTM 模型 (1997 年)

LSTM 提出尽管该模型在序列建模上的特性非常突出, 但由于正处于 ANN 的下坡期, 也没有引起足够的重视。

深层信度网络 (2006 年)

2006 年, DL 元年。

Hinton 提出了深层网络训练中梯度消失问题的解决方案: **无监督预训练对权值进行初始化, 并进行有监督训练微调。**但是由于没有特别有效的实验验证, 该论文并没有引起重视。

背景与现状

ANN 到 DL 的技术发展

ReLU 激活函数 (2011 年)

该激活函数能够有效的抑制梯度消失问题。

语音识别突破 (2011 年)

微软首次将 DL 应用在语音识别上，取得了重大突破。

ImageNet 竞赛夺冠 (2012 年)

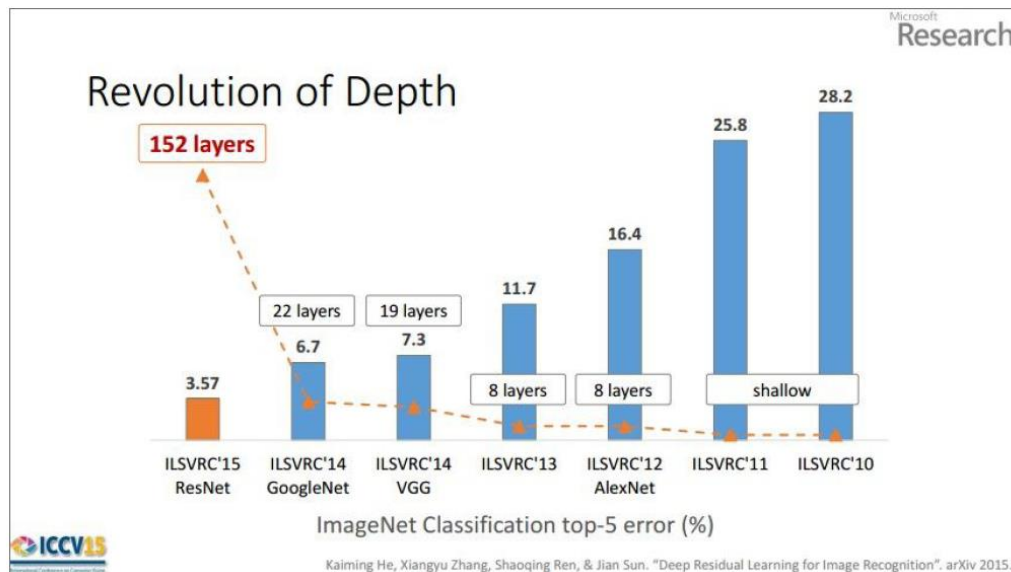
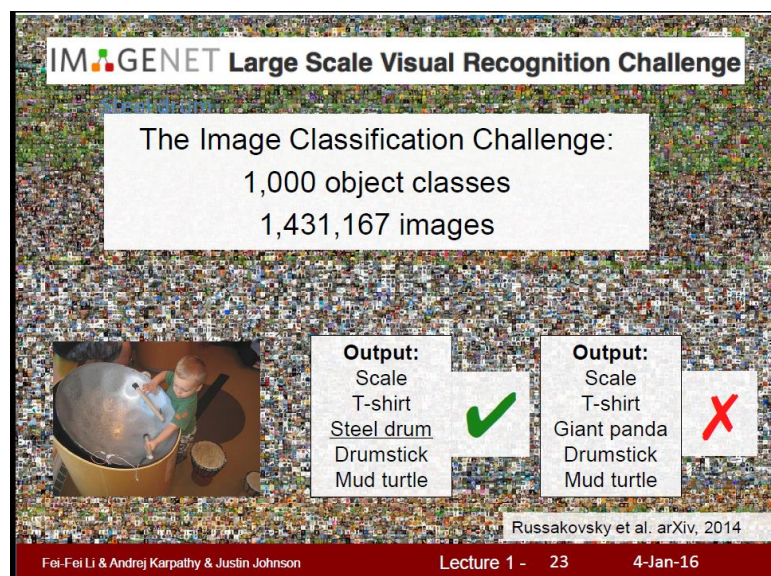
Hitton 团队首次参加 ImageNet 图像识别比赛，其通过构建的 **AlexNet** 网络一举夺得冠军。

背景与现状

ANN 到 DL 的技术发展

截至 2015 年 ImageNet 竞赛结果，ResNet: 3.6% top 5 error.

2013,2014,2015 年，通过 ImageNet 图像识别比赛，DL 的网络结构，训练方法，GPU 硬件的不断进步，促使其在其他领域也在不断的征服战场。

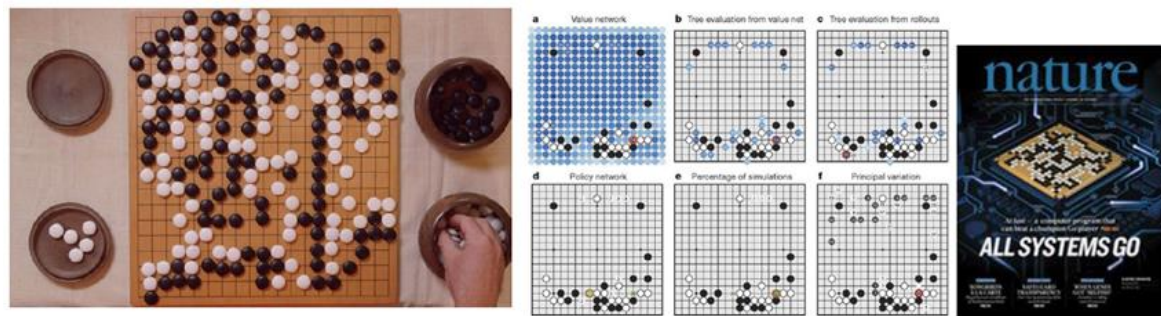


背景与现状

ANN 到 DL 的技术发展

2016、2017 年 AlphaGo (强化学习)

2016 年 3 月人工智能围棋比赛，谷歌（Google）旗下 DeepMind 公司的戴维·西尔弗、艾佳·黄和戴密斯·哈萨比斯与他们的团队开发的 AlphaGo 战胜了世界围棋冠军、职业九段选手李世石，并以 4:1 的总比分获胜。



ANN 到 DL 的技术发展

(2017 年--) 深度学习的技术突破

- 生成对抗 (GAN)
- 注意力机制 (Attentional Mechanism)
- 预训练模型: BERT

DL 在 A I 成功应用

语音识别

2009 年，Hinton 把深层神经网络介绍给做语音识别的学者们。

2010 年，语音识别就产生了巨大突破。本质上是把传统的混合高斯模型（GMM）替换成了深度神经网络（DNN）模型，但相对识别错误率一下降低 20% 多，这个改进幅度超过了过去很多年的总和。这里的关键是把原来模型中通过 GMM 建模的手工特征换成了通过 DNN 进行更加复杂的特征学习。

在此之后，在深度学习框架下，人们还在不断利用更好的模型和更多的训练数据进一步改进结果。现在语音识别已经真正变得比较成熟，并且被广泛商用，且 **目前所有的商用语音识别算法没有一个不是基于深度学习的。**

DL 在 A I 成功应用

计算视觉

- 视觉系统的信息处理是分级

1981 年的诺贝尔医学奖，分发给了 David Hubel、Torsten Wiesel 和 Roger Sperry。前两位的主要贡献是，发现了人的视觉系统的信息处理是分级。

