

Data Mining

Lecture Notes for Chapter 4

Artificial Neural Networks

Introduction to Data Mining , 2nd Edition

by

Tan, Steinbach, Karpatne, Kumar

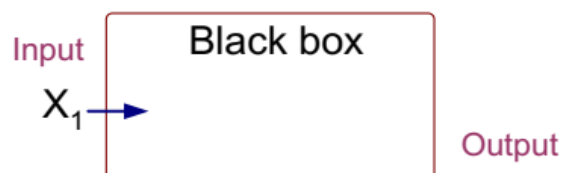
02/14/2018

Introduction to Data Mining, 2nd Edition

1

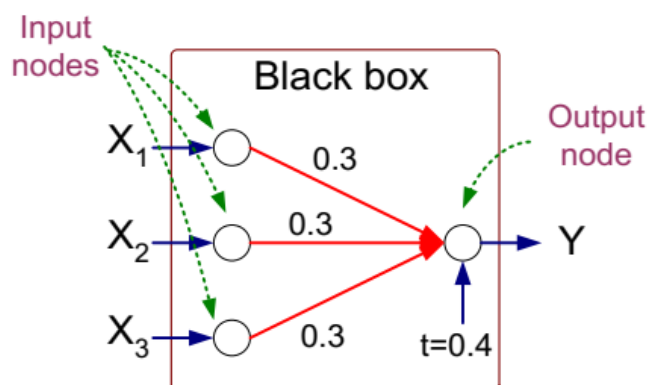
Artificial Neural Networks (ANN)

X_1	X_2	X_3	Y
1	0	0	-1
1	0	1	1
1	1	0	1



Artificial Neural Networks (ANN)

X_1	X_2	X_3	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	-1
0	1	0	-1
0	1	1	1
0	0	0	-1



$$Y = \text{sign}(0.3X_1 + 0.3X_2 + 0.3X_3 - 0.4)$$

$$\text{where } \text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

02/14/2018

Introduction to Data Mining, 2nd Edition

3

02/14/2018 数据挖掘导论，第 2 版 4

人工神经网络

●模型是以下各项的集合

互联节点

和加权链接

●输出节点总结

它的每个输入值

根据重量

它的链接

●比较输出节点

针对某个阈值测试

X1

X2

X3

Y

黑盒子

w1

t

输出

结节

投入

节点

w2

w3

感知器模型

$\sqrt{}$

$\sqrt{}$

=

=

=

=-

d

i

我

d

i

我

符号 w X

y 形符号 w X t

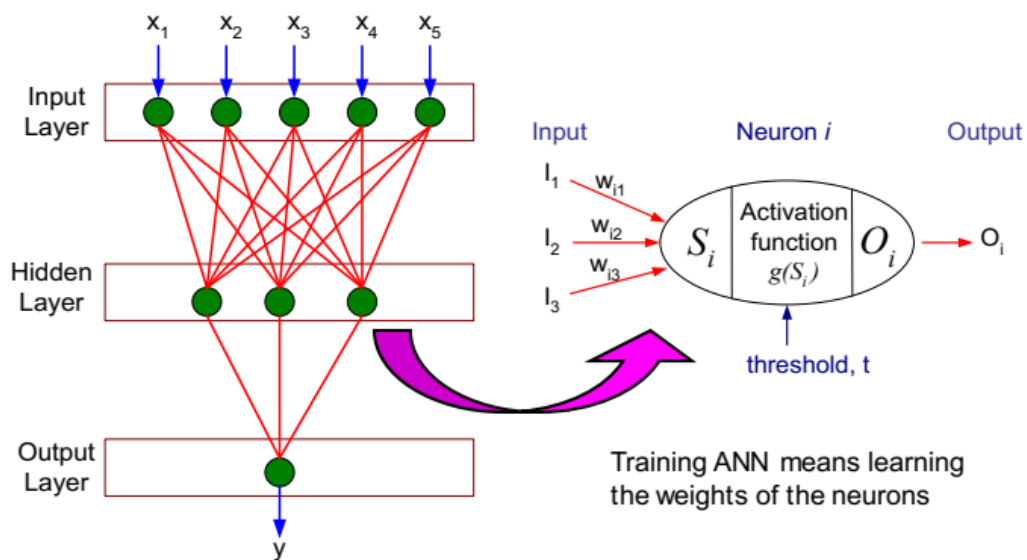
0

1

()

()

General Structure of ANN



02/14/2018

Introduction to Data Mining, 2nd Edition

5

02/14/2018 数据挖掘导论，第 2 版 6

人工神经网络

- 各种类型的神经网络拓扑

单层网络(感知机)与

多层网络

前馈与递归网络

- 各种类型

激活功能(f)

