深度学习基本知识

孙千惠 18200100078 百度飞桨 4 队

一、深度学习发展历史

深度学习是近十年来人工智能领域取得的重要突破。它在语音识别、自然语言处理、计算机视觉、图像与视频分析、多媒体等诸多领域的应用取得了巨大成功。现有的深度学习模型属于神经网络。神经网络的起源可追溯到 20 世纪 40 年代，曾经在八九十年代流行。神经网络试图通过模拟大脑认知的机理解决各种机器学习问题。1986 年，鲁梅尔哈特（Rumelhart）、欣顿（Hinton）和威廉姆斯（Williams）在《自然》杂志发表了著名的反向传播算法用于训练神经网络，该算法直到今天仍被广泛应用。

神经网络有大量参数，经常发生过拟合问题，虽然其识别结果在训练集上准确率很高，但在测试集上效果却很差。这是因为当时的训练数据集规模都较小，加之计算资源有限，即便是训练一个较小的网络也需要很长的时间。与其他模型相比，神经网络并未在识别准确率上体现出明显的优势。

因此更多的学者开始采用支持向量机、Boosting、最近邻等分类器。这些分类器可以用具有一个或两个隐含层的神经网络模拟，因此被称为浅层机器学习模型。在这种模型中，往往是针对不同的任务设计不同的系统，并采用不同的手工设计的特征。例物体识别采用尺度不变特征转换（Scale Invariant Feature Transform，SIFT），人脸识别采用局部二值模式（Local Binary Patterns，LBP），行人检测采用方向梯度直方图（Histogram of Oriented Gradient，HOG）特征。

2006 年，欣顿提出了深度学习。之后深度学习在诸多领域取得了巨大成功，受到广泛关注。神经网络能够重新焕发青春的原因有几个方面：首先，大规模训练数据的出现在很大程度上缓解了训练过拟合的问题。例如，ImageNet 训练集拥有上百万个有标注的图像。其次，计算机硬件的飞速发展为其提供了强大的计算能力，一个 GPU 芯片可以集成上千个核。这使得训练大规模神经网络成为可能。第三，神经网络的模型设计和训练方法都取得了长足的进步。例如，为了改进神经网络的训练，学者提出了非监督和逐层的预训练，使得在利用反向传播算法对网络进行全局优化之前，网络参数能达到一个好的起始点，从而在训练完成时能达到一个较好的局部极小点。

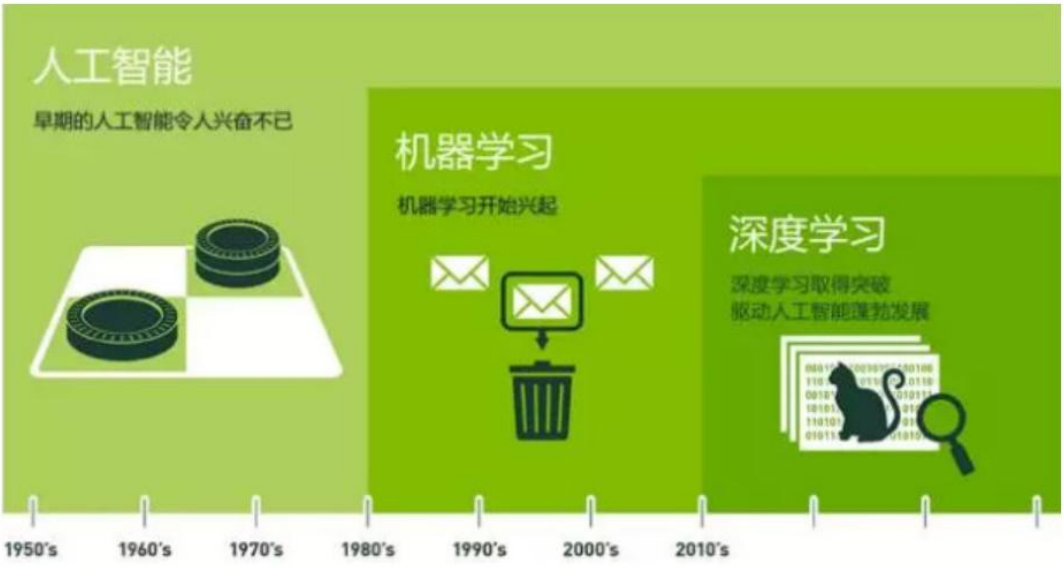
深度学习在计算机视觉领域最具影响力的突破发生在 2012 年，欣顿的研究小组采用深度学习赢得了 ImageNet 图像分类比赛的冠军。排名第 2 到第 4 位的小组采用的都是传统的计算机视觉方法、手工设计的特征，他们之间准确率的差别不超过 1%。欣顿研究小组的准确率超出第二名 10%以上，这个结果在计算机视觉领域产生了极大的震动，引发了深度学习的热潮。