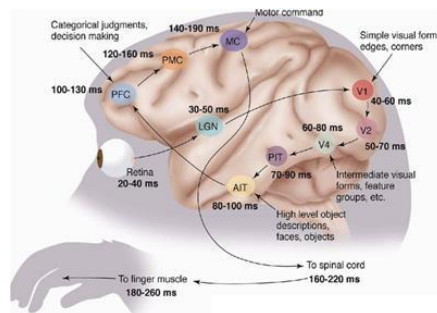


图 1 人的视觉处理系统

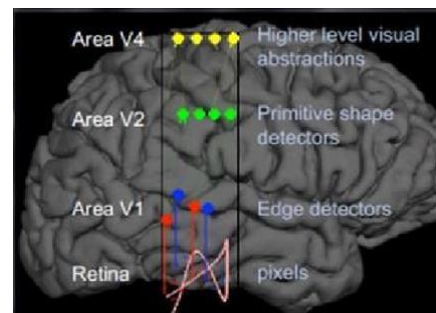


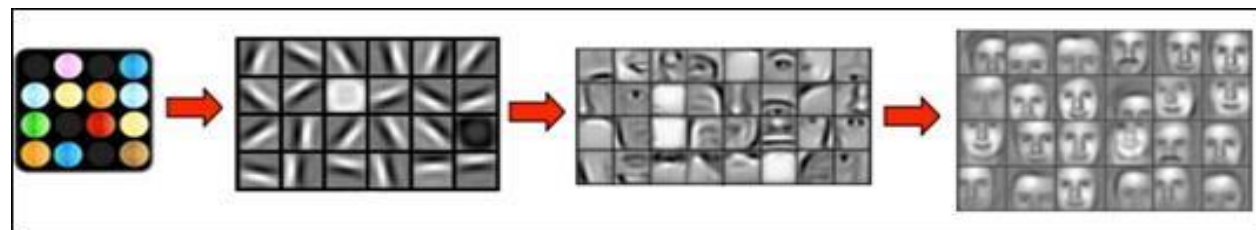
图 2 视觉系统分层处理结构

如图 1 所示，从视网膜（Retina）出发，经过低级的 V1 区提取边缘特征，到 V2 区的基本形状或目标的局部，再到高层 V4 的整个目标（如判定为一张人脸），以及到更高层的 PFC（前额叶皮层）进行分类判断等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表达越来越抽象和概念化。

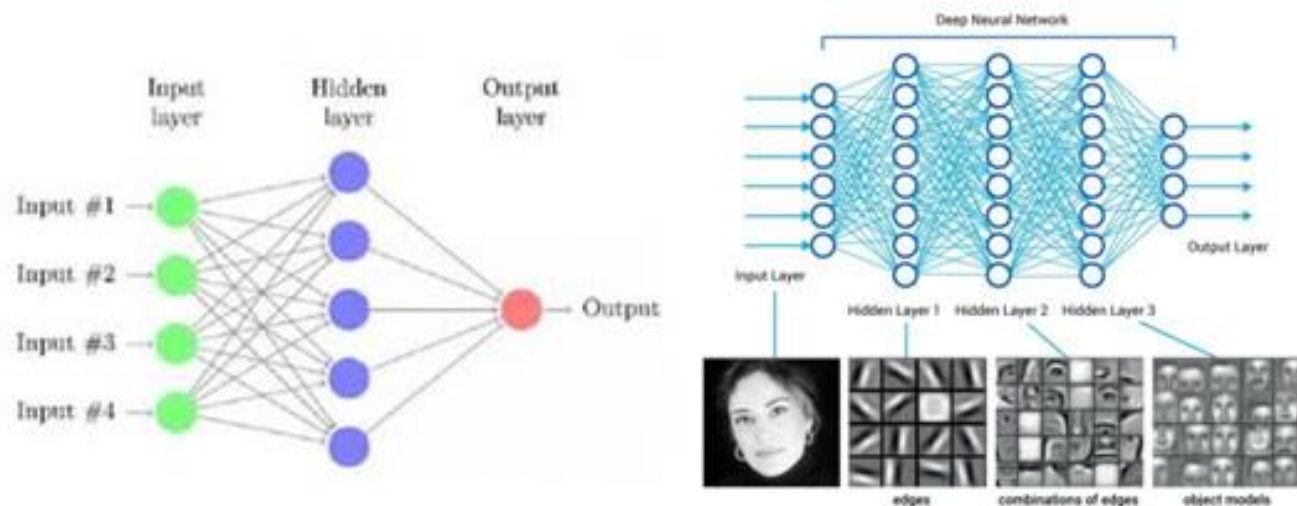
背景与现状

DL 在 A I 成功应用

- 大脑是一个深度架构，认知过程也是深度的：



而深度学习，恰恰就是通过组合低层特征形成更加抽象的高层特征



DL 在 A I 成功应用

- **D L 在图像识别**

Yann LeCun 早在 1989 年就开始了卷积神经网络的研究，取得了在一些小规模（手写字）的图像识别的成果，但在像素丰富的图片上迟迟没有突破；

直到 2012 年 Hinton 和他学生在 ImageNet 上的突破，使识别精度提高了一大步；截至 2015 年最好的模型 ResNet (He et al)；

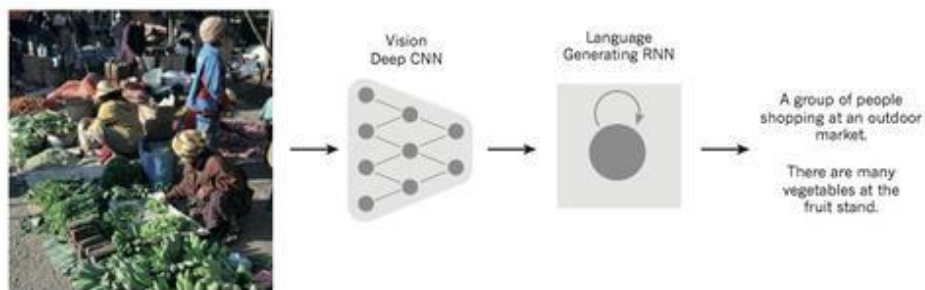
2012 年 Google Brain 用 16000 个 CPU 核的计算平台训练 10 亿神经元的深度网络，无外界干涉下自动识别了“Cat”；

2014 年香港中文大学教授汤晓鸥研究组 DeepID 的深度学习模型，在 LFW 数据库上获得了 99.15% 的识别率，人用肉眼在 LFW 上的识别率为 97.52%，深度学习在学术研究层面上已经超过了人用肉眼的识别。

背景与现状

DL 在 A I 成功应用

- DL 在图像理解



背景与现状

DL 在 AI 成功应用

自然语言处理

- **词向量表示学习**

词向量是指通过对大量文本的无监督学习，根据前后文自动学习到每个词的紧凑向量表达，包括 NNML、Word2Vector、预训练模型等。

预训练模型：ELMo、GPT 和 BERT 等，全线提升自然语言领域的多项任务的 Baseline.

- **递归神经网络 (RNN)**

文本的各个词之间是有顺序的，RNN 能更好的挖掘和利用这个性质，在自然语言各个领域都在尝试进行中。已经 BPTT、LSTM 等。

背景与现状

DL 在 A I 成功应用

主要学者

Hinton, LeCun, Bengio, Andrew Ng



背景与现状

DL 在 A I 成功应用

主推技术的公司

Google, IBM, Microsoft, NVIDIA(英伟达), 国内 BAT

国际会议

NIPS, ICML, ICLR

深度学习的未来

2015 年, Yann LeCun, Yoshua Bengio 和 Geoffrey Hinton 发表于《Nature》的综述文章“Deep Learning”。

