深度学习概述

阿里云

深度学习（Deep Learning，DL）

- 是机器学习的分支；
- 是一种以人工神经网络为架构；
- 对资料进行表征学习的算法。
- 表征学习（特征学习）：将原始数据转换成为能够被机器学习有效开发的一种技术的集合。

传统机器学习与深度学习对比

优点

- 学习能力强
- 覆盖范围广、适应性好
- 数据驱动、上限高
- 可移植性好

缺点

- 计算量大、便携性差
- 硬件需求高
- 模型设计复杂
- 容易存在偏见

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

深度学习框架的定义及特点

阿里云

- 一种界面、库或工具；
- 能够让开发人员利用预先构建和优化好的组件集合定义模型；
- 更容易、更快速地构建深度学习模型；
- 一个良好的深度学习框架应具备以下5个关键特征；
- 常用深度学习框架包括TensorFlow、Torch、Caffe等。

01 针对性能进行优化

02 易于理解与编码

03 强大的社区生态

04 并行化进程加快运算

05 自动计算渐变

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

TensorFlow框架的特点

阿里云

TensorFlow深度学习框架

- 一个利用数据流图（Data Flow Graphs）进行数值计算的开源软件库；
- 可以在众多异构的系统上方便地移植。

机动性

可适性强

多种编程语言可选

最优化表现

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

Torch框架的特点

阿里云

Torch深度学习框架

- 包含大量的机器学习、计算机视觉、信号处理、并行计算、图像、视频、音频的库；
- 和Caffe类似，拥有大量的训练好的深度学习模型。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

Caffe框架的特点

阿里云

Caffe深度学习框架

- 一个清晰而高效的深度学习框架；
- 基于C++/CUDA的架构，支持命令行、Python和Matlab接口；
- 可以在CPU和GPU直接无缝切换，并支持多GPU。

广泛地应用于
前沿的工业界和学术界

可完成计算机视觉
领域的多种任务

第一个主流的工业级
深度学习框架

在各种硬件环境编译
并具有良好的移植性

提供了Python语言
接口pycaffe

使用许多顺序连接的层来描述
神经网络结构

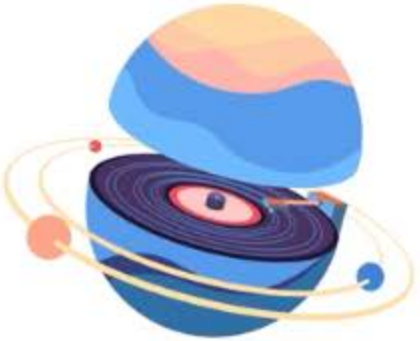
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

深度学习框架的选择

深度学习框架的选择往往需要考虑多方面的因素，如学习门槛、上手难度、开发速度、易用性等。

01 学习门槛

02 上手难度



03 开发速度

04 易用性

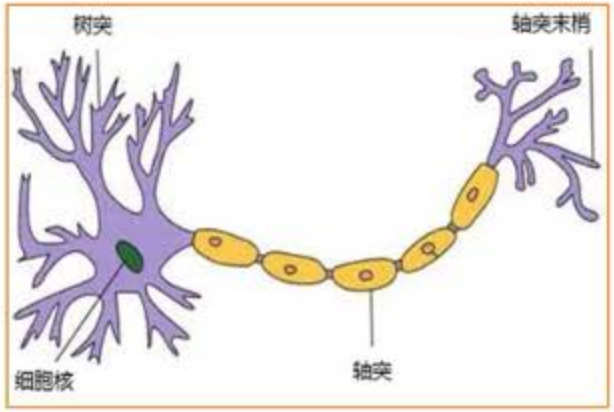
05 文档齐全规范

神经网络基础

阿里云

定义

- 人工神经网络（Artificial Neural Network，ANN），简称神经网络（Neural Network，NN）；
- 在计算机领域中，是一种模仿生物神经网络的结构和功能的数学或计算模型；
- 目的是模拟大脑的某些机理与机制，实现某个方面的功能，例如图像识别、语音识别。



生物神经元

神经元抽象模型

神经网络抽象模型

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

神经网络组成

阿里云

神经元

- 负责计算和处理输入信号

网络连接

- 负责将不同神经元连接起来，形成神经网络
- 连接两端各为一个神经元
- 一个神经元的输出为另一个神经元的输入，例如信号a
- 网络连接有加权参数w，经过加权计算后，信号变成a*w

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

感知机概述

阿里云

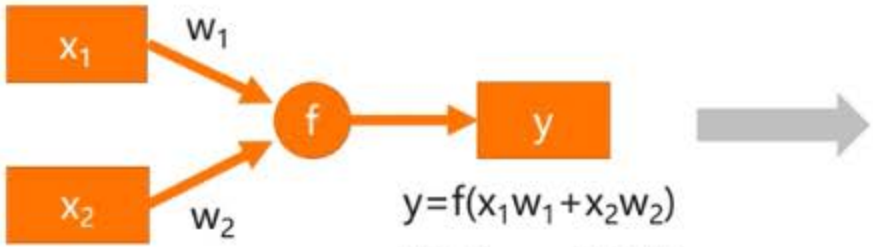
感知机的定义

- 人工智能最早的模型
- 一种有监督的学习算法
- 本质上是一个二分类问题
- 是神经网络和支持向量机的基础

感知机的缺点

- 感知机只能解决单纯的线性问题

感知机的过程



$y=f(x_1w_1+x_2w_2)$

f为Sigmoid函数

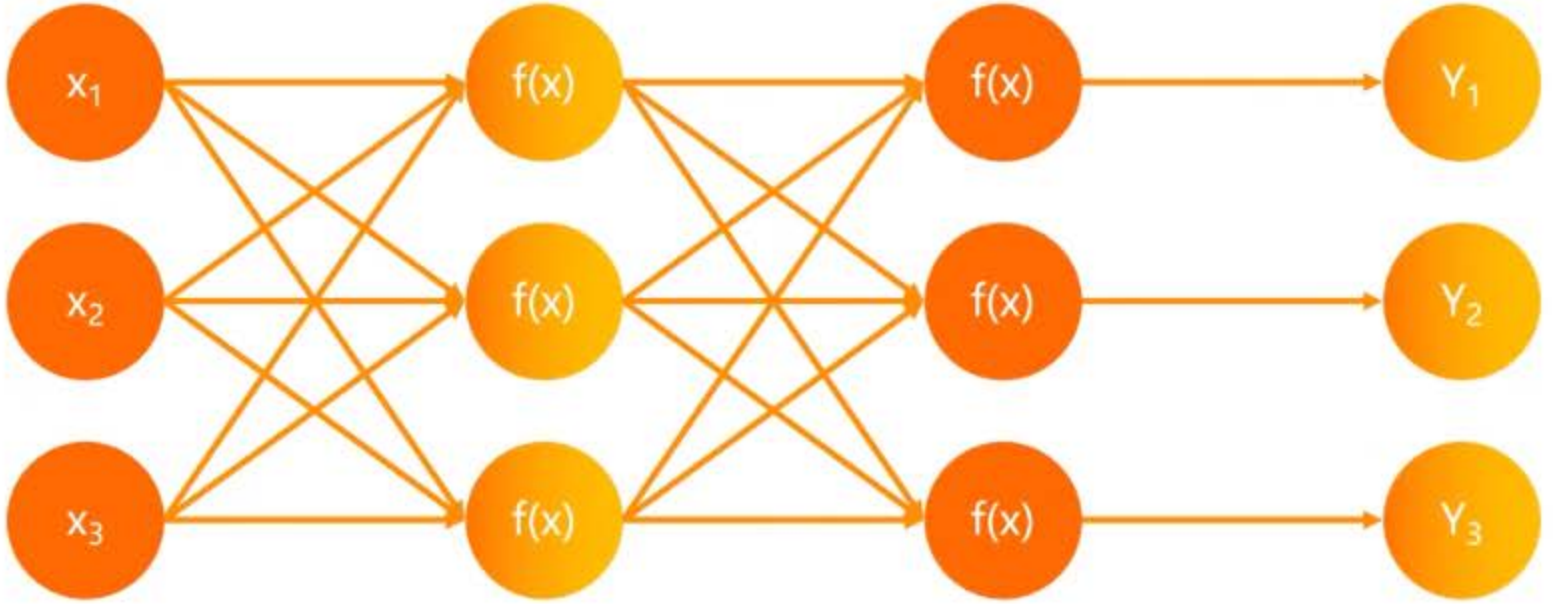
$$y=\begin{cases}0 & (w_1x_1+w_2x_2\le 0)\\ 1 & (w_1x_1+w_2x_2>0)\end{cases}$$

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

多层感知机的层级结构

阿里云

多层感知机的层级结构主要包含**输入层、隐藏层和输出层**，可以用于拟合非线性函数。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