计算机视觉领域另一个重要的挑战是人脸识别。有研究表明，如果只把不包

括头发在内的人脸的中心区域给人看，人眼在户外脸部检测数据库（Labeled Faces in the Wild，LFW）上的识别率是 97.53%。如果把整张图像，包括背景和头发给人看，人眼的识别率是 99.15%。经典的人脸识别算法 Eigenface 在 LFW 测试集上只有 60%的识别率。在非深度学习算法中，最高的识别率是 96.33%。目前深度学习可以达到 99.47%的识别率。

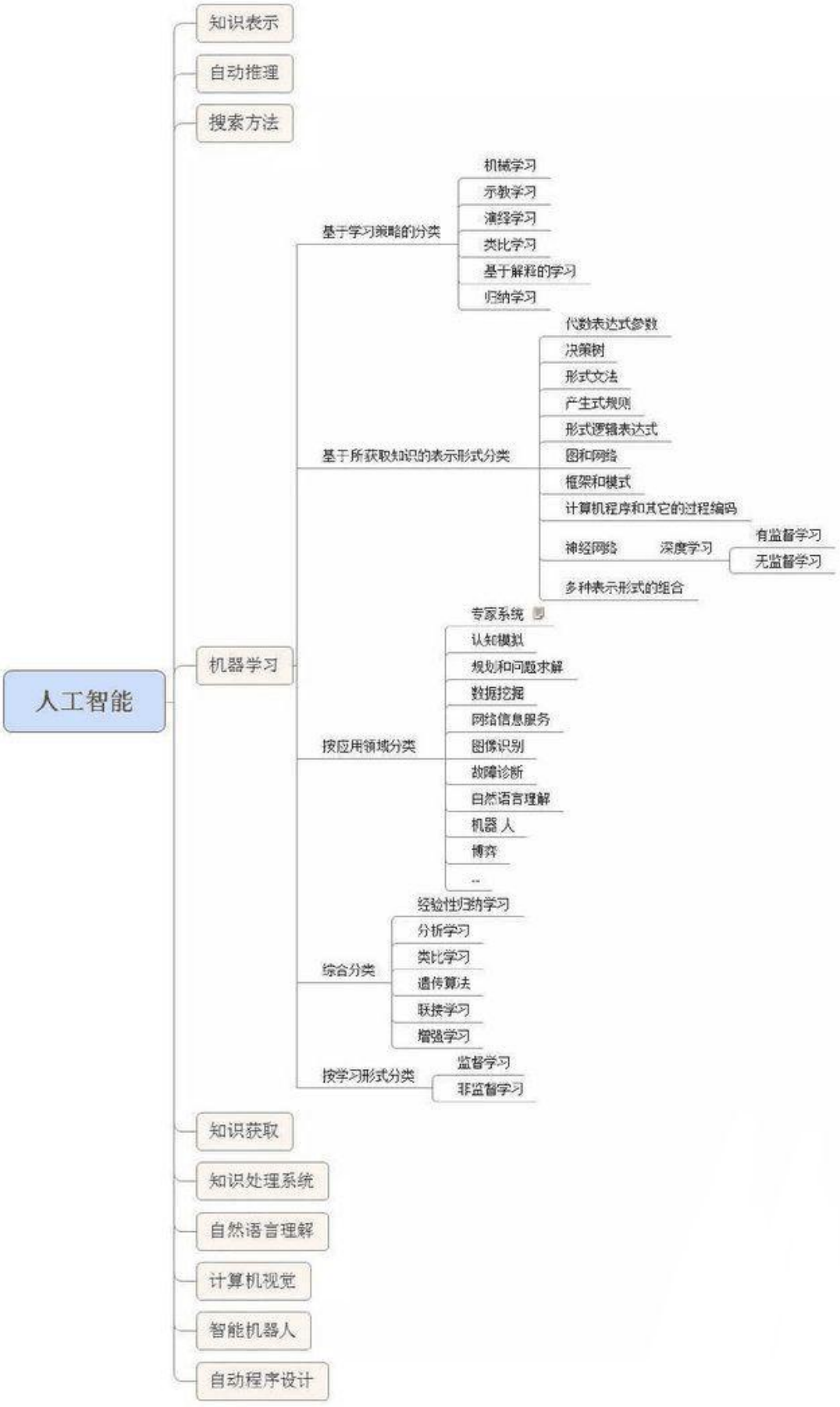
在欣顿的科研小组赢得 ImageNet 比赛冠军之后的 6 个月，谷歌和百度都发布了新的基于图像内容的搜索引擎。他们采用深度学习模型，应用在各自的数据上，发现图像搜索准确率得到了大幅度提高。百度在 2012 年成立了深度学习研究院，2014 年 5 月又在美国硅谷成立了新的深度学习实验室，聘请斯坦福大学著名教授吴恩达担任首席科学家。脸谱于 2013 年 12 月在纽约成立了新的人工智能实验室，聘请深度学习领域的著名学者、卷积网络的发明人雅恩·乐昆（Yann LeCun）作为首席科学家。2014 年月，谷歌抛出四亿美金收购了深度学习的创业公司 DeepMind。鉴于深度学习在学术界和工业界的巨大影响力，2013 年，《麻省理工科技评论》（MIT Technology Review）将其列为世界十大技术突破之首。