第十一章 神经网络与深度学习

14.1 概述

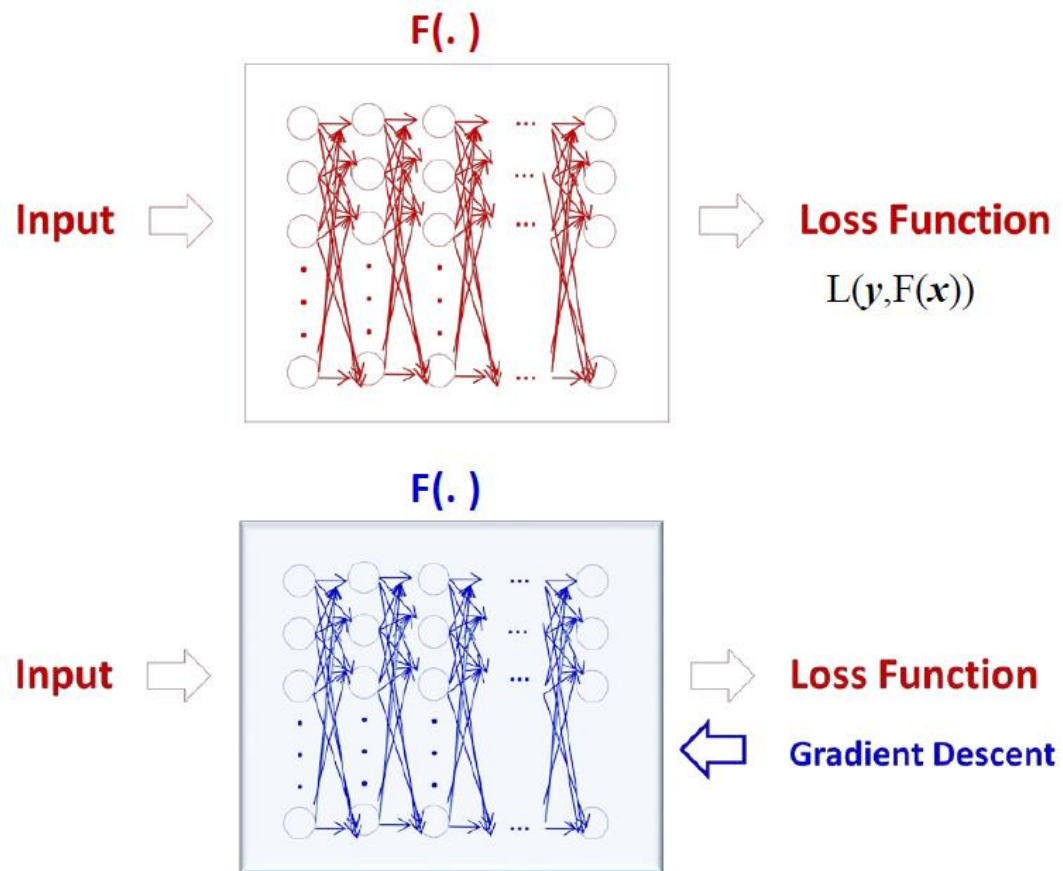
14.1.1 背景与现状

14.1.2 神经网络模型概述

14.1.3 深度网络模型概述

神经网络模型概述

神经网络模型学习框架



神经网络模型概述

神经网络模型学习框架

常用损失函数

Quadratic Loss

$$L(y, \hat{y}) = (y - f(x, \theta))^2$$

例如：LMS 方法

神经网络模型概述

神经网络模型学习框架

常用损失函数

Cross-entropy Loss

$$L(y, f(x, \theta)) = - \sum_{i=1}^C y_i \log f_i(x, \theta).$$

$$f_i(x, \theta) \in [0, 1], \quad \sum_{i=1}^C f_i(x, \theta) = 1$$

神经网络模型概述

神经网络模型学习框架

Softmax:

With **softmax**, the posterior probability of $y = c$ is

$$P(y = c|\mathbf{x}) = \mathbf{softmax}(\mathbf{w}_c^T \mathbf{x}) = \frac{\exp(\mathbf{w}_c^T \mathbf{x})}{\sum_{i=1}^C \exp(\mathbf{w}_i^T \mathbf{x})}.$$

To represent class c by one-hot vector

$$\mathbf{y} = [I(1 = c), I(2 = c), \dots, I(C = c)]^T,$$

where $I()$ is indicator function.

神经网络模型概述

神经网络模型学习框架

Given training set $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}), 1 \leq i \leq N$, the cross-entropy loss is

$$\mathcal{J}(W) = - \sum_{i=1}^N \sum_{c=1}^C y_c^{(i)} \log \hat{y}_c^{(i)}$$

We use one-hot vector \mathbf{y} to represent class c in which $y_c = 1$ and other elements are 0.

$$L(y, f(x, \theta)) = -\log f_y(x, \theta).$$

神经网络模型概述

神经网络模型学习框架

参数学习

$$\begin{aligned}\theta^* &= \arg \min_{\theta} \mathcal{R}(\theta_t) \\ &= \arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y^{(i)}, f(x^{(i)}, \theta)).\end{aligned}$$

梯度下降

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{a}_t - \lambda \frac{\partial \mathcal{R}(\theta)}{\partial \theta_t} \\ &= \mathbf{a}_t - \lambda \sum_{i=1}^N \frac{\partial \mathcal{R}(\theta_t; x^{(i)}, y^{(i)})}{\partial \theta}\end{aligned}$$

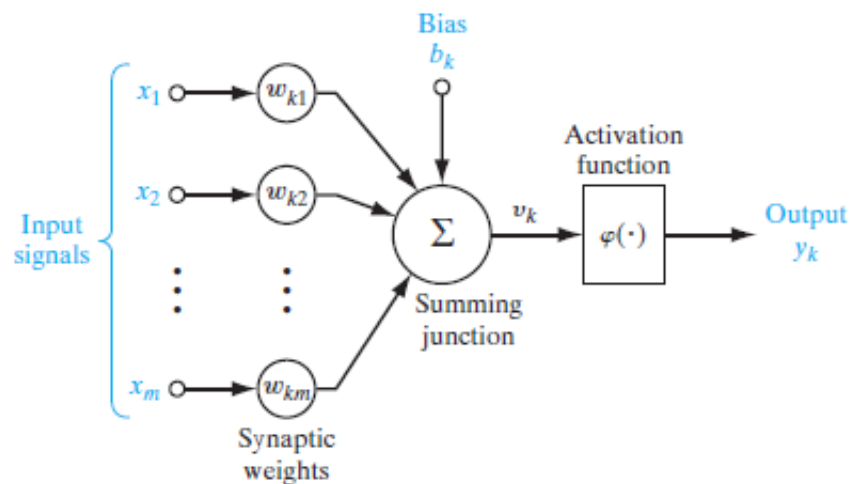
随机梯度下降

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \lambda \frac{\partial \mathcal{R}(\theta_t; x^{(t)}, y^{(t)})}{\partial \theta}$$