激活函数概述

阿里云

定义： 激活函数是一种在人工智能神经网络的神经单元上运行的函数，旨在帮助网络学习数据中的复杂模式，负责将神经元节点的输入映射到输出端。

常见激活函数： Sigmoid函数、Tanh函数、ReLU函数等等。

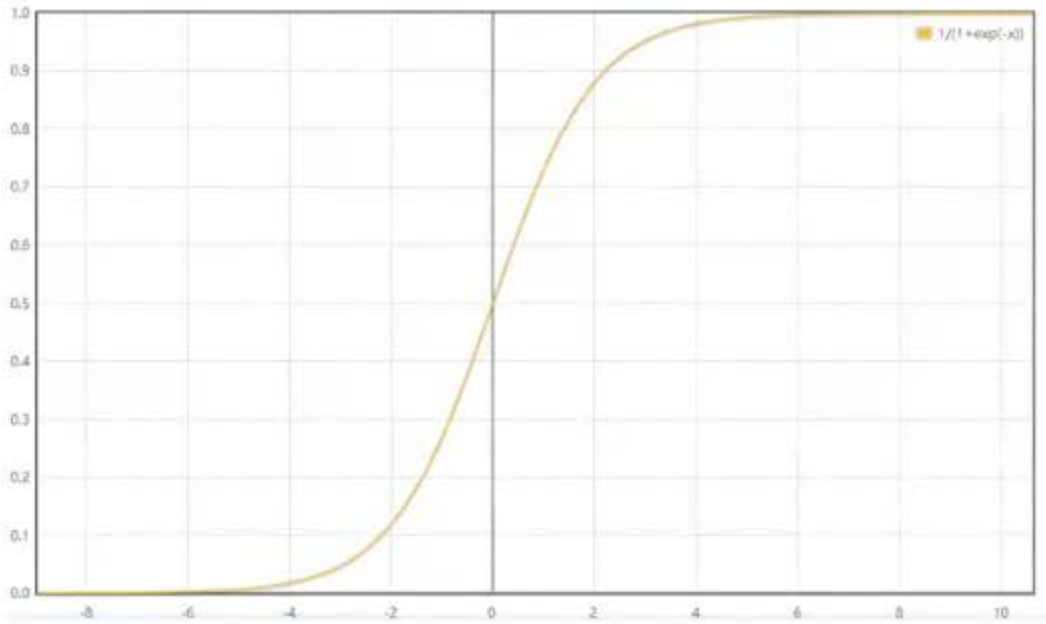
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

Sigmoid激活函数

阿里云

Sigmoid函数

- Sigmoid函数也叫Logistic函数，用于隐藏层神经元输出，取值范围为0到1，可以将一个实数映射到0到1区间，可用于二分类。
- 输出不是0均值
- 存在梯度消失的情况



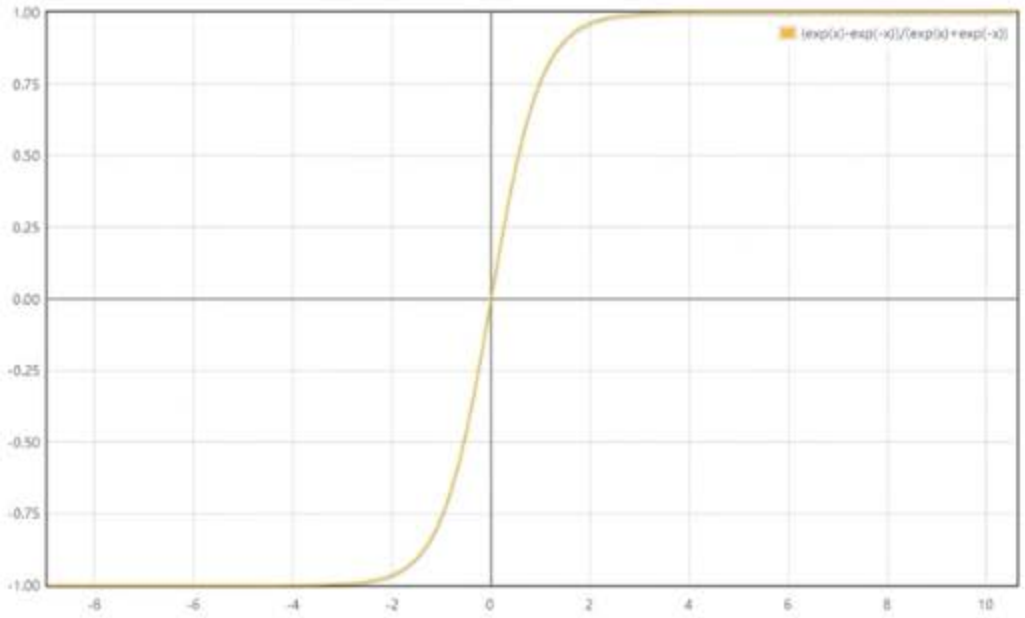
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

Tanh激活函数

阿里云

Tanh函数

- Tanh函数解决了Sigmoid函数的不是零均值化输出问题，在特征相差明显时的效果更好，在循环过程中会不断扩大特征效果。
- 存在梯度消失的问题



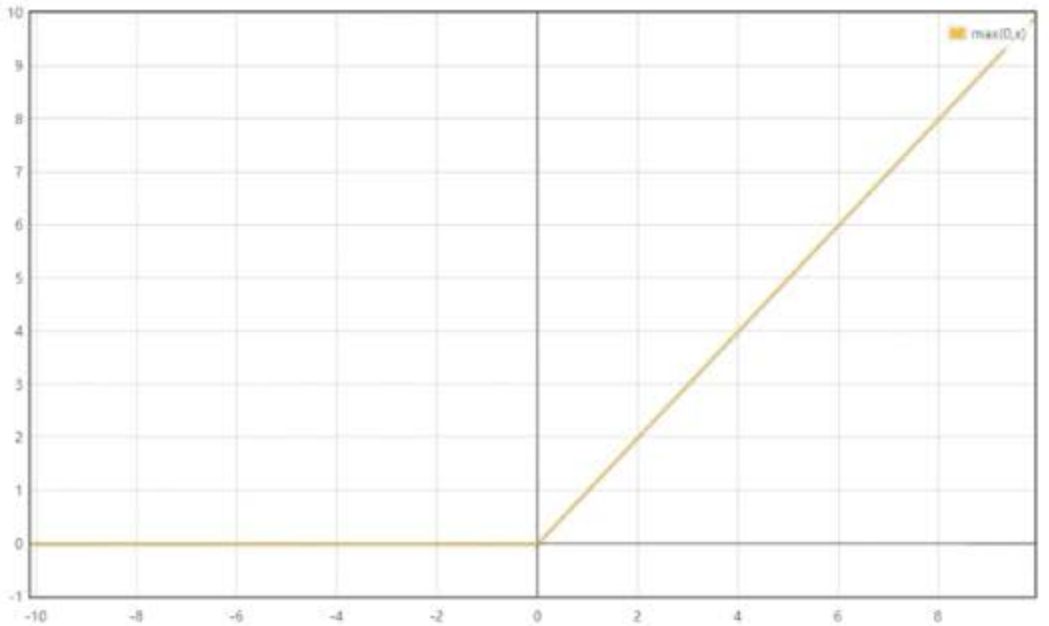
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

ReLU激活函数

阿里云

ReLU函数

- ReLU函数是最常用的激活函数，它解决了梯度消失的问题，计算速度非常快，收敛速度远快于Sigmoid和Tanh。



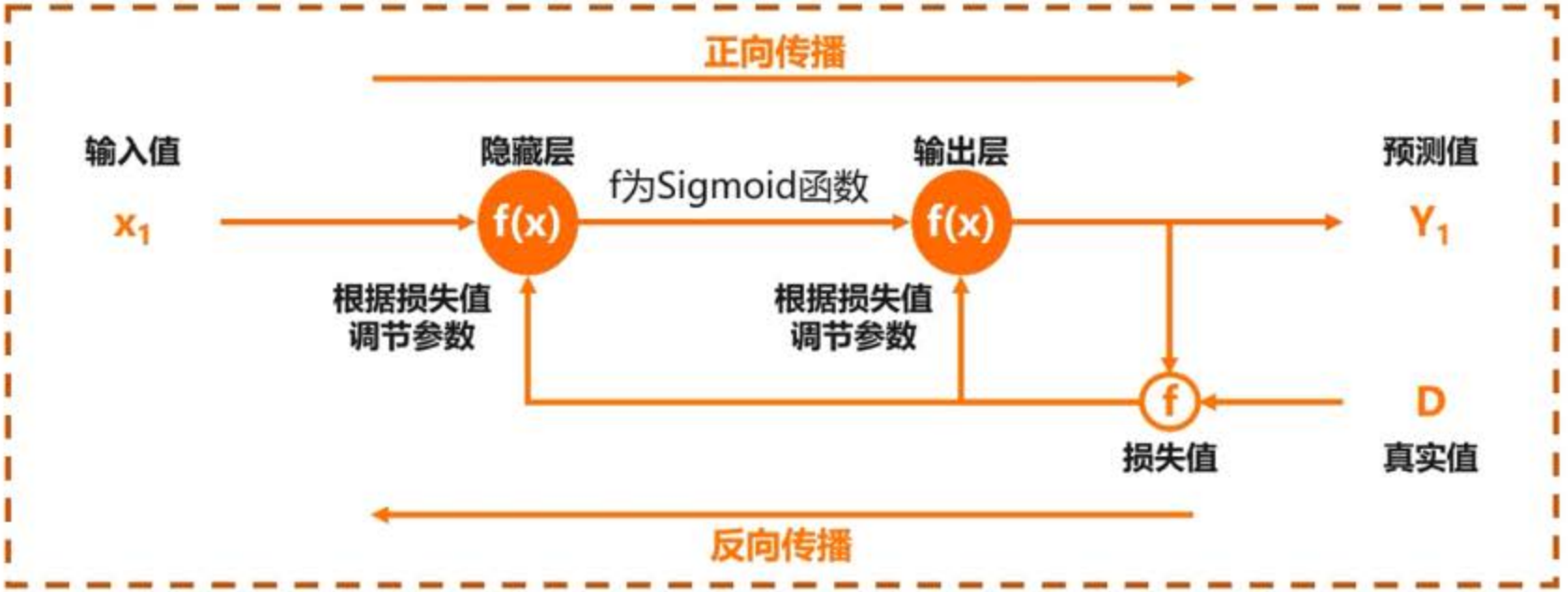
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