二、人工智能、机器学习和深度学习有什么区别和联系



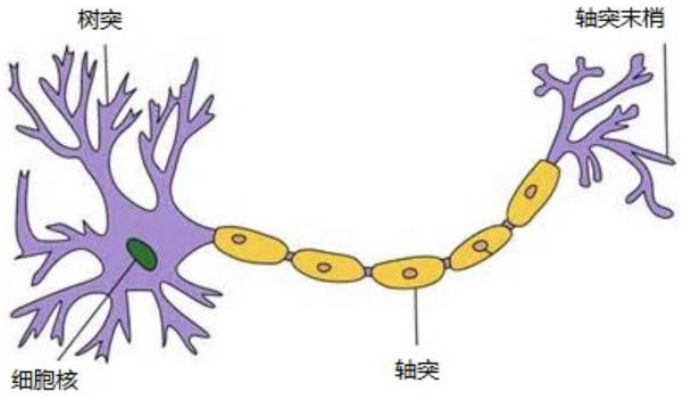
人工智能（Artificial Intelligence）是为机器赋予人的智能，让机器能像人一样处理问题，并且可以解决人类难以解决的问题。机器学习则是一种实现人工智能的方法，最基本的做法，是使用算法来解析数据、从中学习，然后对真实世界中的事件做出决策和预测。深度学习是一种实现机器学习的技术，它使得机器学习能够实现众多的应用，并拓展了人工智能的领域范围。