神经网络模型概述

1 单个神经元模型

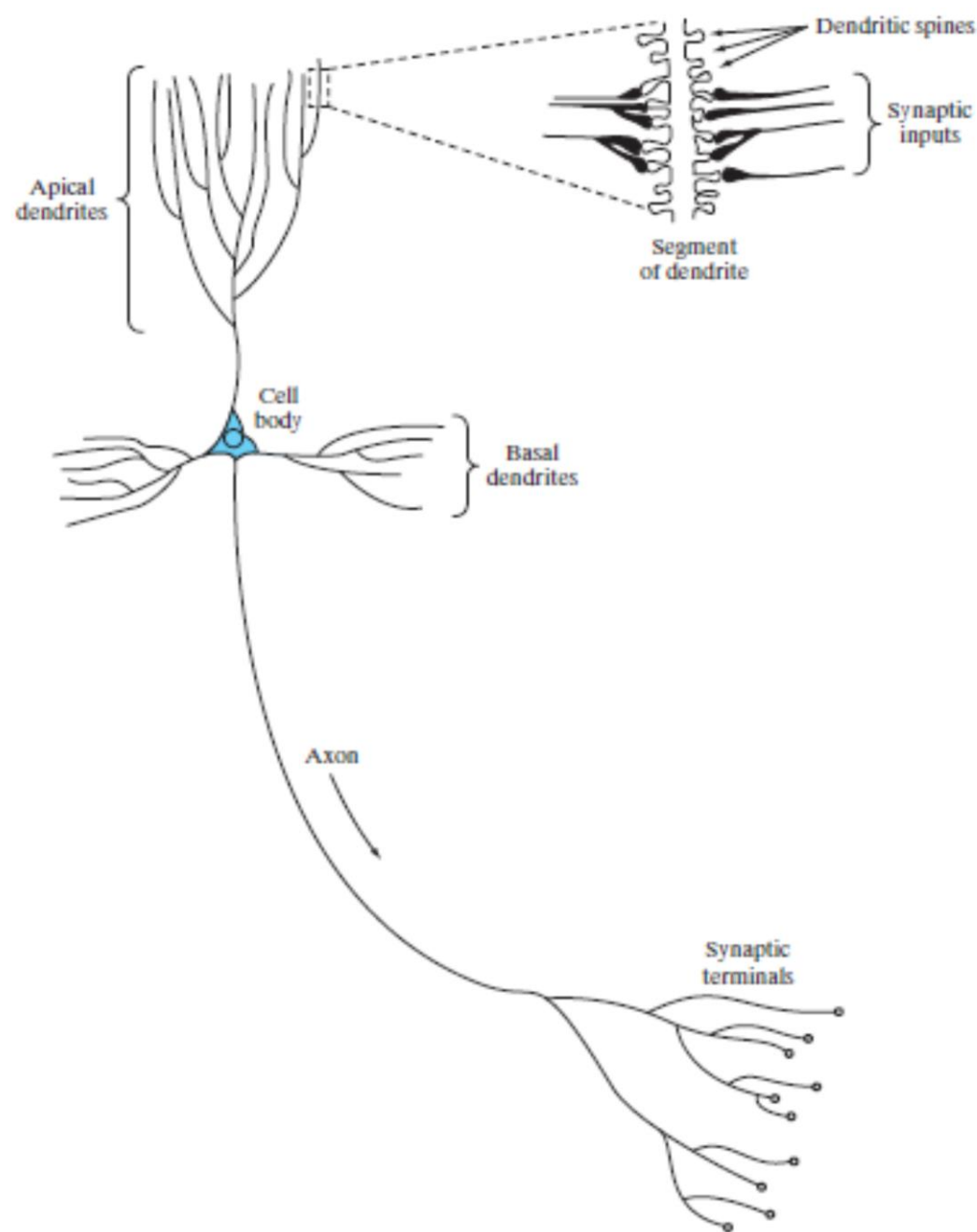
神经元



$$u_k = \sum_{j=1}^m w_{kj} x_j$$

$$y_k = \varphi(u_k + b_k)$$

基本的成分：突触连接 w ，加法器，激活函数。

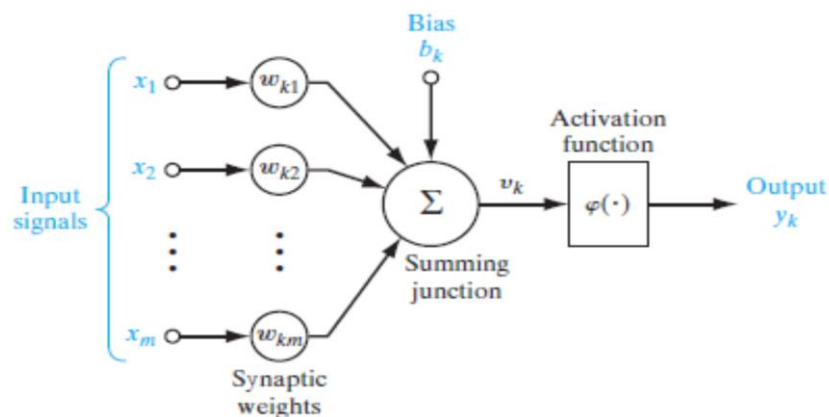


神经网络模型概述

1 单个神经元模型

主要模型

- 感知机
- 最小方差回归
- Logistic 模型



神经网络模型概述

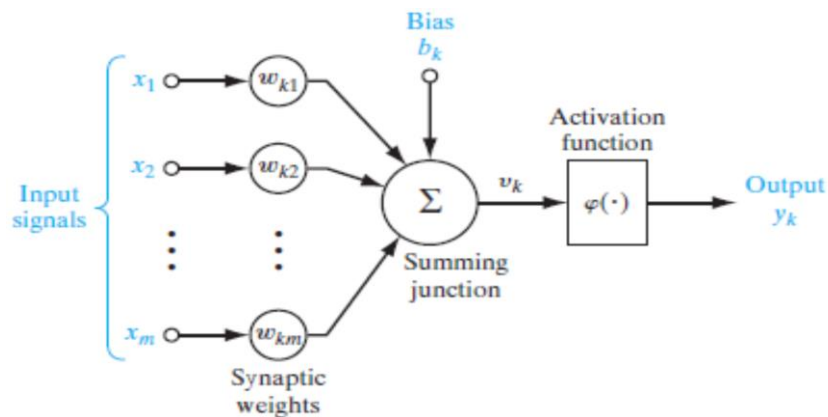
1 单个神经元模型

主要模型

- 感知机
- 最小方差回归
- Logistic 模型

损失函数

$$J(\mathbf{w}) = \sum_{\mathbf{x}(n)} -\mathbf{w}^T \mathbf{x}(n) (d(n) - y(n))$$