$= (\sum)$

i

$w_i x_i$ Y f

02/14/2018 数据挖掘导论，第 2 版 7

感知器

- 单层网络

仅包含输入和输出节点

- 激活功能: $f = \text{符号}(w \times x)$

- 应用模型很简单

$X_1 = 1, X_2 = 0, X_3 = 1 \Rightarrow y = \text{符号}(0.2) = 1$

$\{ \int_{<=}$

$= + +$

1 if 0

1 if 0

其中()

$(0.3 \ 0.3 \ 0.3 \ 0.4) \ 1 \ 2 \ 3$

x

x
符号 x
y 符号 X X X

02/14/2018 数据挖掘导论，第 2 版 8
感知器学习规则
●初始化重量(w_0, w_1, \dots, w_d)
●重复
对于每个训练例子(x_i , y_i)
u 计算机 $f(x_i, w)$
u 更新重量:
●直到满足停止条件
$$[I . I . k . k . w . y . f(w, x) x()() = ++\lambda$$

02/14/2018 数据挖掘导论，第 2 版 9
感知器学习规则
●重量更新公式:
●直觉:
基于误差更新重量:
如果 $y=f(x, w)$, $e=0$:不需要更新
如果 $y>f(x, w)$, $e=2$:重量必须增加
 $f(x, w)$ 会增加
如果 $y<f(x, w)$, $e=-2$:重量必须降低
 $f(x, w)$ 会减少
$$[(,) ; \text{学习率}() \lambda I \lambda k I k k w = w + y f w x + x$$

$$[(,) () I k I e = y f w x$$

Example of Perceptron Learning

$$w^{(k+1)} = w^{(k)} + \lambda [y_i - f(w^{(k)}, x_i)] x_i$$

$$Y = \text{sign}(\sum_{i=0}^d w_i X_i)$$

$$\lambda = 0.1$$

X ₁	X ₂	X ₃	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	-1
0	1	0	-1
0	1	1	1
0	0	0	-1

	w ₀	w ₁	w ₂	w ₃
0	0	0	0	0
1	-0.2	-0.2	0	0
2	0	0	0	0.2
3	0	0	0	0.2
4	0	0	0	0.2
5	-0.2	0	0	0
6	-0.2	0	0	0
7	0	0	0.2	0.2
8	-0.2	0	0.2	0.2

Epoch	w ₀	w ₁	w ₂	w ₃
0	0	0	0	0
1	-0.2	0	0.2	0.2
2	-0.2	0	0.4	0.2
3	-0.4	0	0.4	0.2
4	-0.4	0.2	0.4	0.4
5	-0.6	0.2	0.4	0.2
6	-0.6	0.4	0.4	0.2