BP神经网络算法

阿里云

定义

- BP（Back Propagation）神经网络是一种按误差逆传播算法训练的多层神经网络；
- 正向传播求损失，反向传播回传误差；
- 根据误差信号修正每层的权重。

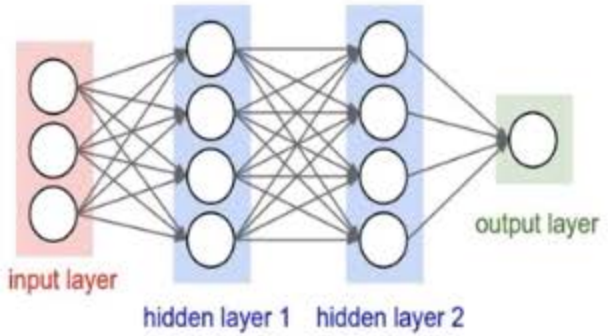


卷积神经网络概述

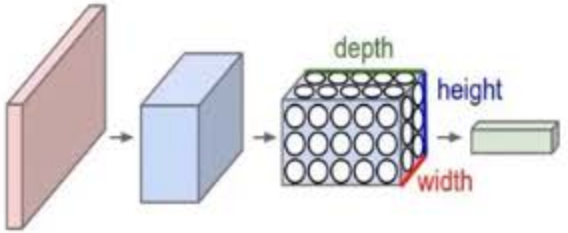
阿里云

卷积神经网络（Convolutional Neural Network，CNN）

- 一种带有卷积结构的深度神经网络，通过特征提取和分类识别完成对输入数据的判别；
- 在1989年提出，早期被成功用于手写字符图像识别；
- 2012年更深层次的AlexNet网络取得成功，此后卷积神经网络被广泛用于各个领域。



传统神经网络



卷积神经网络

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络——输入层

阿里云

输入层（Input Layer） 即接收数据的输入，可以处理多维数据，也能对输入特征进行标准化处理，有利于提升卷积神经网络的学习效率和表现。

进行预处理的主要原因

- 输入数据单位不一致
- 数据范围影响
- 激活函数值域影响

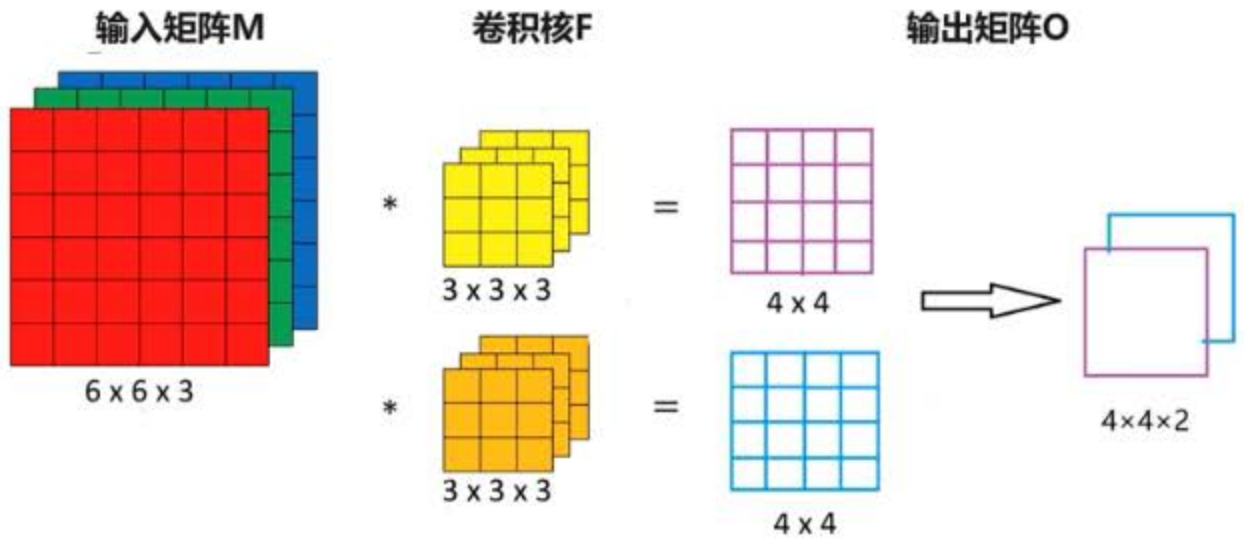
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络——卷积层

阿里云

卷积层（Convolutional Layer）

- 提取一个局部区域的特征，不同的卷积核相当于不同的特征提取器；
- 主要应用在图像处理上，而图像为二维结构，因此为了更充分地利用图像的局部信息；
- 通常将神经元组织为三维结构的神经层，其大小为高度 M× 宽度N× 深度 D，由 D 个 $M \times N$ 大小的特征映射构成。



卷积神经网络——池化层

阿里云

池化层（Pooling Layer）

- 包含预设定的池化函数；
- 将特征图中单个点的结果替换为其相邻区域的特征图统计量；
- 对数据进行降维，减少数据特征，减少网络参数和运算次数，避免过拟合，常用方法有**最大值池化**和**均值池化**。

原始数据

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

最大值池化

3	4
7	8

均值池化

2	3
6	7

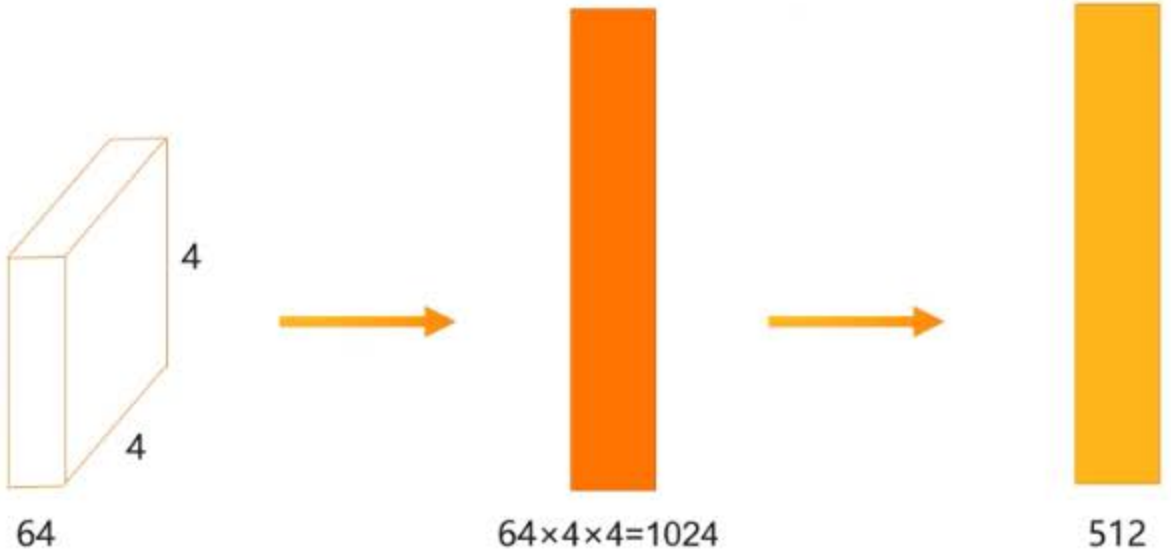
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络——全连接层

阿里云

全连接层（Fully-connected Layer）

- 神经元排成一列，这些神经元与前一层神经元通过权值互连，呈全连接结构；
- 等价于传统前馈神经网络中的隐含层，通常位于卷积神经网络的最后部分，并只向其它全连接层传递信号。



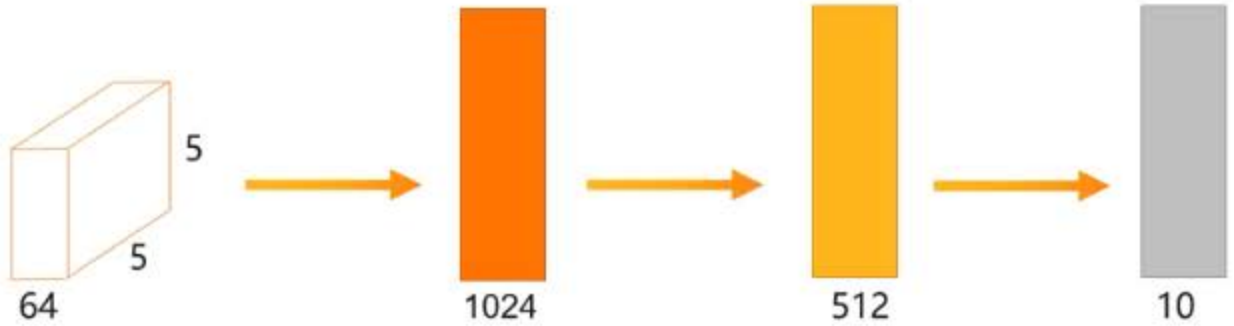
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络——输出层

阿里云

输出层（Output Layer）通常是全连接层，因此其结构和工作原理与传统前馈神经网络中的输出层相同。

假设对于10分类问题，输出层如下



分类标签

中心坐标、大小、分类

每个像素分类结果

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

经典卷积神经网络

阿里云

开山之作——LeNet-5

- 于上世纪90年代提出，第一个卷积神经网络；
- 共包含7层网络结构，分别为2个卷积层、2个池化层、2个全连接层和1个输出层；
- 卷积核大小全部为5*5。

王者归来——AlexNet

- 在2012年的ImageNet ILSVRC竞赛中以Top5错误率16.4%夺得冠军；
- 共包含8层网络结构，分别为5层卷积层和3层全连接层；
- 卷积核大小分别为11*11、5*5和3*3。