神经网络模型概述

1 单个神经元模型

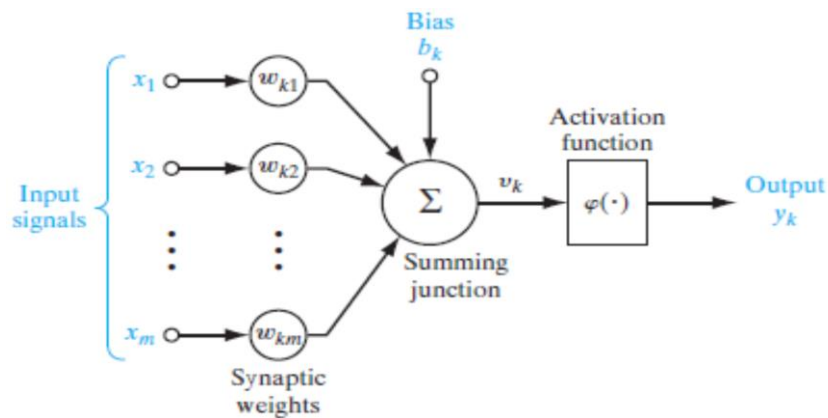
主要模型

- 感知机
- 最小方差回归
- Logistic 模型

损失函数

$$\mathcal{E}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^n e^2(i)$$

$$e(i) = d(i) - y(i)$$

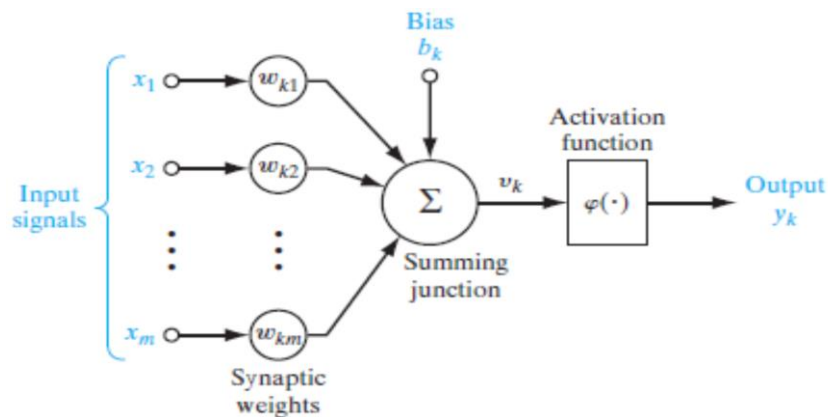


神经网络模型概述

1 单个神经元模型

主要模型

- 感知机
- 最小方差回归
- **Logistic 模型**



优化目标

The posterior probability of $y = 1$ is

$$P(y = 1|\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x})},$$

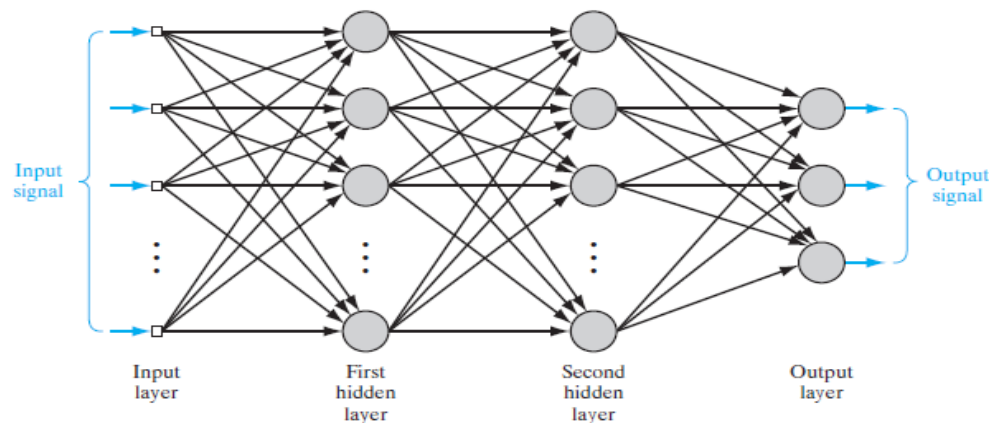
where, $\sigma(\cdot)$ is logistic function.

The posterior probability of $y = 0$ is $P(y = 0|\mathbf{x}) = 1 - P(y = 1|\mathbf{x})$.

最大似然估计

神经网络模型概述

2 多层感知机

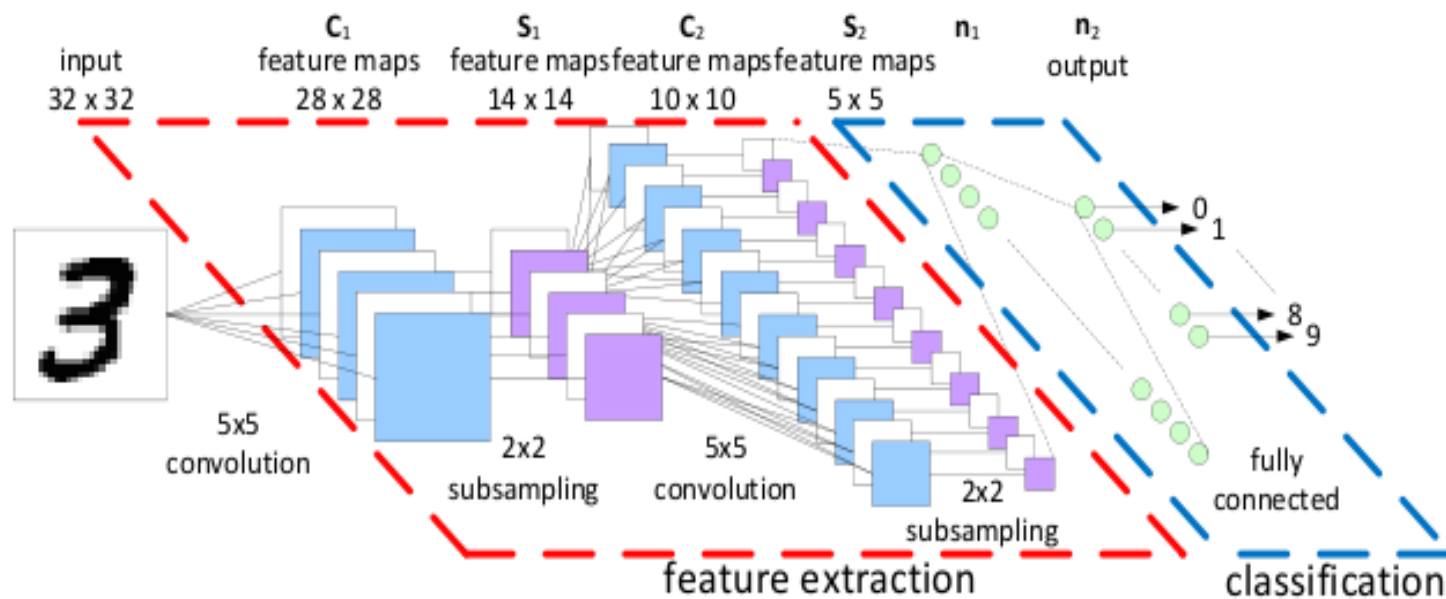


$$\mathbf{x} = \mathbf{a}^{(0)} \rightarrow \mathbf{z}^{(1)} \rightarrow \mathbf{a}^{(1)} \rightarrow \mathbf{z}^{(2)} \rightarrow \dots \rightarrow \mathbf{a}^{(L-1)} \rightarrow \mathbf{z}^{(L)} \rightarrow \mathbf{a}^{(L)} = \mathbf{y}$$

$$\begin{aligned} J(\mathbf{W}, \mathbf{b}) &= \sum_{i=1}^N L(y^{(i)}, f(\mathbf{x}^{(i)} | \mathbf{W}, \mathbf{b})) + \frac{1}{2} \lambda \|\mathbf{W}\|_F^2, \\ &= \sum_{i=1}^N J(\mathbf{W}, \mathbf{b}; \mathbf{x}^{(i)}, y^{(i)}) + \frac{1}{2} \lambda \|\mathbf{W}\|_F^2, \end{aligned}$$

神经网络模型概述

3 卷积网络

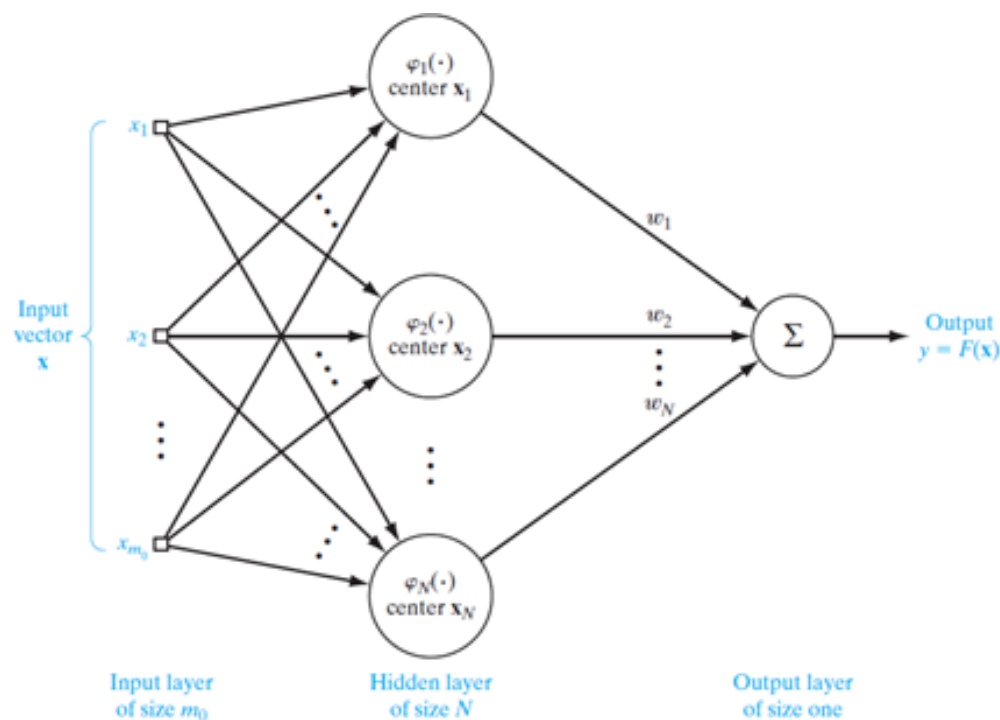


神经网络模型概述

4 核函数网络

- 径向基网络
- 支持向量机

单隐层神经网络、非线性体现在径向基核函数



- W 的求取采用最小均方在线学习

$$F(\mathbf{x}) = \sum_{i=1}^N w_i \varphi(\|\mathbf{x} - \mathbf{x}_i\|)$$

- 简化模型 (k-means 混合模型)