下图可以更为直观地看出这三者之间的关系。


人工智能
知识表示
自动推理
搜索方法
机器学习
基于学习策略的分类
机械学习
示教学习
演绎学习
类比学习
基于解释的学习
归纳学习
基于所获取知识的表示形式分类
代数表达式参数
决策树
形式文法
产生式规则
形式逻辑表达式
图和网络
框架和模式
计算机程序和其它的过程编码
神经网络
深度学习
有监督学习
无监督学习
多种表示形式的组合
按应用领域分类
专家系统
认知模拟
规划和问题求解
数据挖掘
网络信息服务
图像识别
故障诊断
自然语言理解
机器人
博弈
...
综合分类
经验性归纳学习
分析学习
类比学习
遗传算法
联接学习
增强学习
按学习形式分类
监督学习
非监督学习
知识获取
知识处理系统
自然语言理解
计算机视觉
智能机器人
自动程序设计

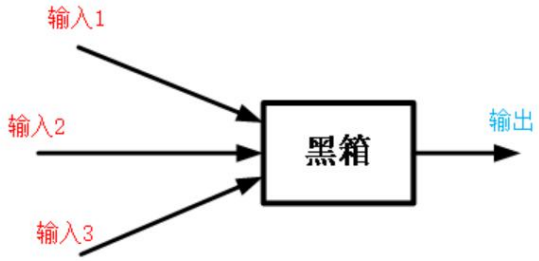
三、神经元、单层感知机、多层感知机

1、神经元



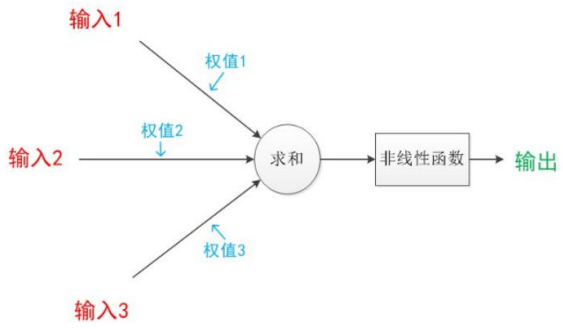
一个神经元通常具有多个树突，主要用来接受传入信息；而轴突只有一条，轴突尾端有许多轴突末梢可以给其他多个神经元传递信息。轴突末梢跟其他神经元的树突产生连接，从而传递信号。这个连接的位置在生物学上叫做“突触”。突触之间的交流通过神经递质实现。

下面对上面的这个模型进行抽象处理。首先考虑到神经元结构有多个树突，一个轴突可将其抽象为下图的黑箱结构：



神经元黑箱模型

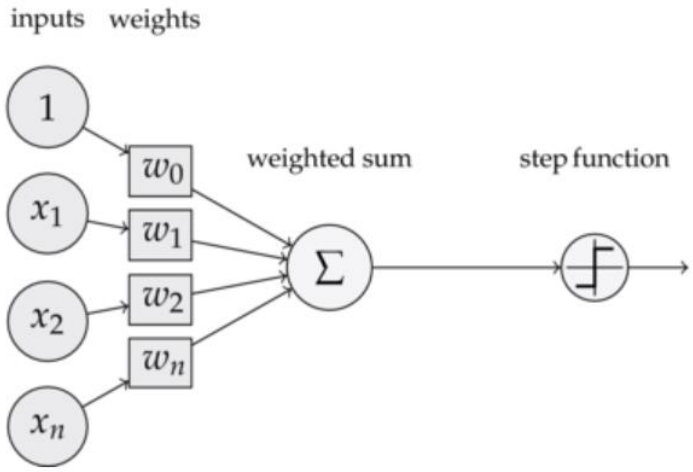
但是黑箱结构有诸多不便，首先是不知道黑箱中的函数结构就不能为我们所用，其次是输入输出与黑箱的关系也无法量化。因此考虑将上述结构简化，首先把树突到细胞核的阶段简化为线性加权的过程（当然了，该过程也有可能是非线性的，但是我们可以把其非线性过程施加到后面的非线性函数以及多层网络结构中），其次把突触之间的信号传递简化为对求和结果的非线性变换，那么上述模型就变得清晰了：