里程碑式创新——ResNet

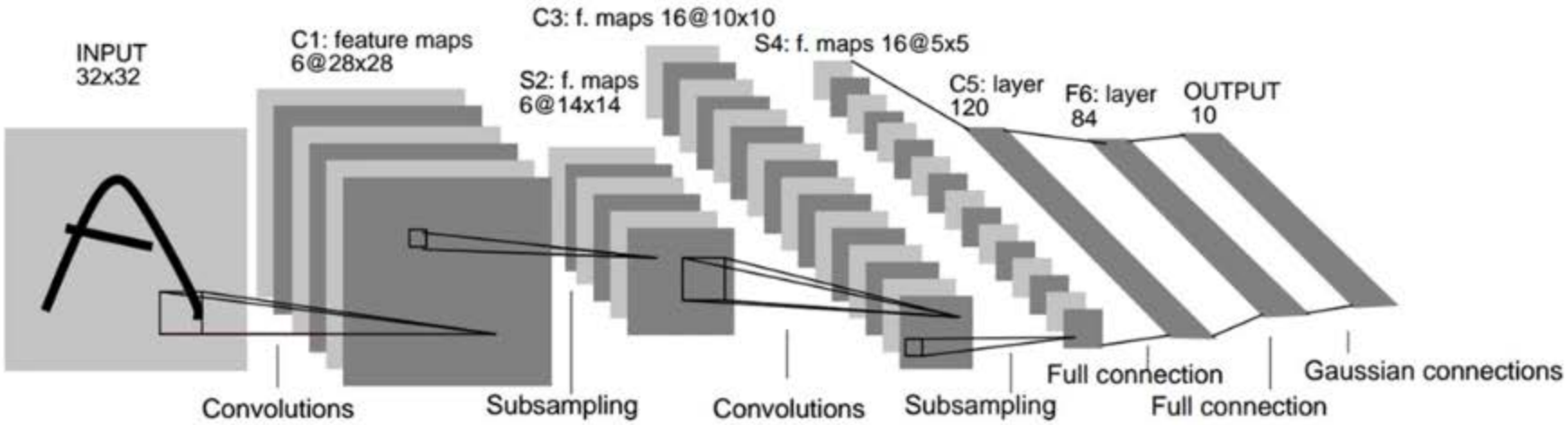
- 残差网络（Residual Network）是ILSVRC2015的胜利者；
- 共包含152层网络结构，分别为151层卷积层和1层全连接层；
- 它使用了跳跃链接，并大量使用了批量归一化。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

LeNet-5

阿里云

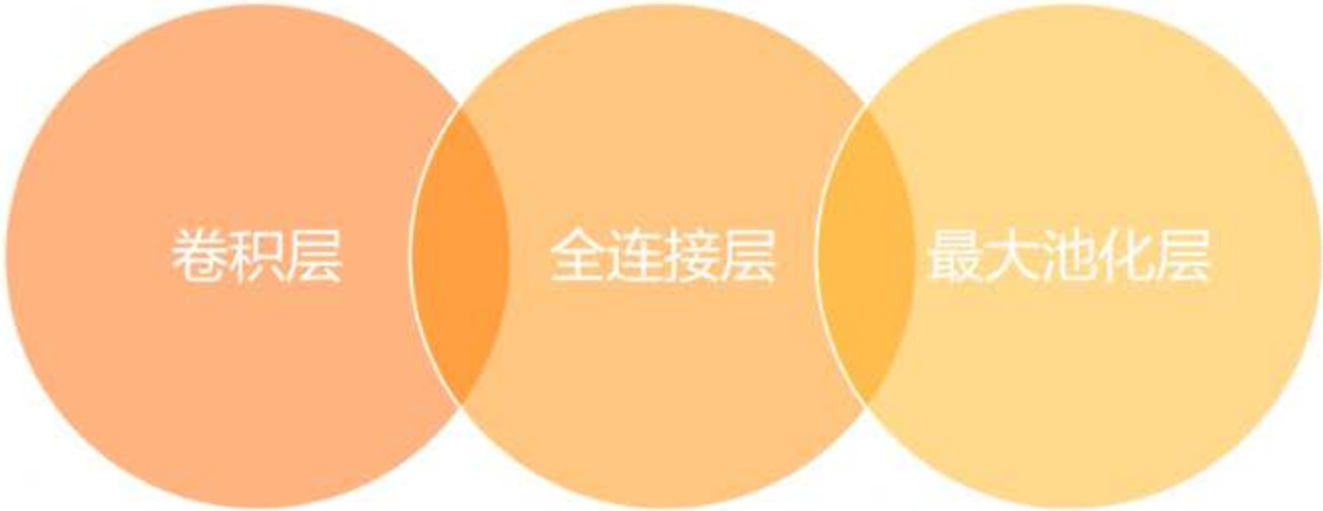
LeNet-5虽然提出的时间比较早，但它是一个非常成功的神经网络模型。基于LeNet-5的手写数字识别系统在20世纪90年代被美国很多银行使用，用来识别支票上面的手写数字。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

AlexNet

AlexNet是第一个现代深度卷积网络模型，其首次使用了很多现代深度卷积网络的技术方法，比如使用GPU进行并行训练，采用ReLU作为非线性激活函数，使用Dropout防止过拟合，使用数据增强来提高模型准确率等。

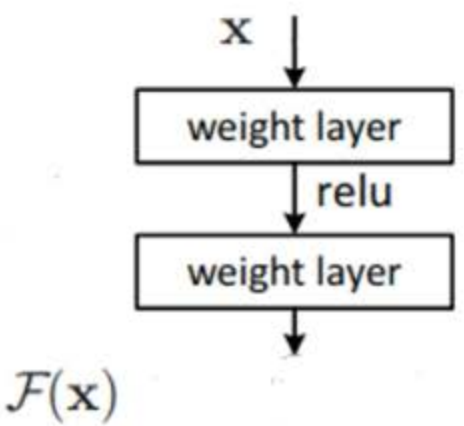


AlexNet网络结构

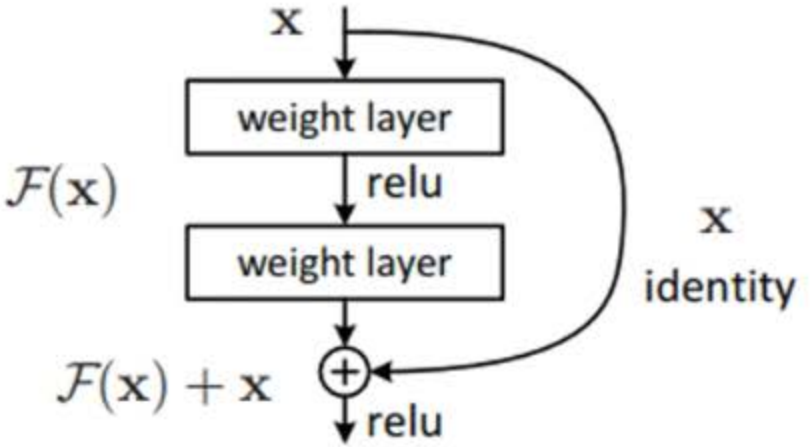
ResNet

阿里云

ResNet通过使用残差单元成功训练出了152层的神经网络。残差网络的特点是容易优化，并且能够通过增加相当的深度来提高准确率。其内部的残差块使用了跳跃连接，缓解了在深度神经网络中增加深度带来的梯度消失问题。



Original block



Residual learning: a building block.

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络的应用

阿里云

一维卷积

序列模型、自然语处理模型...

文本分类

将不同领域的文本
内容进行归类存放。

语音识别

通过输入语音进行
识别并转为文本内容。

机器翻译

通过计算机将一种语言
转为另外一种语言。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络的应用

阿里云

二维卷积

图像处理、计算机视觉领域...