02/14/2018

Introduction to Data Mining, 2nd Edition

10

02/14/2018 数据挖掘导论，第2版 11

感知器学习规则

- 因为 $f(w, x)$ 是线性的

输入组合

变量，决策

边界是线性的

- 对于非线性可分问题，感知器

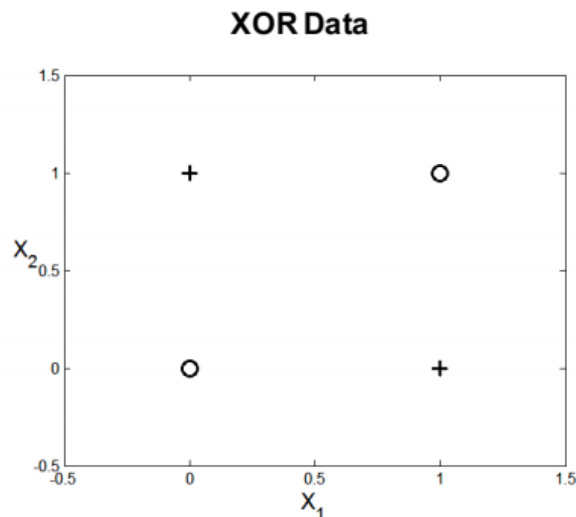
学习算法会失败，因为没有线性

超平面可以完美地分离数据

Nonlinearly Separable Data

$$y = x_1 \oplus x_2$$

x_1	x_2	y
0	0	-1
1	0	1
0	1	1
1	1	-1

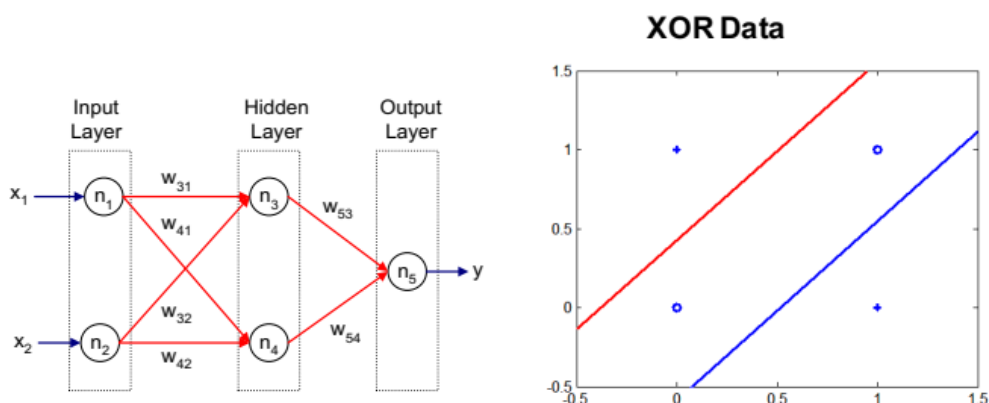


Multilayer Neural Network

- Hidden layers
 - intermediary layers between input & output layers

Multi-layer Neural Network

- Multi-layer neural network can solve any type of classification task involving nonlinear decision surfaces



02/14/2018 数据挖掘导论，第 2 版 15

学习多层神经网络

●我们能把感知器学习规则应用到每一个吗
节点，包括隐藏节点？

感知器学习规则计算误差项

$e = yf(w, x)$, 并相应更新权重

u 问题: 如何确定 y 的真实值

隐藏节点？

隐藏节点中的近似错误由中的错误引起

输出节点

u 问题:

不清楚隐藏节点中的调整如何影响整体误差

无法保证收敛到最优解

02/14/2018 数据挖掘导论，第 2 版 16

多层神经网络的梯度下降

●重量更新:

●误差函数:

●激活函数 f 必须是可微的

●对于乙状结肠功能:

●随机梯度下降(更新权重

立即)

j

$k j$

k

$j w$

E

$w w$

∂

$\partial = + \lambda \frac{\partial}{\partial}$

$\sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^p (w_{ij} - w_{ij}^{old})^2 = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^p (w_{ij} - w_{ij}^{old})^2$

$= + \infty +$

i

$i i i i j$

$k j$

k

$j w w (t o) o (o) x () () \lambda$

02/14/2018 数据挖掘导论，第 2 版 17

多层神经网络的梯度下降

●对于输出神经元，

重量更新公式为

和以前一样

(梯度下降

感知器)

●对于隐藏神经元:

\sqrt

\sqrt

$\in \varphi$

$\in \varphi$
 $+$
 $= -$
 $= -$
 $= +$
 j
 i
 k
 $j j j k j k$
 $j j j j j$
 j
 $i i j i j p_i$
 k
 p_i
 k
 p_i
 $o o w$
 $o o t o$
 $w w o o w x$
 $\delta \delta$
 δ
 $\lambda \delta$
 隐藏神经元:(1)
 输出神经元:(1)()
 (1)()
 批发物价指数
 wq_i
 神经元 I
 神经元 p
 神经元 q
 神经元 x
 神经元 y
 威克斯
 w_{iy}
 隐蔽层
 $k-1$
 隐蔽层
 k
 隐蔽层
 $k+1$

02/14/2018 数据挖掘导论，第 2 版 18

人工神经网络中的设计问题

●输入层的节点数

每个二进制/连续属性一个输入节点

每个分类属性的 k 个或 $\log_2 k$ 个节点

k 值

●输出层中的节点数

二进制类问题的一个输出

k 类问题的 k 或 $\log_2 k$ 节点

●隐藏层中的节点数

- 初始权重和偏差

02/14/2018 数据挖掘导论，第 2 版 19

人工神经网络的特点

- 多层人工神经网络是通用的近似器，但如果网络太大，就会遭受过度拟合
- 梯度下降可能会收敛到局部最小值
- 模型构建可能非常耗时，但是测试可能非常快
- 可以处理冗余属性，因为权重自动学习
- 对训练数据中的噪声敏感
- 难以处理缺失的属性

02/14/2018 数据挖掘导论，第 2 版 20

人工神经网络最近值得注意的发展

- 用于深度学习和无监督功能
- 学问
- 寻求自动学习好的
- 来自未标记的输入的数据

- 谷歌大脑项目

通过观察学会了“猫”的概念

YouTube 上未标记的图片

十亿个连接网络