2、单层感知机

上面我们介绍的神经元的基本模型实际就是一个感知机的模型，该词最早出现于 1958 年，计算科学家 Rosenblatt 提出的由两层神经元组成的神经网络。

对前面的模型进一步符号化，如下图所示：



可以看到，感知机的基本模型包括：

• **输入**：x_1，x_2，...，x_n 实际可能比这更多,此处添加了一个偏置 1，是为了平衡线性加权函数总是过零点的问题。

• **权值**：对应于每个输入都有一个加权的权值 w_1，w_2，...，w_n

• **激活函数**：激活函数 f 对应于一个非线性函数，其选择有很多，本文后面会详细介绍

• **输出** y：由激活激活函数进行处理后的结果，往往是区分度较大的非连续值用于分类。

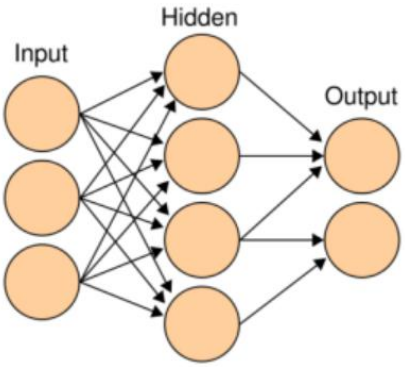
$$y = f(w_0 + x_1 * w_1 + x_2 * w_2 + \cdots + x_n * w_n)$$

考虑向量的点乘过程，上式又可以简化为：

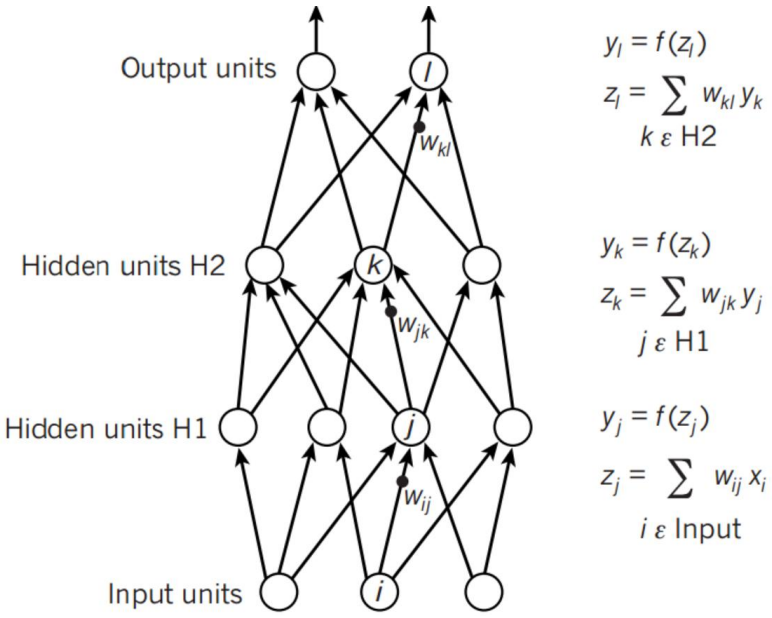
$$y = f(W \bullet x + w_0)$$

3、多层感知机

多层感知器（MLP，Multilayer Perceptron）是一种前馈人工神经网络模型，其将输入的多个数据集映射到单一的输出的数据集上。



四、什么是前向传播（图文）



如图所示，这里讲得已经很清楚了，前向传播的思想比较简单。

举个例子，假设上一层结点 i,j,k,…等一些结点与本层的结点 w 有连接，那么结点 w 的值怎么算呢？就是通过上一层的 i,j,k 等结点以及对应的连接权值进行加权和运算，最终结果再加上一个偏置项（图中为了简单省略了），最后在通过一个非线性函数（即激活函数），如 ReLu，sigmoid 等函数，最后得到的结果就是本层结点 w 的输出。

最终不断的通过这种方法一层层的运算，得到输出层结果。

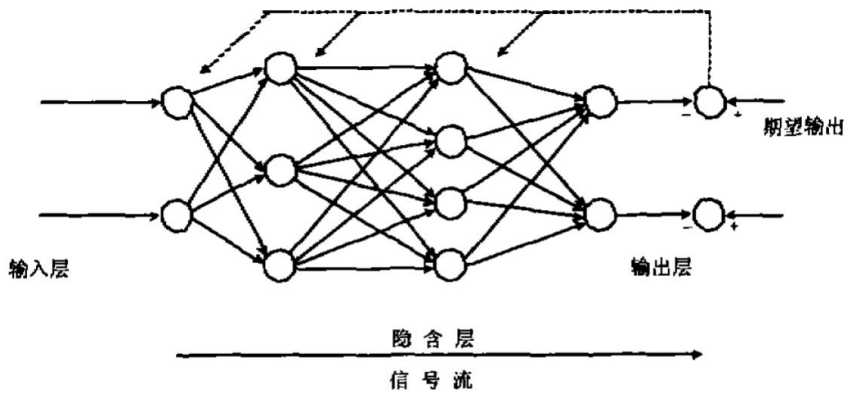
对于前向传播来说，不管维度多高，其过程都可以用如下公式表示：

$$a^2 = \sigma(z^2) = \sigma(a^1 * W^2 + b^2)$$

五、什么是反向传播（图文）

反向传播算法，简称 BP 算法，适合于多层神经元网络的一种学习算法，它建立在梯度下降法的基础上。BP 网络的输入输出关系实质上是一种映射关系：一个 n 输入 m 输出的 BP 神经网络所完成的功能是从 n 维欧氏空间向 m 维欧氏空间中一有限域的连续映射，这一映射具有高度非线性。它的信息处理能力来源于简单非线性函数的多次复合，因此具有很强的函数复现能力。这是 BP 算法得以应用的基础。

BP 算法(即误差反向传播算法)适合于多层神经元网络的一种学习算法，它建立在梯度下降法的基础上。BP 网络的输入输出关系实质上是一种映射关系：一个 n 输入 m 输出的 BP 神经网络所完成的功能是从 n 维欧氏空间向 m 维欧氏空间中一有限域的连续映射，这一映射具有高度非线性。它的信息处理能力来源于简单非线性函数的多次复合，因此具有很强的函数复现能力。这是 BP 算法得以应用的基础。



图为 BP 网络的结构

反向传播算法主要由两个环节(激励传播、权重更新)反复循环迭代，直到网络的对输入的响应达到预定的目标范围为止。

BP 算法的学习过程由正向传播过程和反向传播过程组成。在正向传播过程中，输入信息通过输入层经隐含层，逐层处理并传向输出层。如果在输出层得不到期望的输出值，则取输出与期望的误差的平方和作为目标函数，转入反向传播，逐层求出目标函数对各神经元权值的偏导数，构成目标函数对权值向量的梯量，作为修改权值的依据，网络的学习在权值修改过程中完成。误差达到所期望值时，网络学习结束。

激励传播

每次迭代中的传播环节包含两步：

1、(前向传播阶段)将训练输入送入网络以获得激励响应；

2、(反向传播阶段)将激励响应同训练输入对应的目标输出求差，从而获得隐层和输出层的响应误差。

权重更新

对于每个突触上的权重，按照以下步骤进行更新：

1、将输入激励和响应误差相乘，从而获得权重的梯度；

2、将这个梯度乘上一个比例并取反后加到权重上。

3、这个比例将会影响到训练过程的速度和效果，因此称为“训练因子”。梯度的方向指明了误差扩大的方向，因此在更新权重的时候需要对其取反，从而减小权重引起的误差。