垃圾分类

通过图像分类技术

判断垃圾所属类别。

车牌识别

通过图像检测技术

识别车辆车牌信息。

图像搜索

通过图像搜索技术

给用户提供更多检索信息。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

卷积神经网络的应用

阿里云

三维卷积

医学领域

CT影像的检测，协助医生快速准确的对患者病患进行初步判断。

视频处理领域

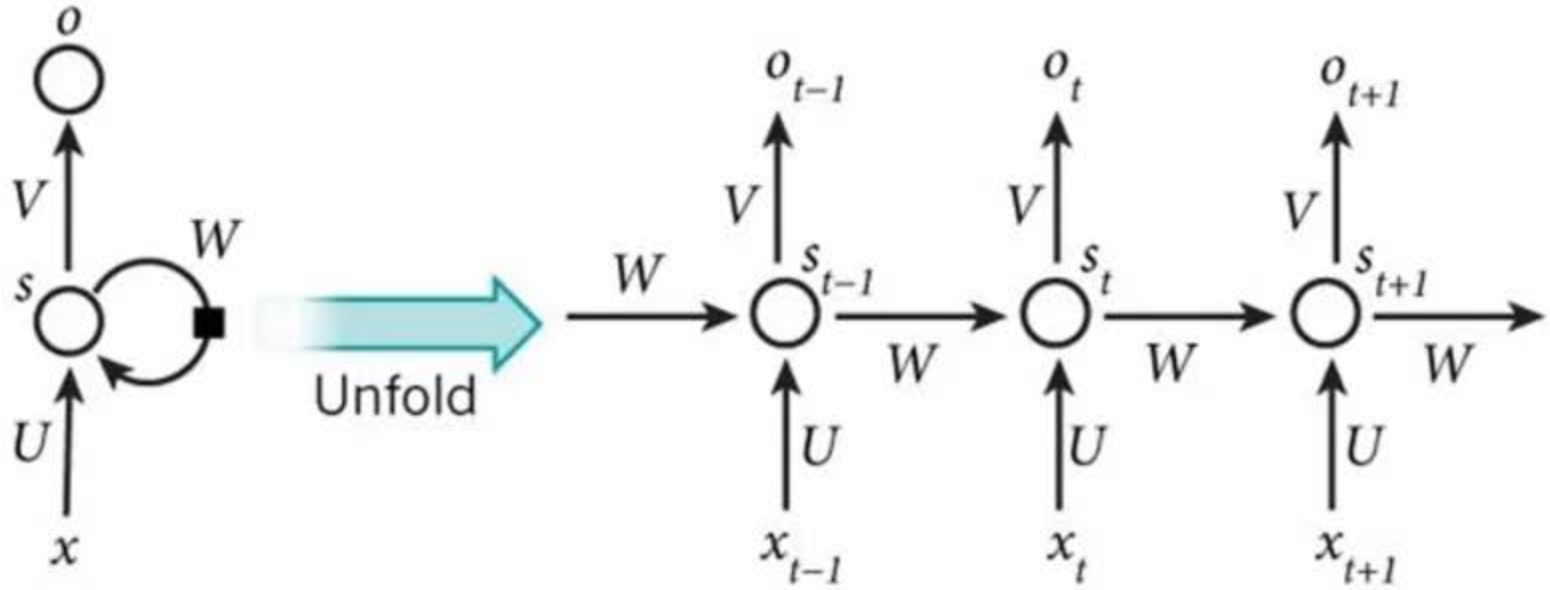
在家庭监控视频中自动识别是否有婴儿或宠物出现在画面中，并可以对婴儿表情进行识别，用AI守护家庭安全。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

循环神经网络概述

阿里云

循环神经网络（Recurrent Neural Network，RNN）是一类以序列（sequence）数据为输入，在序列的演进方向进行递归（recursion）且所有节点（循环单元）按链式连接的递归神经网络（recursive neural network）。



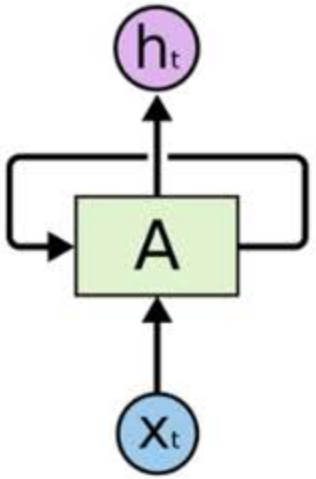
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

循环神经网络的结构（1/3）

阿里云

循环神经网络结构

- 一个典型的RNN网络包含一个输入x，一个输出h和一个神经网络单元A；
- 与普通的神经网络不同的是，RNN网络的神经网络单元A不仅仅与输入和输出存在联系，其与自身也存在一个回路；
- 这种网络结构就揭示了RNN的实质：上一个时刻的网络状态信息将会作用于下一个时刻的网络状态。



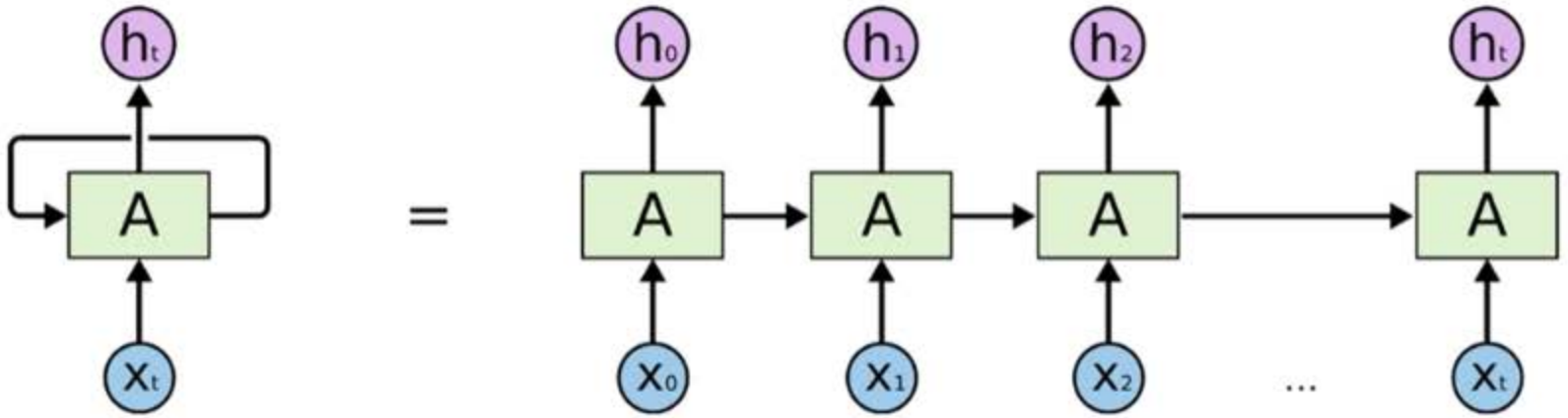
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

循环神经网络的结构（2/3）

阿里云

循环神经网络结构——拓展

- 将RNN的自循环结构展开，像是将同一个网络复制并连成一条线的结构，将自身提取的信息传递给下一个继承者；
- 这种链式的结构揭示了RNN与序列和列表类型的数据密切相关。



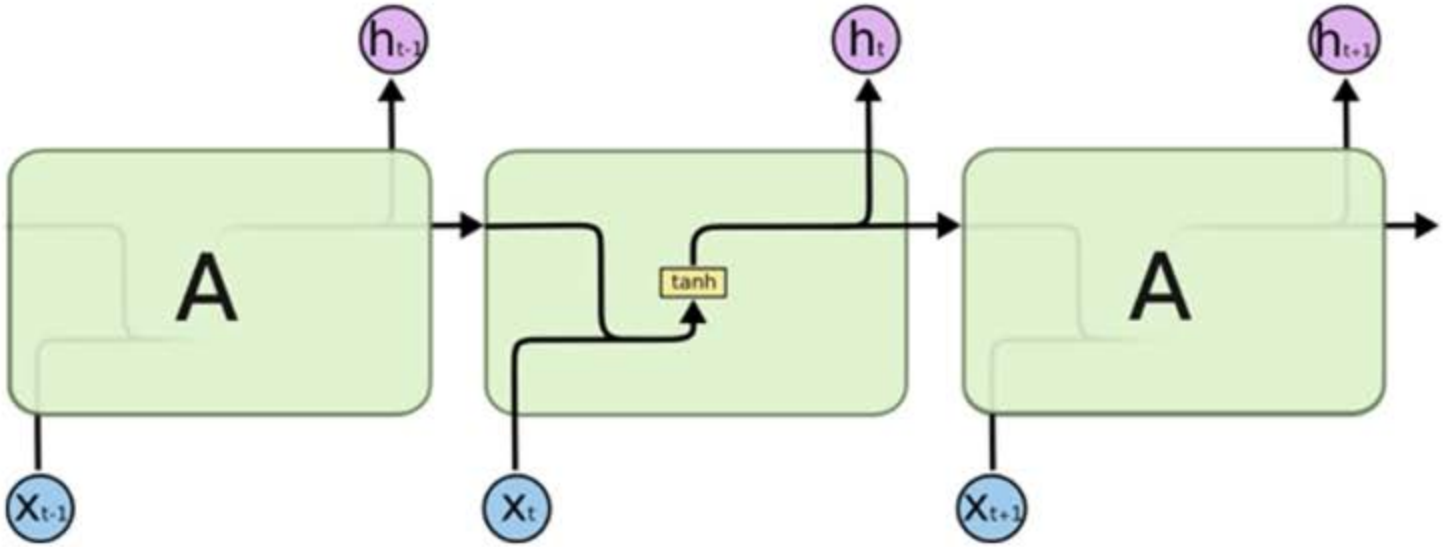
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

循环神经网络的结构（3/3）

阿里云

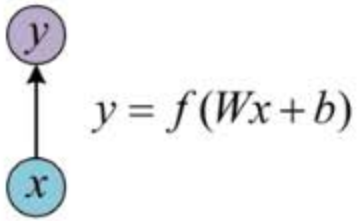
循环神经网络的神经元结构

- 下图依然是一个RNN神经网络的时序展开模型，中间t时刻的网络模型揭示了RNN的结构，原始的RNN网络的内部结构非常简单。神经元A在t时刻的状态仅仅是t-1时刻神经元状态与t时刻网络输入的双曲正切函数的值，这个值不仅仅作为该时刻网络的输出，也作为该时刻网络的状态被传入到下一个时刻的网络状态中，这个过程叫做RNN的正向传播。

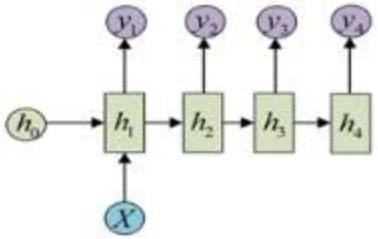


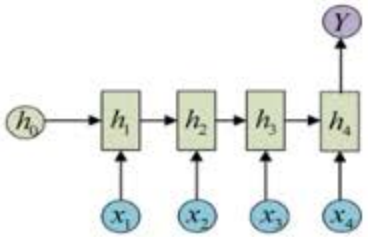
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