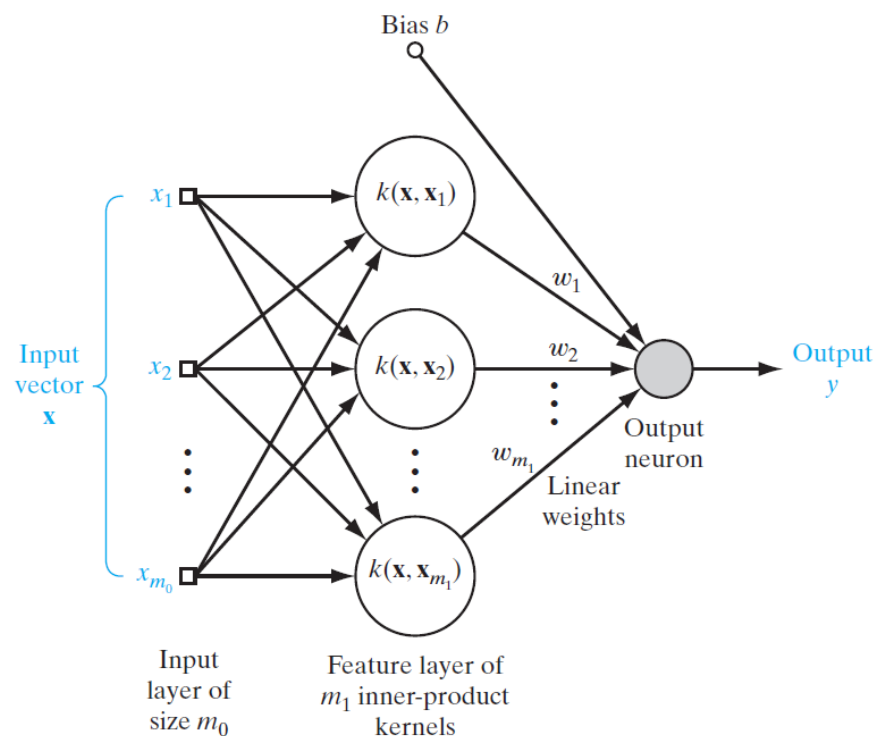
$$F(\mathbf{x}) = \sum_{j=1}^K w_j \varphi(\mathbf{x}, \mathbf{x}_j)$$

神经网络模型概述

4 核函数网络

- 径向基网络
- 支持向量机

单隐层神经网络、非线性体现在径向基核函数

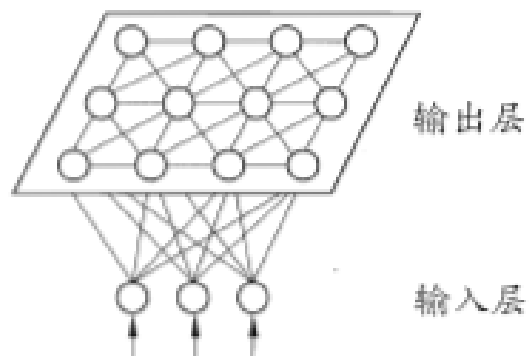


- W 的求取采用最大间隔学习

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$
$$s.t. \quad y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0, \quad i=1, 2, \dots, l$$

神经网络模型概述

5 自组织映射



竞争过程:

$$i(\mathbf{x}) = \arg \min_j \| \mathbf{x}(n) - \mathbf{w}_j \|, \quad j = 1, 2, \dots, l$$

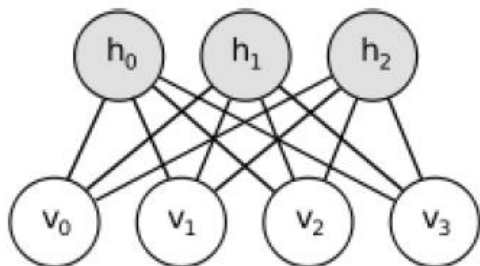
合作过程:

$$h_{j,i(\mathbf{x})} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right), \quad j \in \mathcal{A}$$

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h_{j,i(\mathbf{x})}(n) (\mathbf{x}(n) - \mathbf{w}_j(n))$$

神经网络模型概述

6 RBM



- 同层神经元间无连接，并彼此相互独立
- 是一个无向图(权值对称)，即连接可看作双向的
- h 层为隐层，v 层为可见层

神经元的条件概率分布

The energy function $E(v, h)$ of an RBM is defined as:

$$E(v, h) = -b'v - c'h - h'Wv$$

$$P(h_i = 1|v) = \text{sigm}(c_i + W_i v)$$

$$P(v_j = 1|h) = \text{sigm}(b_j + W_j' h)$$

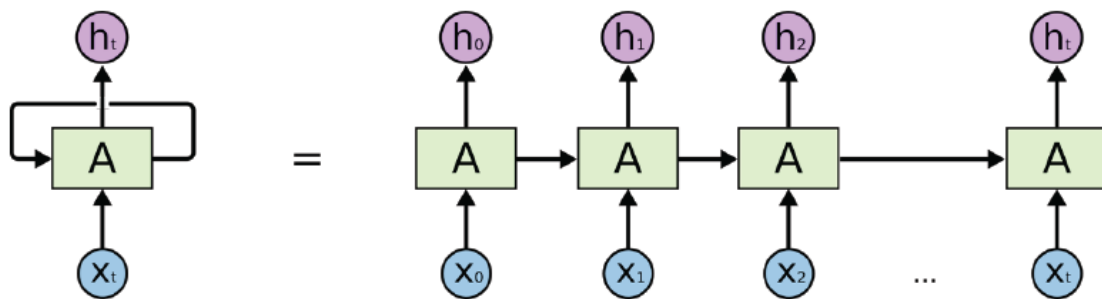
\Rightarrow

$$p(h|v) = \prod_i p(h_i|v)$$

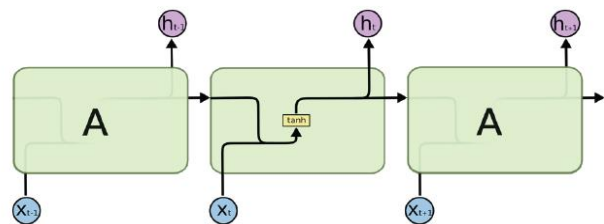
$$p(v|h) = \prod_j p(v_j|h).$$

神经网络模型概述

7 递归网络



- Unit Structure 1: Normal RNN



$$h_t = \sigma \left(W^{(hh)} h_{t-1} + W^{(hx)} x_{[t]} \right)$$
$$\hat{y}_t = \text{softmax} \left(W^{(S)} h_t \right)$$

第十一章 神经网络与深度学习

14.1 概述

14.1.1 背景与现状

14.1.2 神经网络模型概述

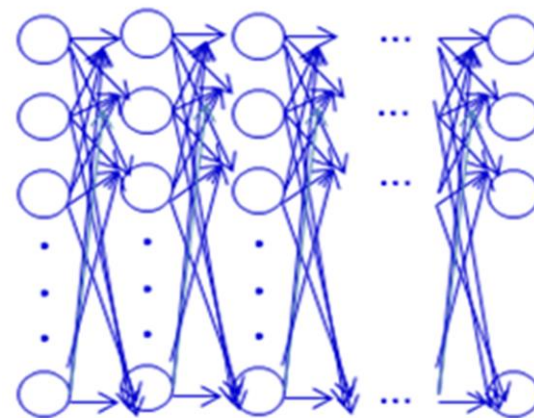
14.1.3 深度网络模型概述

深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+CNN+FL
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks

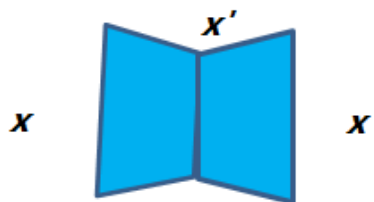


深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks



深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- **预训练+全连接网络 Au+FL**
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks

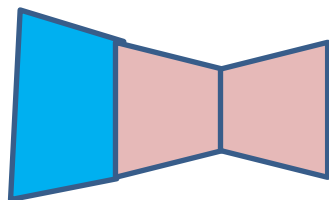


深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks

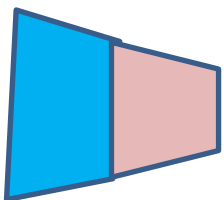


深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks

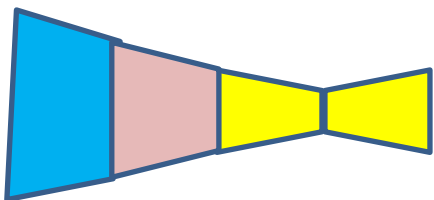


深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks

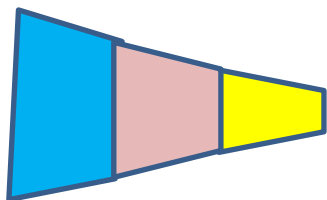


深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- **预训练+全连接网络 Au+FL**
- 卷积+全连接网络 CNN+FL
- CNN + FL+ ReLu + Tricks



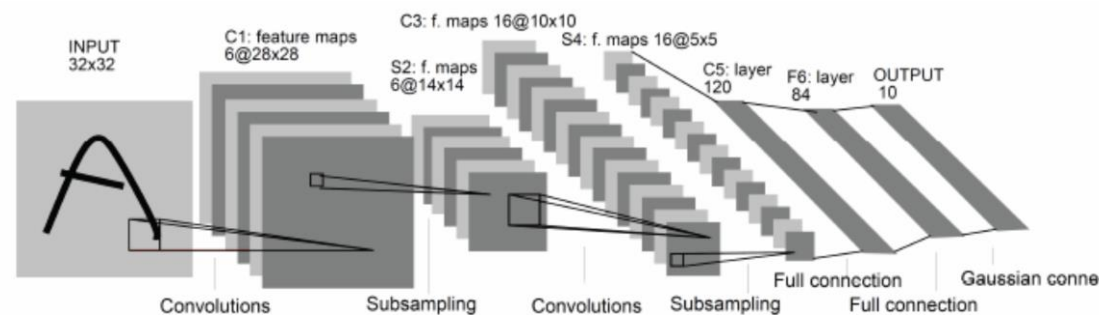
预训练相当于特征提取

深度网络模型概述

深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- **卷积+全连接网络 CNN+FL**
- CNN + FL+ ReLu + Tricks



深度网络模型概述

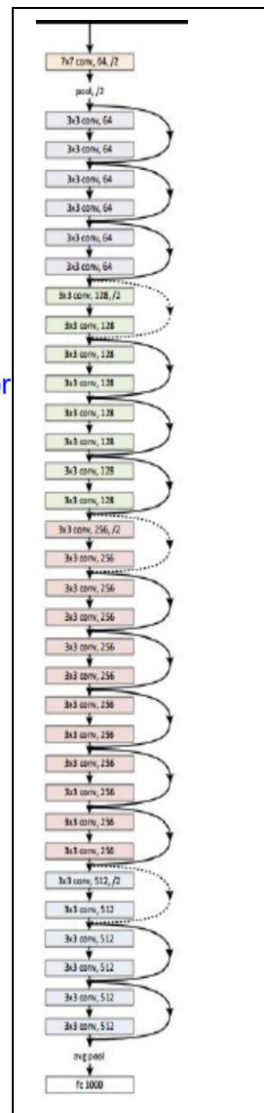
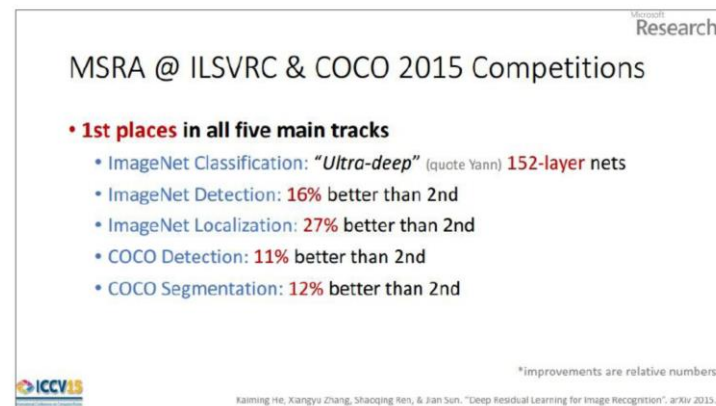
深度前馈网络

常见的结构

- 全连接网络 DFL
- 预训练+全连接网络 Au+FL
- 卷积+全连接网络 CNN+FL
- **CNN + FL+ ReLu + Tricks**

ImageNet 评测

Case Study: ResNet [He et al., 2015]
ILSVRC 2015 winner (3.6% top 5 error)

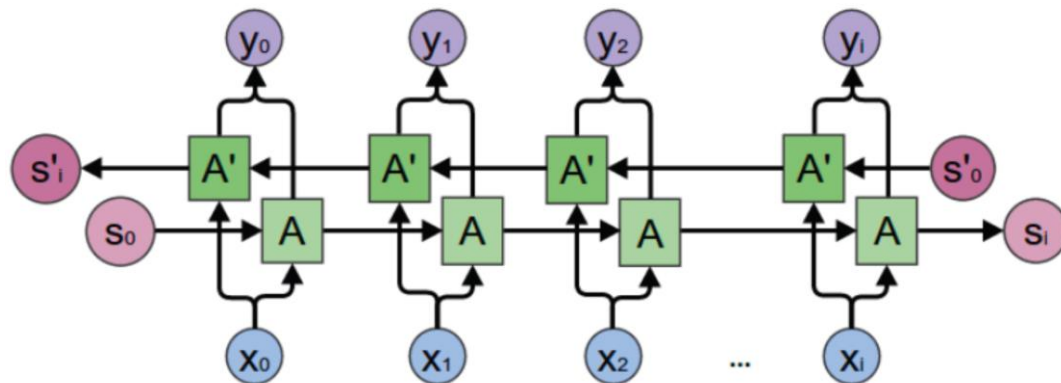


深度网络模型概述

递归神经网络

常见的结构

- Bi 结构
- Deep 结构
- CNN+RNN 结构
- ...

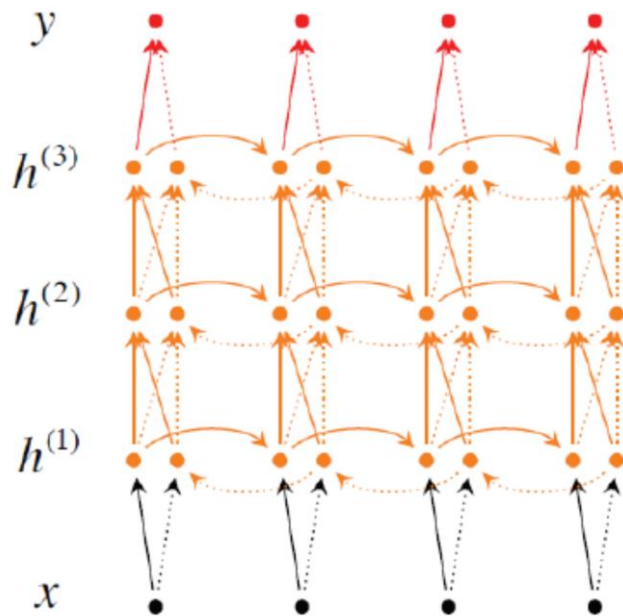


深度网络模型概述

递归神经网络

常见的结构

- Bi 结构
- **Deep 结构**
- CNN+RNN 结构
- ...