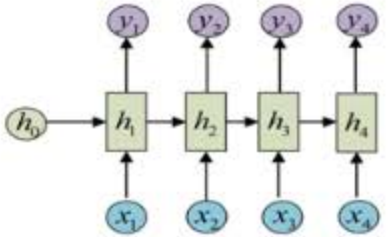
循环神经网络的类别



$$y = f(Wx + b)$$







01

最基本的单层网络，输入是x，经过变换Wx+b和激活函数f得到输出y。

一对一

02

能够处理的问题包含从图像生成文字，输入的X为图像的特征，输出的y序列为一段句子。

一对多

03

常用于处理序列分类问题，如输入一段文字判别所属的类别等。

多对一

04

最经典的RNN结构，输入和输出序列必须是等长的。

多对多

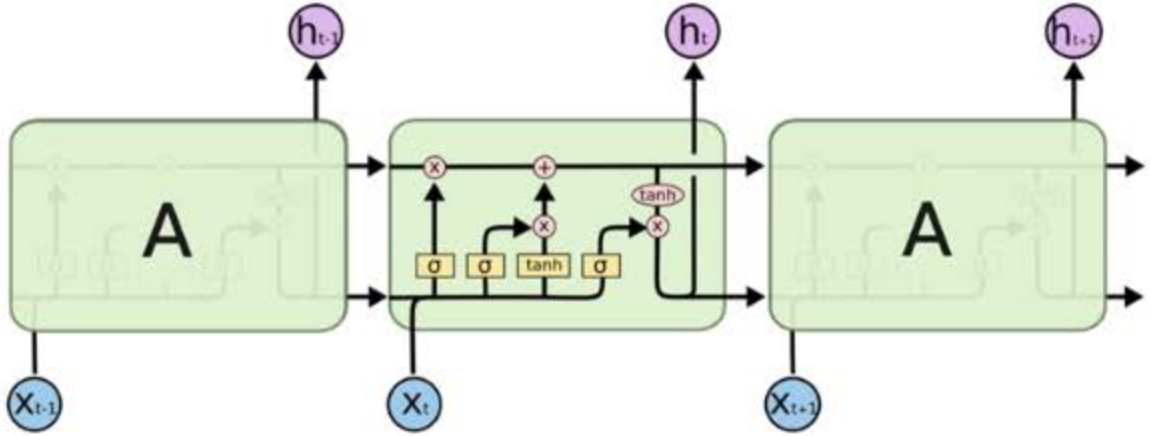
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

长短期记忆网络LSTM

阿里云

长短期记忆网络（Long Short Term Memory，LSTM）

- 循环神经网络最著名和成功的拓展；
- 增强循环神经网络的学习能力，缓解网络的梯度消失等问题；
- 对有价值的信息进行长期记忆，减小循环神经网络的学习难度。



LSTM神经网络结构

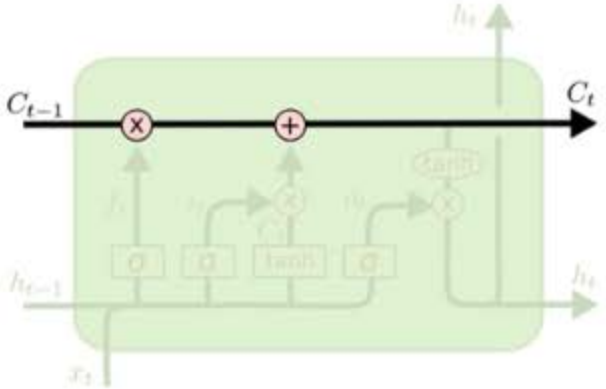
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

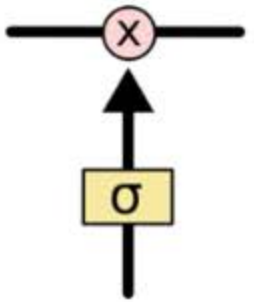
LSTM的核心

阿里云

LSTM的核心

- 细胞状态，用贯穿细胞的水平线表示，细胞状态像传送带一样；
- 贯穿整个细胞却只有很少的分支，这样能保证信息不变的流过整个RNN；
- LSTM网络能通过一种被称为门的结构对细胞状态进行删除或者添加信息，门能够有选择性地决定让哪些信息通过；
- 门的结构很简单，就是一个Sigmoid层和一个点乘操作的组合；
- 因为Sigmoid层的输出是0-1的值，这代表有多少信息能够流过Sigmoid层；
- 一个LSTM里面包含三个门来控制细胞状态。



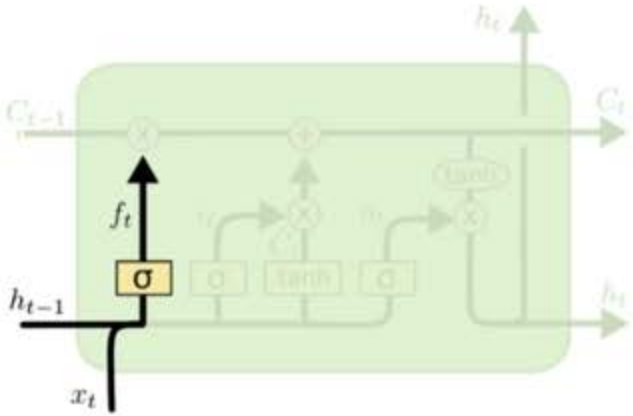


阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

LSTM神经元（1/4）

阿里云

LSTM的第一步就是决定细胞状态需要丢弃哪些信息，通过一个称为**忘记门**的Sigmoid单元来处理的，通过查看h_{t-1}和x_t信息来输出一个0-1之间的向量，元胞状态C_{t-1}中的每个数字都要与该系数相乘；该向量里面的0-1值表示细胞状态中的哪些信息保留或丢弃多少，0表示不保留，1表示都保留。

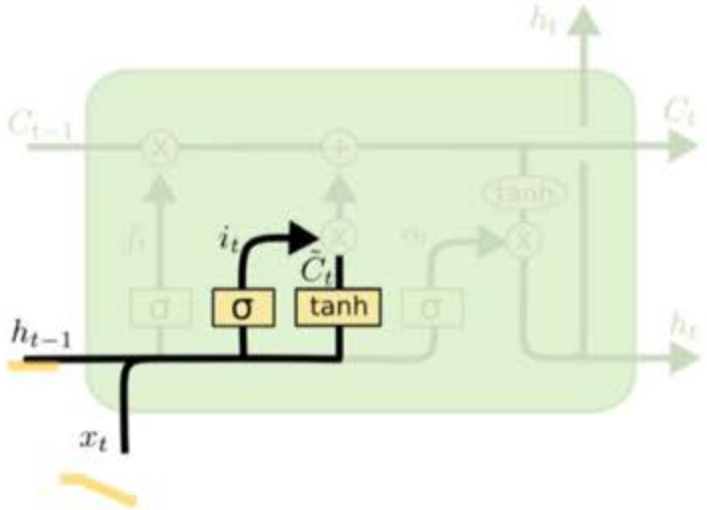


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

LSTM神经元（2/4）

阿里云

LSTM的第二步是决定给细胞状态添加哪些新的信息，分为两个步骤，首先利用h_{t-1}和x_t通过一个称为**输入门**的操作来决定更新哪些信息，接着利用h_{t-1}和x_t通过一个Tanh层得到新的候选细胞信息C_t，这些信息可能会被更新到细胞信息中。



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

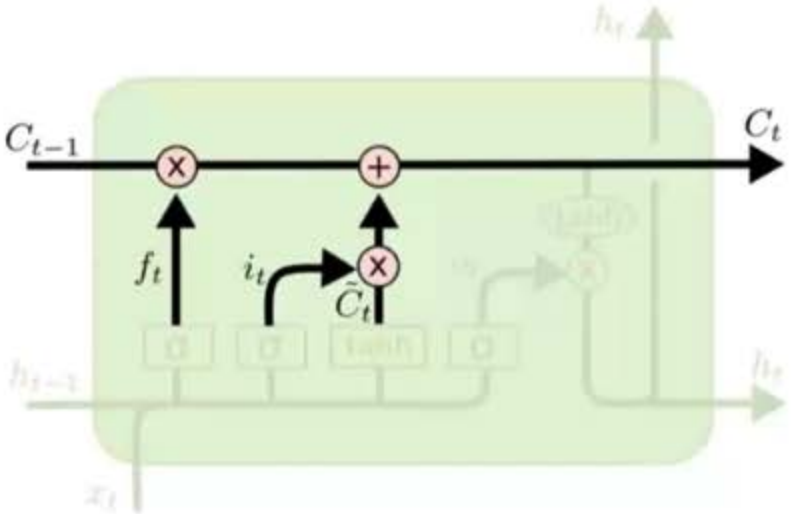
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

LSTM神经元（3/4）

阿里云

LSTM的第三步是更新旧的细胞信息C_{t-1}，变为新的细胞信息C_t，更新的规则就是通过忘记门选择忘记旧细胞信息的一部分，通过输入门选择添加候选细胞信息的一部分得到新的细胞信息C_t。



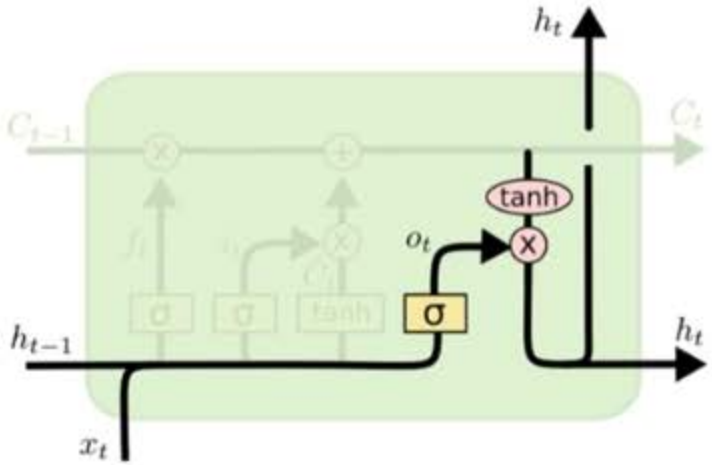
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

LSTM神经元（4/4）

阿里云

LSTM的第四步需要根据输入的h_{t-1}和x_t来判断输出细胞的哪些状态特征，首先将输入经过一个称为**输出门**的Sigmoid层得到判断条件，接着将细胞状态经过Tanh层得到一个-1~1之间值的向量，该向量与输出门得到的判断条件相乘就得到了最终该RNN单元的输出。



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

循环神经网络的应用

阿里云

循环神经网络主要用于序列数据的处理问题，在自然语言处理领域就是一个典型的应用场景。

文本生成	机器翻译	语音识别	生成图像描述	视频标记
根据给出的前后文，预测空格中间的词。	将一种文本语言转为另外一种语言的文本，词的顺序直接影响翻译的结果。	根据输入的音频判断对应的文字内容，生成相应文本。	类似看图说话，给定一张图，能够描述出图片中的内容。	首先将视频分解为图片，接着使用图像描述来描述图片的内容。

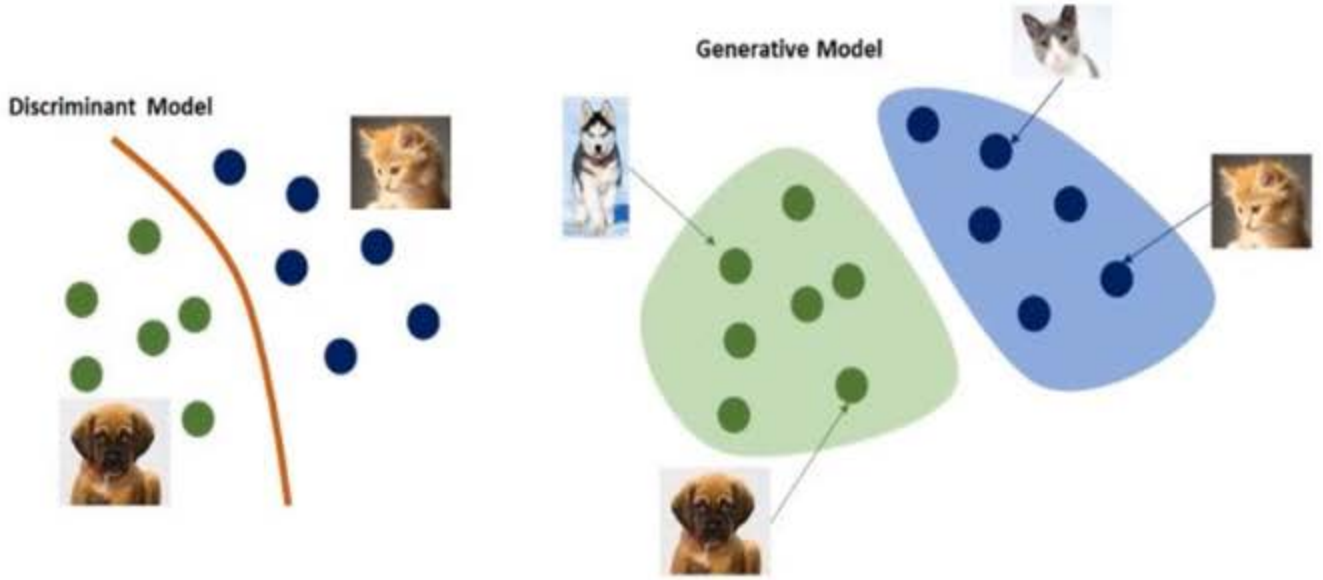
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

生成对抗网络定义

阿里云

生成对抗网络（Generative Adversarial Network，GAN）

- 一种深度学习模型，通过判别模型（Discriminative Model）和生成模型（Generative Model）的相互博弈学习，生成接近真实数据的数据分布或对输入数据进行分类；
- 近年来复杂分布上无监督学习最具前景的方法之一。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