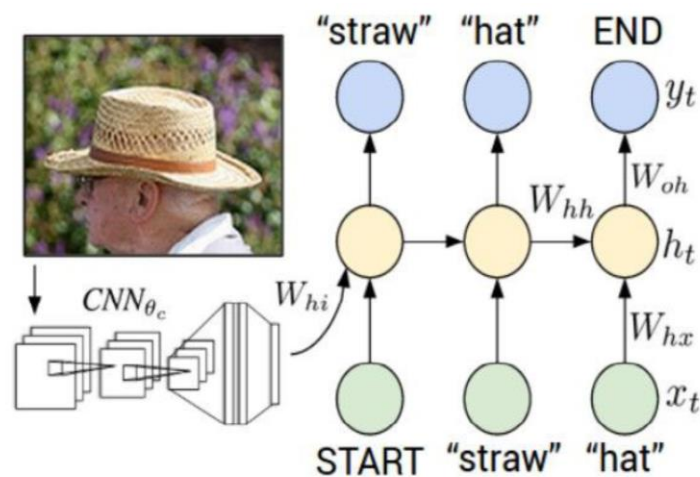


深度网络模型概述

递归神经网络

常见的结构

- Bi 结构
- Deep 结构
- **CNN+RNN 结构**
- ...



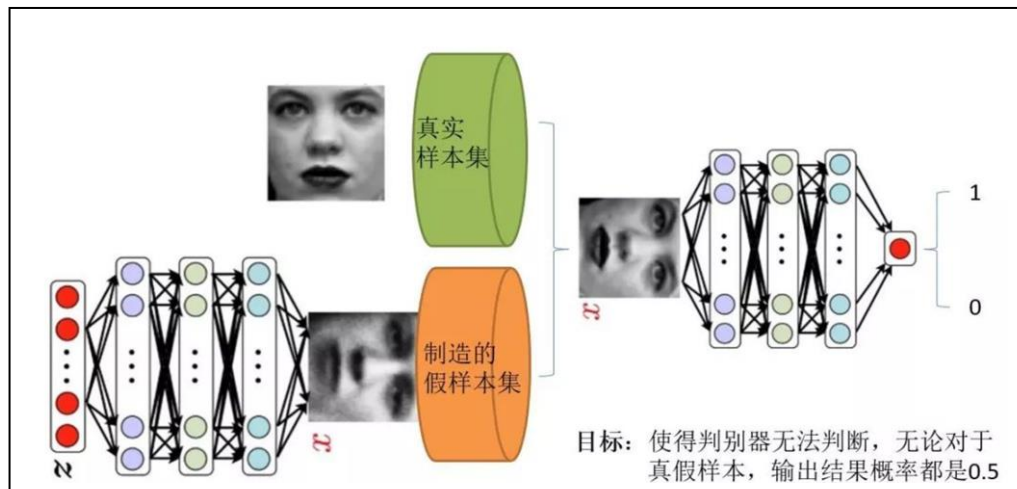
深度网络模型概述

生成对抗网(GAN)

两个网络博弈：G (Generator) 和 D (Discriminator) 。

G 是一个生成图片的网络，它接收一个随机的噪声 z ，通过这个噪声生成图片，记做 $G(z)$ 。

D 是一个判别网络，判别一张图片是不是“真实的”。它输入一张图片 x ，输出 $D(x)$ 代表 x 为真实图片的概率，如果为 1，就代表 100%是真实的图片，而输出为 0，就代表不可能是真实的图片。



深度网络模型概述

生成对抗网(GAN)

Generative Modeling

- Density estimation



- Sample generation



Training examples

Model samples

Which of these are real photos ?

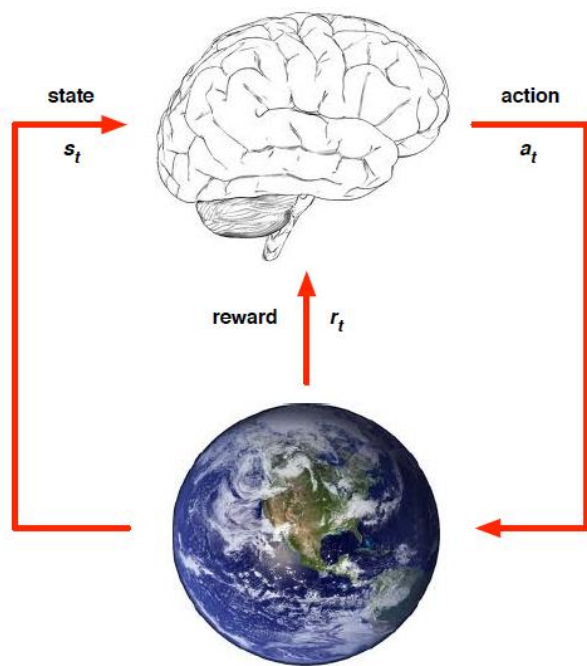


(work by vue.ai covered by Quartz)

深度网络模型概述

深度强化学习

强化学习



Episodes:

$S_1, a_1, R_2, S_2, a_2, R_3, S_3, a_3, R_4, \dots$

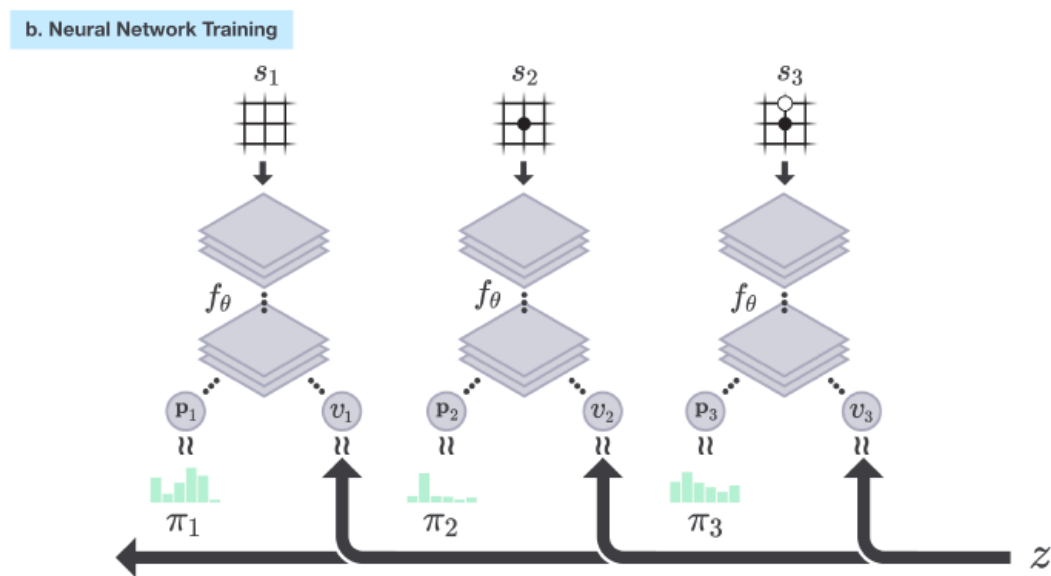
学习目标：策略概率 $p(a/s)$

深度网络模型概述

深度强化学习

值函数网络、Deep Q-Learning

策略网络、Deep Policy Network



本章的重点内容

- 多层感知机
- 卷积网络
- Recurrent 网络
- 深度学习

本讲参考文献

1. Simon Haykin, Neural Network and Learning Machine. 3rd
2. 邱锡鹏,《深度学习与自然语言处理》 Slides@CCF ADL 20160529。
3. Stanford Class-CS231n: Convolutional Neural Networks for Visual Recognition。
4. 第十一届中国中文信息学会暑期学校, 暨中国中文信息学会《前沿技术讲习班》, 201607。