生成对抗网络结构

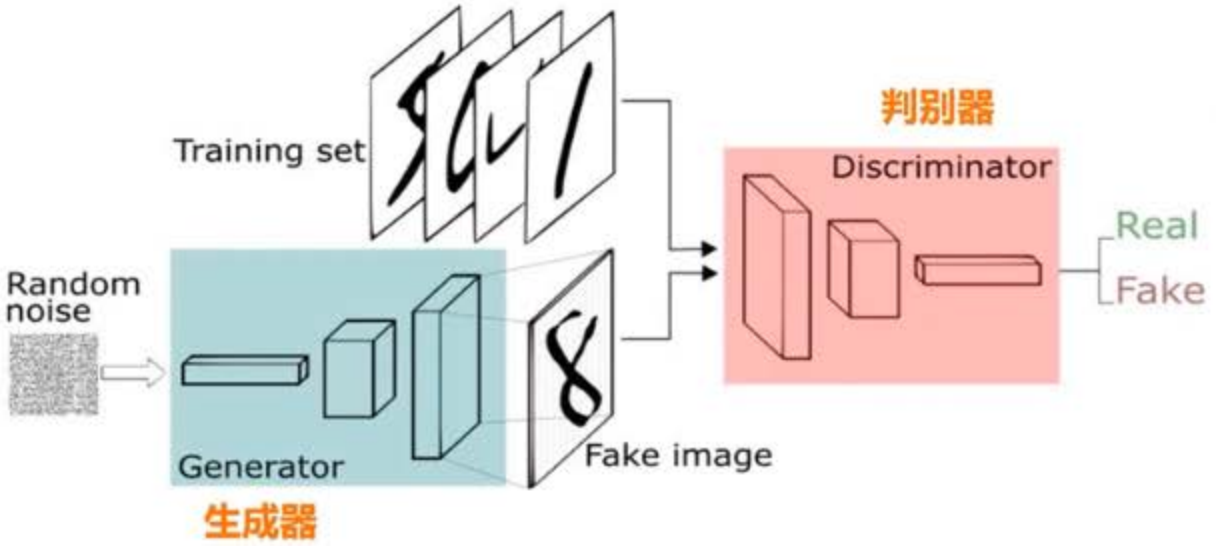
阿里云

- **生成器**

在给定输入数据时，理解输入，生成类似的输出。

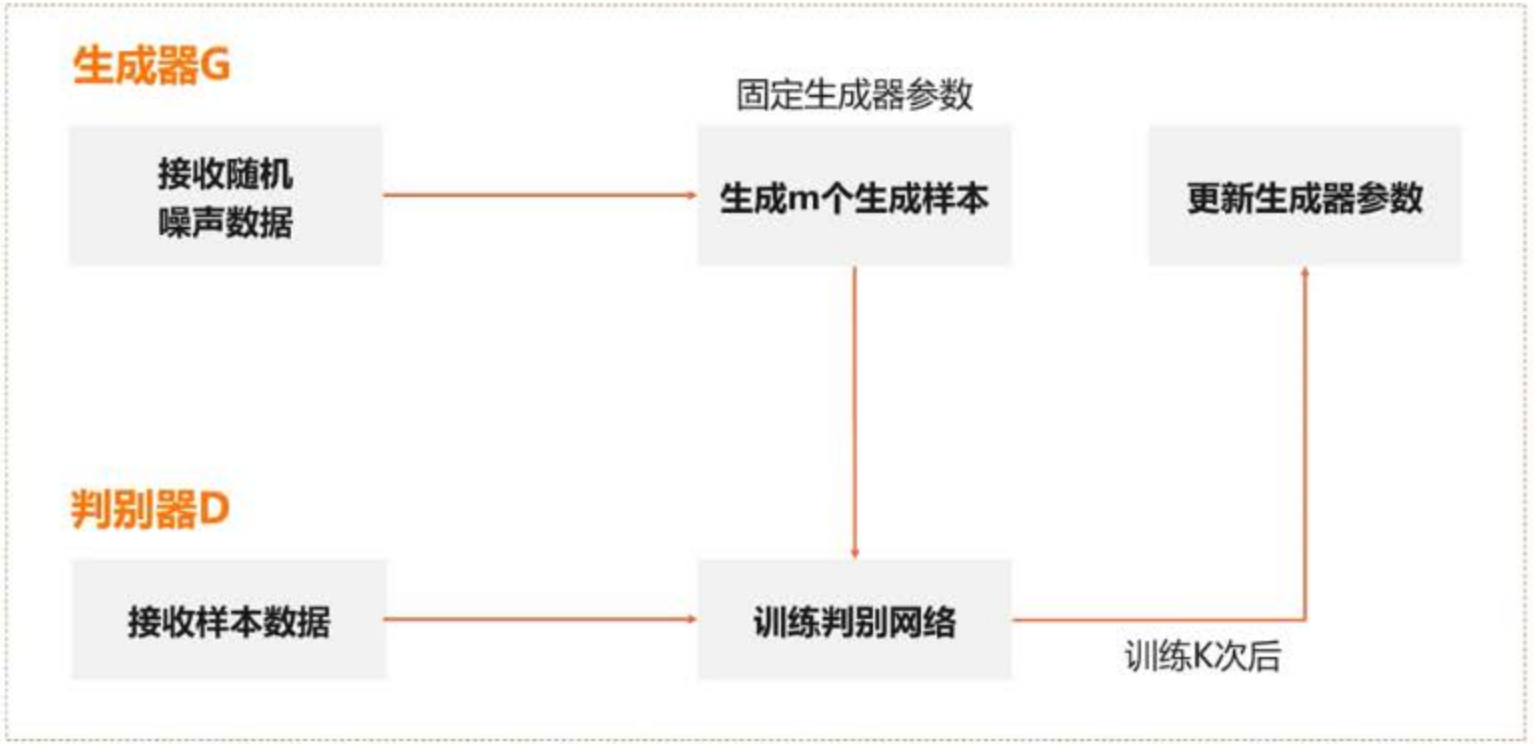
- **判别器**

在给定输入数据时，将输入数据正确地分类。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

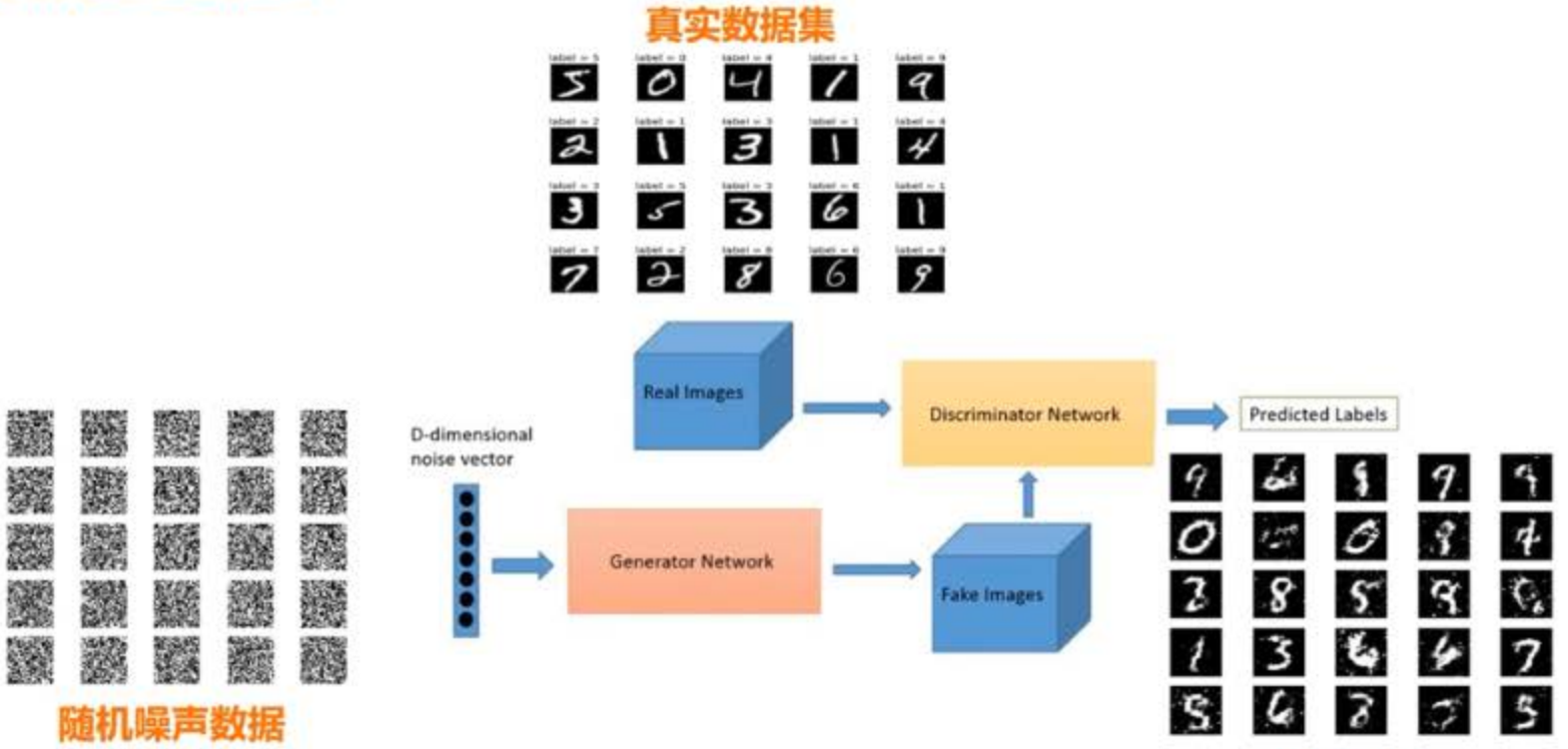
生成对抗网络工作流程



N轮迭代，直至判别器分辨不出样本是生成的还是真实的

生成对抗网络应用示例

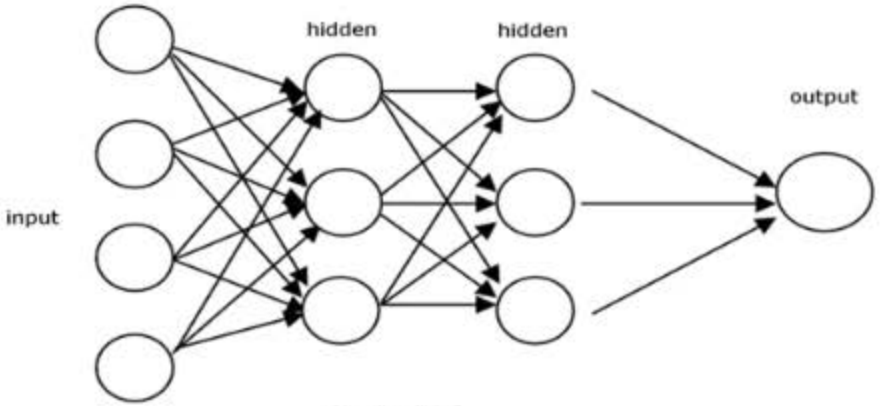
- 生成对抗网络生成图片



生成对抗网络对比

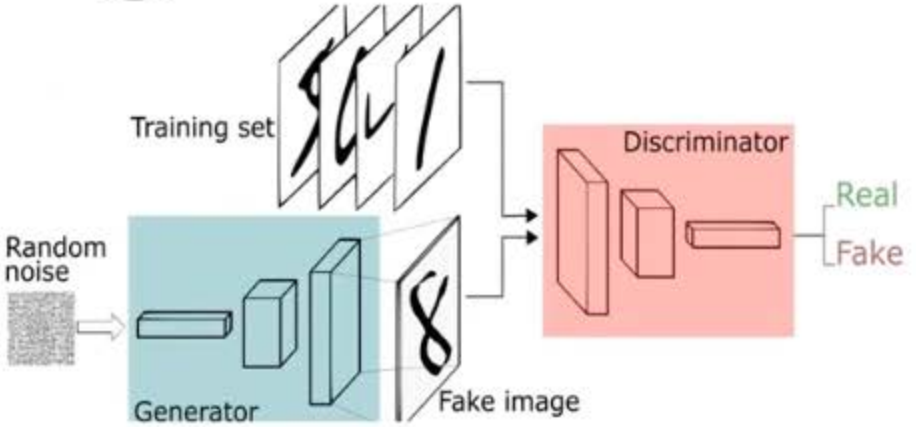
传统神经网络

- 根据输入数据的特征，预测输入数据的标签；
- 一个神经网络是一个训练模型；
- 网络训练时，依赖于输入数据样本更新参数梯度。



生成对抗网络

- 根据输入数据的标签，生成接近真实的输入分布；
- 一个网络包含生成器和判别器两个模型；
- 网络训练时，生成器模型梯度更新依赖于判别器模型；
- 生成器和判别器可以是CNN、RNN神经网络。



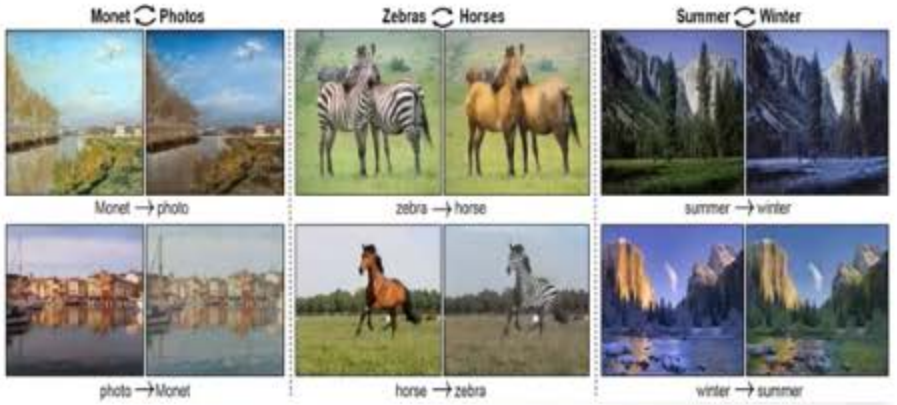
生成对抗神经网络类别

阿里云

生成对抗网络类别	特征
标准生成对抗网络（GAN）	2014年由Ian Goodfellow等人提出，采用多层感知机作为网络结构
深度卷积生成对抗网络（DCGAN）	采用卷积神经网络作为GAN的网络结构，用于生产图片
条件生成对抗网络（CGAN）	针对多类别的输入数据，可以指定生成数据的分类类别
风格迁移生成对抗网络（CycleGAN）	包含两个生成器和两个判别器，实现图像到图像的转换
基于风格的生成对抗网络（StyleGAN）	关注GAN网络的稳定性，提高GAN对生成图像的精确控制能力
大型生成对抗网络（BigGAN）	由DeepMind提出，用强大的深度学习技术训练GAN网络，生成效果高度逼真

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

生成对抗网络应用场景（1/2）



图像生成

生成真假难辨的高分辨率图像

图像转换

实现不同风格的图像转换

生成对抗网络应用场景（2/2）

阿里云

人脸合成

- 根据一张人脸的图像，合成出不同角度的人脸图像
- 用于提高人脸识别的精度

半监督学习

- 缺少训练数据的场景下，利用生成对抗网络生成数据加入到训练集中
- 提升训练模型精度

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER