

人工智能 算法 计算机 神经网络

如何简单形象又有趣地讲解神经网络是什么？

有网友在 Quora 上提问：对于那些非计算机科学行业的人，你会如何向他们解释机器学习和数据挖掘？ 斯坦福大学的印度学生、机器学习爱好者 Parart...显示全部

5 条评论 分享 邀请回答 ...

46 个回答

默认排序



王小龙 数学 话题的优秀回答者

收录于 编辑推荐 · 知乎周刊 · 5253 人赞同了该回答
神经网络很萌的！

0. 分类

神经网络最重要的用途是分类，为了让大家对分类有个直观的认识，咱们先看几个例子：

- 垃圾邮件识别：现在有一封电子邮件，把出现在里面的所有词汇提取出来，送进一个机器里，机器需要判断这封邮件是否是垃圾邮件。
- 疾病判断：病人到医院去做了一大堆肝功、尿检测验，把测验结果送进一个机器里，机器需要判断这个病人是否得病，得的什么病。
- 猫狗分类：有一大堆猫、狗照片，把每一张照片送进一个机器里，机器需要判断这幅照片里的东西是猫还是狗。

这种能自动对输入的东西进行分类的机器，就叫做分类器。

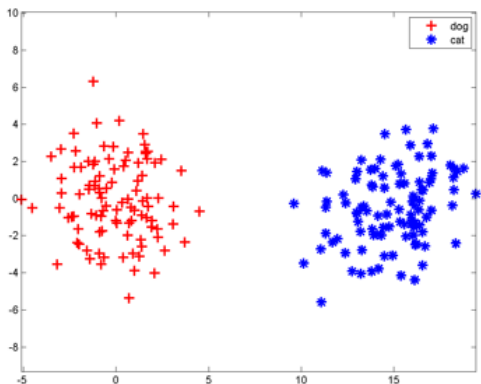
分类器的输入是一个数值向量，叫做特征（向量）。在第一个例子里，分类器的输入是一堆0、1值，表示字典里的每一个词是否在邮件中出现，比如向量(1,1,0,0,0.....)就表示这封邮件里只出现了两个词abandon和abnormal；第二个例子里，分类器的输入是一堆化验指标；第三个例子里，分类器的输入是照片，假如每一张照片都是320*240像素的红绿蓝三通道彩色照片，那么分类器的输入就是一个长度为320*240*3=230400的向量。

分类器的输出也是数值。第一个例子中，输出1表示邮件是垃圾邮件，输出0则说明邮件是正常邮件；第二个例子中，输出0表示健康，输出1表示有甲肝，输出2表示有乙肝，输出3表示有饼干等等；第三个例子中，输出0表示图片中是狗，输出1表示是猫。

分类器的目标就是让正确分类的比例尽可能高。一般我们需要首先收集一些样本，人为标记上正确分类结果，然后用这些标记好的数据训练分类器，训练好的分类器就可以在新来的特征向量上工作了。

1. 神经元

咱们假设分类器的输入是通过某种途径获得的两个值，输出是0和1，比如分别代表猫和狗。现在有一些样本：



大家想想，最简单地把这两组特征向量分开的方法是谁？当然是在两组数据中间画一条竖直线，直线左边是狗

659

36 条评论

分享

收藏

感谢

收起

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验

相关问题

右边的都是猫。

一条直线把平面一分为二，一个平面把三维空间一分为二，一个n-1维超平面把n维空间一分为二，两边分属不同的两类，这种分类器就叫做神经元。

大家都知道平面上的直线方程是 $ax + by + c = 0$ ，等式左边大于零和小于零分别表示点 (x, y) 在直线的一侧还是另一侧，把这个式子推广到n维空间里，直线的高维形式称为超平面，它的方程是：

$$h = a_1x_1 + a_2x_2 + \dots + a_nx_n + a_0 = 0$$

神经元就是当h大于0时输出1，h小于0时输出0这么一个模型，它的实质就是**把特征空间一切两半，认为两瓣分别属两个类**。你恐怕再也想不到比这更简单的分类器了，它是McCulloch和Pitts在1943年想出来了。

这个模型有点像人脑中的神经元：从多个感受器接受电信号 x_1, x_2, \dots, x_n ，进行处理（加权相加再偏移一点，即判断输入是否在某条直线 $h = 0$ 的一侧），发出电信号（在正确的那侧发出1，否则不发信号，可以认为是发出0），这就是它叫神经元的原因。

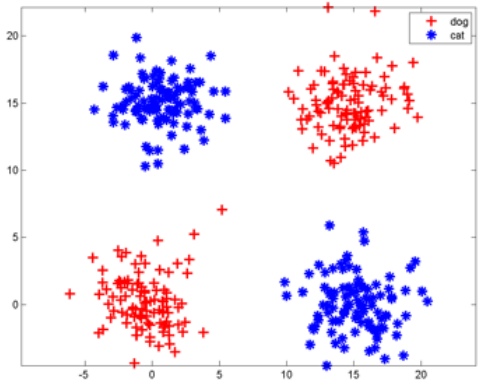
当然，上面那幅图我们是开了上帝视角才知道“一条竖直线能分开两类”，在实际训练神经元时，我们并不知道特征是怎么抱团的。神经元模型的一种学习方法称为Hebb算法：

先随机选一条直线/平面/超平面，然后把样本一个个拿过来，如果这条直线分错了，说明这个点**分错边了**，就稍微把直线移动一点，让它靠近这个样本，争取跨过这个样本，让它跑到直线正确的一侧；如果直线分对了，它就暂时停下不动。因此训练神经元的过程就是这条直线不断在跳舞，最终跳到两个类之间的竖直线位置。

2. 神经网络

MP神经元有几个显著缺点。首先它把直线一侧变为0，另一侧变为1，这东西不可微，不利于数学分析。人们用一个和0-1阶跃函数类似但是更平滑的函数Sigmoid函数来代替它（Sigmoid函数自带一个尺度参数，可以控制神经元对离超平面距离不同的点的响应，这里忽略它），从此神经网络的训练就可以用梯度下降法来构造了，这就是有名的反向传播算法。

神经元的另一个缺点是：它只能切一刀！你给我说说一刀怎么能把下面这两类分开吧。



解决办法是多层神经网络，底层神经元的输出是高层神经元的输入。我们可以在中间横着砍一刀，竖着砍一刀，然后把左上和右下的部分合在一起，与右上的左下部分分开；也可以围着左上角的边沿砍10刀把这一部分先挖出来，然后和右下角合并。

每砍一刀，其实就是使用了一个神经元，把不同砍下的半平面做交、并等运算，就是把这些神经元的输出当作输入，后面再连接一个神经元。这个例子中特征的形状称为异或，这种情况一个神经元搞不定，但是两层神经元就能正确对其进行分类。

只要你能砍足够多刀，把结果拼在一起，什么奇怪形状的边界神经网络都能够表示，所以说神经网络在理论上可以表示很复杂的函数/空间分布。但是真实的神经网络是否能摆动到正确的位置还要看网络初始值设置、样本容量和分布。

神经网络神奇的地方在于它的每一个组件非常简单——把空间切一刀+某种激活函数(0-1阶跃、sigmoid、max-pooling)，但是可以一层一层级联。输入向量连到许多神经元上，这些神经元的输出又连到一堆神经元上，这一过程可以重复很多次。这和人脑中的神经元很相似：每一个神经元都有一些神经元作为其输入，又是另一些神经元的输入，数值向量就像是电信号，在不同神经元之间传导，每一个神经元只有满足了某种条件才会发射信号到下一层神经元。当然，人脑比神经网络模型复杂很多：人工神经网络一般不存在环状结构；人脑神经元的电信号不仅有强弱，还有时间缓急之分，就像莫尔斯电码，在人工神经网络里没有这种复杂的信号模式。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题



神经网络的训练依靠反向传播算法：最开始输入层输入特征向量，网络层层计算获得输出，输出层发现输出和正确的类号不一样，这时它就让最后一层神经元进行参数调整，最后一层神经元不仅自己调整参数，还会勒令连接它的倒数第二层神经元调整，层层往回退着调整。经过调整的网络会在样本上继续测试，如果输出还是老分错，继续来一轮回退调整，直到网络输出满意为止。这很像中国的文艺体制，武媚娘传奇剧组就是网络中的一个神经元，最近刚刚调整了参数。

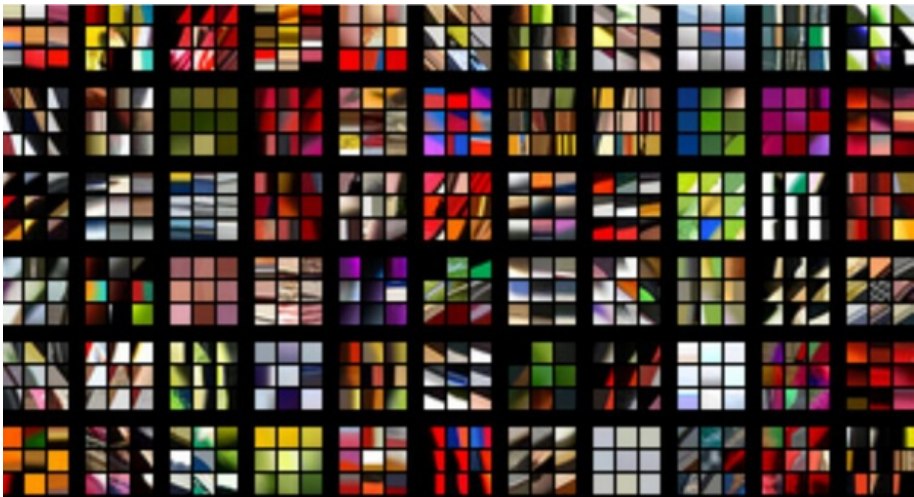
3. 大型神经网络

我们不禁要想了，假如我们的这个网络有10层神经元，第8层第2015个神经元，它有什么含义呢？我们知道它把第七层的一大堆神经元的输出作为输入，第七层的神经元又是以第六层的一大堆神经元做为输入，那么这个特殊第八层的神经元，它会不会代表了某种抽象的概念？

就好比你的大脑里有一大堆负责处理声音、视觉、触觉信号的神经元，它们对于不同的信息会发出不同的信号，那么会不会有这么一个神经元（或者神经元小集团），它收集这些信号，分析其是否符合某个抽象的概念，和其他负责更具体和更抽象概念的神经元进行交互。

2012年多伦多大学的Krizhevsky等人构造了一个超大型卷积神经网络 [1]，有9层，共65万个神经元，6千万个参数。网络的输入是图片，输出是1000个类，比如小虫、美洲豹、救生船等等。这个模型的训练需要海量图片，它的分类准确率也完爆先前所有分类器。纽约大学的Zeiler和Fergusi [2]把这个网络中某些神经元挑出来，把在其上响应特别大的那些输入图像放在一起，看它们有什么共同点。他们发现中间层的神经元响应了某些十分抽象的特征。

第一层神经元主要负责识别颜色和简单纹理



第二层的一些神经元可以识别更加细化的纹理，比如布纹、刻度、叶纹。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

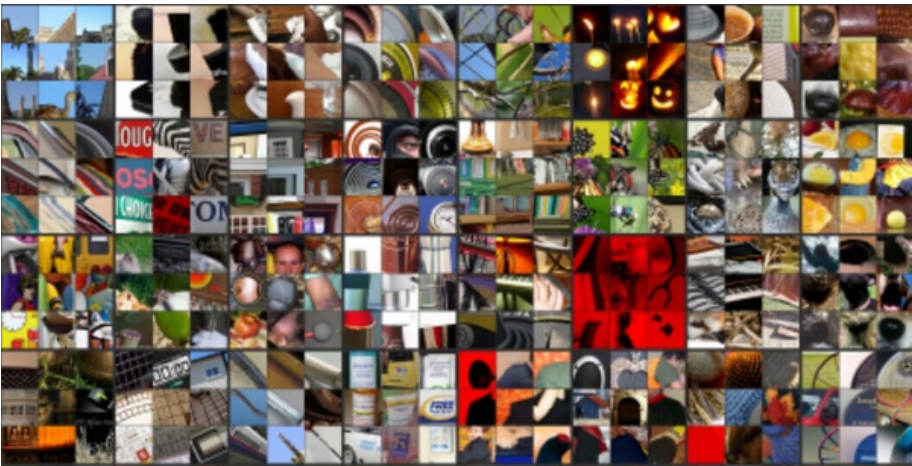


下载知乎客户端
与世界分享知识、经验和见解

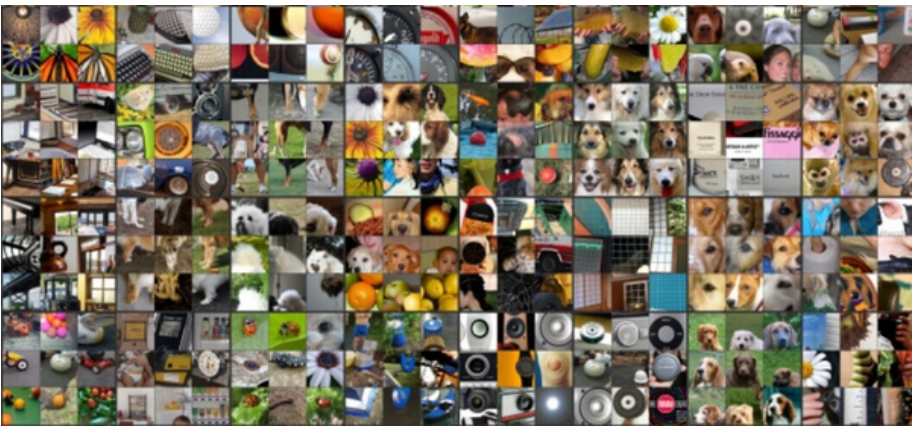
相关问题



第三层的一些神经元负责感受黑夜里的黄色烛光、鸡蛋黄、高光。



第四层的一些神经元负责识别萌狗的脸、七星瓢虫和一堆圆形物体的存在。



第五层的一些神经元可以识别出花、圆形屋顶、键盘、鸟、黑眼圈动物。

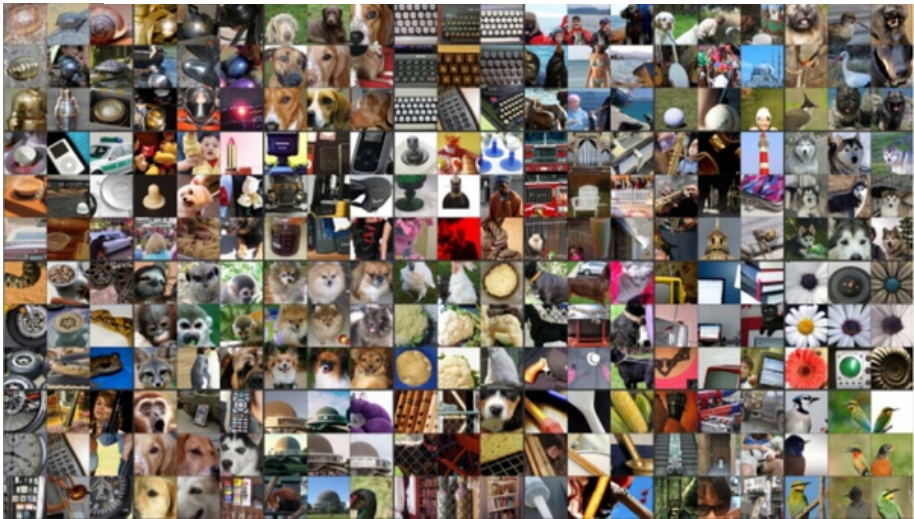
相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 ·
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题



这里面的概念并不是整个网络的输出，是网络中间层神经元的偏好，它们为后面的神经元服务。虽然每一个神经元都傻不拉几的（只会切一刀），但是65万个神经元能学到的东西还真是深邃呢。

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[2] Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional neural networks. *arXiv preprint arXiv:1311.2901*.

编辑于 2017-03-25

▲ 5.3K ▼

● 243 条评论

➤ 分享

★ 收藏

♥ 感谢

收起 ^

 **YJango** ✓

日本会津大学人机界面实验室 博士在读

659 人赞同了该回答

深度学习通俗易懂教程专栏超智能体 - 知乎专栏

阅读列表：

1. 深层神经网络：深度学习为何要“Deep”（上）（由于下篇写的并不通俗，不推荐阅读，用公开课代替）
2. 反向传播算法实例：未编写
3. 深度学习总览：公开课：深层神经网络设计理念
4. 深度学习入门误区：知乎Live（公开课涵盖了Live的内容，若觉得作者辛苦也可参加）
5. **Tensorflow**：TensorFlow整体把握
6. 前馈神经网络（1）：前馈神经网络--代码LV1
7. 前馈神经网络（2）：前馈神经网络--代码LV2
8. 前馈神经网络（3）：前馈神经网络--代码LV3
9. 循环神经网络（1）：循环神经网络--介绍
10. 循环神经网络（2）：循环神经网络--实现LSTM
11. 循环神经网络（3）：循环神经网络--scan实现LSTM
12. 循环神经网络（4）：循环神经网络--双向GRU
13. 卷积神经网络（1）：卷积神经网络--介绍

以下是正文

如何简单形象又有趣地讲解神经网络是什么？

一、基本变换：层

神经网络是由一层一层构建的，那么每层究竟在做什么？

- 数学式子： $\vec{y} = a(W \cdot \vec{x} + b)$ ，其中 \vec{x} 是输入向量， \vec{y} 是输出向量， \vec{b} 是偏移向量， W 是权重矩阵， $a()$ 是激活函数。每一层仅仅是把输入 \vec{x} 经过如此简单的操作得到 \vec{y} 。
- 数学理解：通过如下5种对输入空间（输入向量的集合）的操作，完成 **输入空间** \rightarrow **输出空间** 的变

▲ 659 ▼

● 36 条评论

➤ 分享

★ 收藏

♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区

违法和不良信息举报：010-82716601

联系我们 © 2017 知乎



大脑判定一个人为陌生人的机制？
11 个回答

机器学习有很多关于核函数的说法，核函数的定义和作用是什么？
47 个回答

在自然科学领域，复杂的模型(如神经网络)在逐渐淘汰掉简单的模型吗？
答

是否可以用计算机实现诊断疾病？
相关问题

香农的信息论 关注问题 哪里? 82

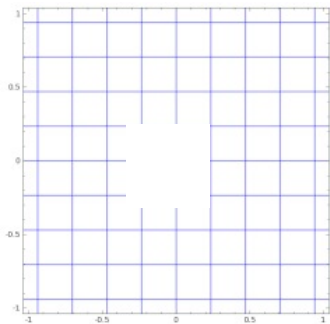
 深度学习入门误区

 我们一起开始机器学习吧

注：用“空间”二字的原因是被分类的并不是单个事物，而是一类事物。空间是指这类事物所有个体的集合。

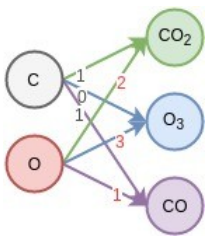
- 1. 升维/降维
- 2. 放大/缩小
- 3. 旋转
- 4. 平移
- 5. “弯曲”

这5种操作中，1,2,3的操作由 $W \cdot \vec{x}$ 完成，4的操作是由 $+\vec{b}$ 完成，5的操作则是由 $\alpha()$ 来实现。(此处有动态图5种空间操作，帮助理解)



每层神经网络的数学理解：用线性变换跟随非线性变化，将输入空间投向另一个空间。

- 物理解释：对 $W \cdot \vec{x}$ 的理解就是通过组合形成新物质。 $\alpha()$ 又符合了我們所处的世界都是非线性的特点。
 - 情景： \vec{x} 是二维向量，维度是碳原子和氧原子的数量 $[C; O]$ ，数值且定为 $[1; 1]$ ，若确定 \vec{y} 是三维向量，就会形成如下网络的形状 (神经网络的每个节点表示一个维度)。通过改变权重的值，可以获得若干个不同物质。右侧的节点数决定了想要获得多少种不同的新物质。（矩阵的行数）



- 1.如果权重W的数值如（1），那么网络的输出□□ 就会是三个新物质，[二氧化碳，臭氧，一氧化碳]。
$$\begin{bmatrix} CO_2 \\ O_3 \\ CO \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} C \\ O \end{bmatrix} \quad (1)$$
- 2.也可以减少右侧的一个节点，并改变权重W至（2），那么输出 \vec{y} 就会是两个新物质， $[O_{0.3}; CO_{1.5}]$ 。
$$\begin{bmatrix} O_{0.3} \\ CO_{1.5} \end{bmatrix} = \begin{bmatrix} 0 & 0.3 \\ 1 & 1.5 \end{bmatrix} \cdot \begin{bmatrix} C \\ O \end{bmatrix} \quad (2)$$
- 3.如果希望通过层网络能够从[C, O]空间转变到 $[CO_2; O_3; CO]$ 空间的话，那么网络的学习过程就是将W的数值变成尽可能接近(1)的过程。如果再加一层，就是通过组合 $[CO_2; O_3; CO]$ 这三种基础物质，形成若干更高层的物质。 4.重要的是这种组合思想，组合成的东西在神经网络中并不需要物理意义。

每层神经网络的物理解释：通过现有的不同物质的组合形成新物质。

二、理解视角：

现在我們知道了每一层的行为，但这种行为又是如何完成识别任务的呢？

数学视角：“线性可分”

- 一维情景：以分类为例，当要分类正数、负数、零，三类的时候，一维空间的直线可以找到两个超平面（比当前空间低一维的子空间。当前空间是直线的话，超平面就是点）分割这三类。但面对
- ▲ 659 ▼

● 36 条评论

➤ 分享

★ 收藏

♥ 感谢

如何利用 AI 打造对话式



大数据是否能改造你



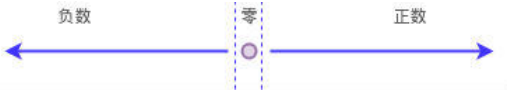
推荐 自动聊天系统的
案与技术挑战

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

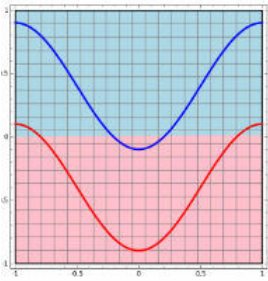


下载知乎客户端
与世界分享知识、经验

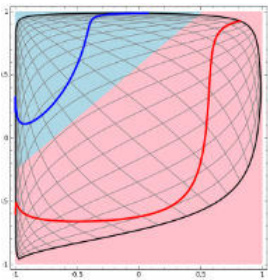
相关问题



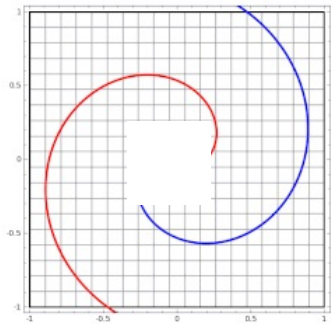
- **二维情景：**平面的四个象限也是线性可分。但下图的红蓝两条线就无法找到一超平面去分割。



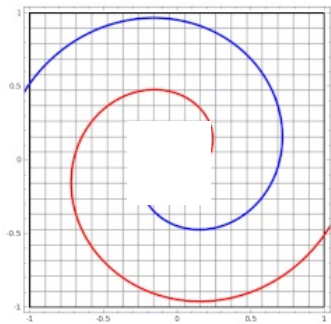
神经网络的解决方法依旧是转换到另外一个空间下，用的是所说的**5种空间变换操作**。比如下图就是经过放大、平移、旋转、扭曲原二维空间后，在三维空间下就可以成功找到一个超平面分割红蓝两线 (同SVM的思路一样)。



上面是一层神经网络可以做到的，如果把 \vec{y} 当做新的输入再次用这5种操作进行第二遍空间变换的话，网络也就变为了二层。最终输出是 $\vec{y} = a_2(W_2 \cdot (a_1(W_1 \cdot \vec{x} + b_1)) + b_2)$ 。设想网络拥有很多层时，对原始输入空间的“扭曲力”会大幅增加，如下图，最终我们可以轻松找到一个超平面分割空间。



当然也有如下图失败的时候，关键在于“如何扭曲空间”。所谓监督学习就是给予神经网络网络大量的训练例子，让网络从训练例子中学会如何变换空间。每一层的权重W就控制着如何变换空间，我们最终需要的也就是训练好的神经网络的所有层的权重矩阵。



这里有非常棒的可视化空间变换demo，一定要打开尝试并感受这种扭曲过程。更多内容请看 [Neural Networks, Manifolds, and Topology](#)

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

线性可分视角：神经网络的学习就是学习如何利用矩阵的线性变换加激活函数的非线性变换，将原始输入空间投向线性可分/稀疏的空间去分类/回归。增加节点数：增加维度，即增加线性转换能力。增加层数：增加激活函数的次数，即增加非线性转换次数。

物理视角：“物质组成”

- 类比：回想上文由碳氧原子通过不同组合形成若干分子的例子。从分子层面继续迭代这种组合思想，可以形成DNA，细胞，组织，器官，最终可以形成一个完整的人。继续迭代还会有家庭，公司，国家等。这种现象在身边随处可见。并且原子的内部结构与太阳系又惊人的相似。不同层级之间都是以类似的几种规则再不断形成新物质。你也可能听过分形学这三个字。可通过观看从1米到150亿光年来感受自然界这种层级现象的普遍性。

- 人脸识别情景：我们可以模拟这种思想并应用在画面识别上。由像素组成菱角再组成五官最后到不同的人脸。每一层代表不同的不同的物质层面 (如分子层)。而每层的W存储着如何组合上一层的物质从而形成新物质。
如果我们完全掌握一架飞机是如何从分子开始一层一层形成的，拿到一堆分子后，我们就可以判断他们是否可以以此形成方式，形成一架飞机。
附：[Tensorflow playground](#) 展示了数据是如何“流动”的。

物质组成视角：神经网络的学习过程就是学习物质组成方式的过程。增加节点数：增加同一层物质的种类，比如118个元素的原子层就有118个节点。增加层数：增加更多层级，比如分子层，原子层，器官层，并通过判断更抽象的概念来识别物体。

三、神经网络的训练

知道了神经网络的学习过程就是学习控制着空间变换方式（物质组成方式）的权重矩阵后，接下来的问题就是如何学习每一层的权重矩阵W。

如何训练：

既然我们希望网络的输出尽可能的接近真正想要预测的值。那么就可以通过比较当前网络的预测值和我们真正想要的目标值，再根据两者的差异情况来更新每一层的权重矩阵（比如，如果网络的预测值高了，就调整权重让它预测低一些，不断调整，直到能够预测出目标值）。因此就需要先定义“如何比较预测值和目标的差异”，这便引出生成函数或目标函数 (loss function or objective function)

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

性越大。那神经网络的训练就变成了尽可能的缩小loss的过程。
所用的方法是**梯度下降（Gradient descent）**：通过使loss值向当前点对应梯度的反方向不断移动，来降低loss。一次移动多少是由**学习速率（learning rate）**来控制的。

梯度下降的问题：

然而使用梯度下降训练神经网络拥有两个主要难题。

1、局部极小值

梯度下降寻找的是loss function的局部极小值，而我们想要全局最小值。如下图所示，我们希望loss值可以降低到右侧深蓝色的最低点，但loss有可能“卡”在左侧的局部极小值中。

试图解决“卡在局部极小值”问题的方法分两大类：

- **调节步伐**：调节学习速率，使每一次的更新“步伐”不同。常用方法有：
- 随机梯度下降（Stochastic Gradient Descent (SGD)）：每次只更新一个样本所计算的梯度
- 小批量梯度下降（Mini-batch gradient descent）：每次更新若干样本所计算的梯度的平均值
- 动量（Momentum）：不仅仅考虑当前样本所计算的梯度；Nesterov动量（Nesterov Momentum）：Momentum的改进
- Adagrad、RMSProp、Adadelta、Adam：这些方法都是训练过程中依照规则降低学习速率，部分也综合动量
- **优化起点**：合理初始化权重（weights initialization）、预训练网络（pre-train），使网络获得一个较好的“起始点”，如最右侧的起始点就比最左侧的起始点要好。常用方法有：高斯分布初始权重（Gaussian distribution）、均匀分布初始权重（Uniform distribution）、Glorot 初始权重、He初始权、稀疏矩阵初始权重（sparse matrix）

2、梯度的计算

机器学习所处理的数据都是高维数据，该如何快速计算梯度、而不是以年来计算。
其次如何更新隐藏层的权重？

解决方法是：计算图：反向传播算法

这里的解释留给非常棒的Computational Graphs: Backpropagation

需要知道的是，反向传播算法是求梯度的一种方法。如同快速傅里叶变换（FFT）的贡献。
而计算图的概念又使梯度的计算更加合理方便。

基本流程图：

下面就结合图简单浏览一下训练和识别过程，并描述各个部分的作用。要结合图解阅读以下内容。
但手机显示的图过小，最好用电脑打开。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

- **收集训练集 (train data)**：也就是同时有input以及对应label的数据。每个数据叫做训练样本 (sample)。label也叫target，也是机器学习中最贵的部分。上图表示的是我的数据库。假设input本别是x的维度是39，label的维度是48。
- **设计网络结构 (architecture)**：确定层数、每一隐藏层的节点数和激活函数，以及输出层的激活函数和损失函数。上图用的是两层隐藏层（最后一层是输出层）。隐藏层所用激活函数a()是ReLu，输出层的激活函数是线性linear（也可看成是没有激活函数）。隐藏层都是1000节点。损失函数L()是用于比较距离MSE: $\text{mean}((\text{output} - \text{target})^2)$ 。MSE越小表示预测效果越好。训练过程就是不断减小MSE的过程。到此所有数据的维度都已确定：
 - 训练数据: $\text{input} \in R^{39}; \text{label} \in R^{48}$
 - 权重矩阵: $W_{h1} \in R^{1000 \times 39}; W_{h2} \in R^{1000 \times 1000}; W_o \in R^{48 \times 1000}$
 - 偏移向量: $b_{h1} \in R^{1000}; b_{h2} \in R^{1000}; b_o \in R^{48}$
 - 网络输出: $\text{output} \in R^{48}$
- **数据预处理 (preprocessing)**：将所有样本的input和label处理成能够使用神经网络的数据，label的值域符合激活函数的值域。并简单优化数据以便让训练易于收敛。比如中心化 (mean subtraction)、归一化 (normlization)、主成分分析 (PCA)、白化 (whitening)。假设上图的input和output全都经过了中心化和归一化。
- **权重初始化 (weights initialization)**： W_{h1}, W_{h2}, W_o 在训练前不能为空，要初始化才能够计算loss从而来降低。 W_{h1}, W_{h2}, W_o 初始化决定了loss在loss function中从哪个点开始作为起点训练网络。上图用均匀分布初始权重 (Uniform distribution)。
- **训练网络 (training)**：训练过程就是用训练数据的input经过网络计算出output，再和label计算出loss，再计算出gradients来更新weights的过程。
 - 正向传递：，算当前网络的预测值
$$\text{output} = \text{linear}(W_o \cdot \text{Relu}(W_{h2} \cdot \text{Relu}(W_{h1} \cdot \text{input} + b_{h1}) + b_{h2}) + b_o)$$
 - 计算loss: $\text{loss} = \text{mean}((\text{output} - \text{target})^2)$
 - 计算梯度：从loss开始反向传播计算每个参数 (parameters) 对应的梯度 (gradients)。这里用Stochastic Gradient Descent (SGD) 来计算梯度，即每次更新所计算的梯度都是从一个样本计算出来的。传统的方法Gradient Descent是正向传递所有样本来计算梯度。SGD的方法来计算梯度的话，loss function的形状如下图所示会有变化，这样在更新中就有可能“跳出”局部最小值。

- **更新权重**：这里用最简单的方法来更新，即所有参数都
$$W = W - \text{learningrate} * \text{gradient}$$
- **预测新值**：训练过所有样本后，打乱样本顺序再次训练若干次。训练完毕后，当再来新的数据input，就可以利用训练的网络来预测了。这时的output就是效果很好的预测值了。下图是一张实际值和预测值的三组对比图。输出数据是48维，这里只取1个维度来画图。蓝色的是实际值，绿色的是实际值。最上方的是训练数据的对比图，而下方的两行是神经网络模型从未见过的数据预测对比图。（不过这里用的是RNN，主要是为了让大家感受一下效果）

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

注：此部分内容**不是**这篇文章的**重点**，但为了理解深层神经网络，需要明白最基本的训练过程。若能理解训练过程是通过梯度下降尽可能缩小loss的过程即可。若有理解障碍，可以用python实践一下[从零开始训练一个神经网络](#)，体会整个训练过程。若有时间则可以再体会一下计算图自动求梯度的方便利用[TensorFlow](#)。

结合Tensorflow playground 理解5种空间操作和物质组成视角

打开网页后，总体来说，蓝色代表正值，黄色代表负值。拿分类任务来分析。

- 数据：在二维平面内，若干点被标记成了两种颜色。黄色，蓝色，表示想要区分的两类。你可以把平面内的任意点标记成任意颜色。网页给你提供了4种规律。神经网络会根据你给的数据训练，再分类相同规律的点。

- 输入：在二维平面内，你想给网络多少关于“点”的信息。从颜色就可以看出来， x_1 左边是负，右边是正， x_1 表示此点的横坐标值。同理， x_2 表示此点的纵坐标值。 x_1^2 是关于横坐标值的“抛物线”信息。你也可以给更多关于这个点的信息。给的越多，越容易被分开。

- 连接线：表示权重，蓝色表示用神经元的原始输出，黄色表示用负输出。深浅表示权重的绝对值大小。鼠标放在线上可以看到具体值。也可以更改。在（1）中，当把 x_2 输出的一个权重改为-1时， x_2 的形状直接倒置了。不过还需要考虑激活函数。（1）中用的是linear。在（2）中，当换成sigmoid时，你会发现没有黄色区域了。因为sigmoid的值域是(0,1)



下载知乎客户端
与世界分享知识、经验和见解

相关问题

(1)

(2)

- 输出：黄色背景颜色都被归为黄点类，蓝色背景颜色都被归为蓝点类。深浅表示可能性的强弱。

上图中所有在黄色背景颜色的点都会被分类为“黄点”，同理，蓝色区域被分成蓝点。在上面的分类分布图中你可以看到每一层通过上一层信息的组合所形成的。权重（那些连接线）控制了“如何组合”。神经网络的学习也就是从数据中学习那些权重。Tensorflow playground所表现出来的现象就是“在我文章里所写的“物质组成思想”，这也是为什么我把Tensorflow playground 放在了那一部分。

不过你要是把Tensorflow的个名字拆开来看的话，是tensor（张量）的flow（流动）。Tensorflow playground的作者想要阐述的侧重点是“张量如何流动”的。

5种空间变换的理解：Tensorflow playground下没有体现5种空间变换的理解。需要打开这个网站尝试：[ConvNetJS demo: Classify toy 2D data](#)

左侧是原始输入空间下的分类图，右侧是转换后的高维空间下的扭曲图。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

最终的扭曲效果是所有绿点都被扭曲到了一侧，而所有红点都被扭曲到了另一侧。这样就可以线性分割（用超平面（这里是一个平面）在中间分开两类）

四、“深层”的思考：真的只有这些原因吗？

文章的最后稍微提一下深层神经网络。深层神经网络就是拥有更多层数的神经网络。

按照上文在理解视角中所述的观点，可以想出下面两条理由关于为什么更深的网络会更加容易识别，增加容纳变异体（variation）（红苹果、绿苹果）的能力、鲁棒性（robust）。

数学视角：变异体（variation）很多的分类的任务需要高度非线性的分割曲线。不断的利用那5种空间变换操作将原始输入空间像“捏橡皮泥一样”在高维空间下捏成更为线性可分/稀疏的形状。 **物理视角：**通过对“**抽象概念**”的判断来识别物体，而非细节。比如对“飞机”的判断，即便人类自己也无法用语言或者若干条规则来解释自己如何判断一个飞机。因为人脑中真正判断的不是是否“有机翼”、“能飞行”等细节现象，而是一个抽象概念。层数越深，这种概念就越抽象，所能涵盖的**变异体**就越多，就可以容纳战斗机，客机等很多种不同种类的飞机。

编辑于 2017-07-14



PENG Bo
稟临科技 withablink.com 联合创始人

1136 人赞同了该回答

最近 Google Tensorflow 做了一个非常直观的神经网络 playground。不夸张地说，现在每个人都可以在自己的浏览器里面直接“玩深度神经网络”了。什么都不用说，自己玩玩就明白。

地址是：[A Neural Network Playground](#) （可能要翻墙）

如果不明白里面的 ReLU，L1 等等是什么，没关系，在搜索引擎查一下都可以查到答案。代码在 Github 上，有兴趣的朋友可以去给它加 Dropout，Convolution，Pooling 之类。

【UPDATE：对 AlphaGo 和围棋人工智能有兴趣的朋友，我最近在写一个系列，介绍如何自制你的"AlphaGo"，请点击：[知乎专栏](#)】

编辑于 2017-01-10

▲ 1.1K ▼

💬 45 条评论

➦ 分享

★ 收藏

♥ 感谢



舒小曼
心理学硕士/数据分析/译者

159 人赞同

▲ 659 ▼

💬 36 条评论

➦ 分享

★ 收藏

♥ 感谢

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

先来认识下真正的神经元。

图1:典型神经元的结构（来自维基百科“树突”词条）

神经元大致可以分为树突、突触、细胞体和轴突。树突为神经元的输入通道，其功能是将其他神经元的动作电位传递至细胞体。其他神经元的动作电位借由位于树突分支上的多个突触传递至树突上。神经细胞可以视为有两种状态的机器，激活时为“是”，不激活时为“否”。神经细胞的状态取决于从其他神经细胞接收到的信号量，以及突触的性质（抑制或加强）。当信号量超过某个阈值时，细胞体就会被激活，产生电脉冲。电脉冲沿着轴突并通过突触传递到其它神经元。（内容来自维基百科“感知机”词条，稍作整理）

我们的视觉系统非常复杂。大致来讲，当光线投射到视网膜后，视觉膜上的光感受细胞的突触直接与双极细胞相连，双极细胞突触则与最外层的节细胞相连，节细胞将动作电位传递到大脑。我们的视网膜有大约1亿3千万个光感受细胞接收光信号，然后通过大约120万个节细胞轴突将信息从视网膜传递到大脑。（内容来自维基百科“视觉系统”词条）

视觉系统使我们有了视知觉能力，将外部世界的二维投射重构为三维世界。

人类的视觉系统可以毫不费力的识别出下面图片中的数字（当然还有文字）。但如果让计算机来识别，就会明显感觉到视觉识别的困难。

图2:街道地址（来源于维基百科“Handwriting recognition”词条）

我们不能简单地把“数字2像一只水面上的天鹅，数字7像一把锄头”这样的规则教给计算机。因此，我们需要把手写识别的问题表述成计算机可以理解的方式。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

图3:水面上的天鹅像极了数字2

假设我们有5x5像素的图片来显示每个数字，例如数字2表示如下：

```
#####
.....@
#####
@.....
#####
```

由于我们的神经网络是以数字组成向量作为输入的，所以我们将每个图片转换成长度为25的向量，其元素的值是1（“这个像素位于该图像中”）或0（“这个像素不在该图像中”）

例如，数字2可以表示为：

```
two_digit=[1,1,1,1,1
            0,0,0,0,1
            1,1,1,1,1
            1,0,0,0,0
            1,1,1,1,1]
```

我们希望神经网络最终的输出是一个具体的数字，因此我们需要10个不同的输出结果。例如，对于数字2，正确的输出结果将是：

```
[0,0,1,0,0,0,0,0,0,0]
```

图4:使用python列表解析式建立一个代表0-9不同数字的列表

其中targets[2]就是输出结果数字2。

那么，神经网络是如何把信息的输入转换成我们想要的输出的呢？

感知器（Perceptron）

感知器是一种二元分类器，它是神经网络的基石。感知器是对神经细胞的模拟，如权量（突触）、偏置（阈值）及激活函数（细胞体）。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

输入以向量的形式表示 $x = (x_0, x_1, x_2)$ ，你可以把它们理解为不同的特征维度，其中的 x_0 是偏置单元（bias unit），相当于线性回归中的常数项。在经过“神经元”（激活函数）的计算后，感知器会输出一个大于0或小于0的数。箭头上的数字代表每个特征的权量（weights），相当于线性回归模型的参数，是收集信息的神经突触。

例如，上图列出了每个特征的权量，于是我们有这样的模型：

$$h\theta(x) = -3 + 2 \cdot x_1 + 2 \cdot x_2$$

当 x_1 和 x_2 同时为1时，输出为1，而不同时都将得到负数。这实际上就是逻辑与或称为与门。

我们的感知器将根据输出大于或小于0来对输入进行分类：

这是一个单层的感知器。通过修改权量，我们也可以建立或门（weights=[2, 2], bias=-1）、非门（weights=[-2], bias=1，只有一个输入端）。但是，无论如何，你都无法通过一个单层的感知器来建立“异或门”（XOR），即两个输入相同时输出1，否则输出0。这种情况下，我们就需要更复杂的神经网络了，即多层神经网络，也被称为前馈神经网络（feedforward neural network）。

前馈神经网络

前馈神经网络是多个感知器的组合，这些感知器以不同的方式产生连接，并由不同的激活函数控制激活。

图6:前馈神经网络示意图

我们来认识下前馈神经网络：

- 它包括输入层（input layer）、输出层（output layer）和一个或多个隐藏层（hidden layers）。上图的神经网络由3个单元的输入层，4个单元的隐藏层和2个单元的输出层组成。单元等于感知器。
- 输入层的单元是隐藏层单元的输入，隐藏层单元的输出是输出层单元的输入。
- 两个感

▲ 659 ▼

● 36 条评论 ➦ 分享 ★ 收藏 ♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

- 第t层的每个感知器与第t-1层的每个感知器相互关联。当然，你也可以设置权量为0，从而在实质上取消连接。
- 在加工输入数据时，你将输入数据赋予输入层的每个单元，而隐藏层的每个单元是输入层每个单元的加权求和。也就是说，输入层的数据会被前向传播到隐藏层的每个单元。同理，隐藏层的输出作为输入会前向传播到输入层，计算得到最后的输出，即神经网络的输出。
- 多个隐藏层的神经网络同理。

超越线性

在“感知器”中我们谈到，单层感知器虽然可以建立与门、或门、非门等，但无法建立更为复杂的异或门（XOR），即两个输入相同时输出1，否则输出0。

为了更为直观地理解这个问题，我们可以看下图：

图7:双输入感知器的决策空间

模型有两个输入（input1和input2），我们可以线性地区分或门的两类情况：即同时为0时在左下角，其它情况在右上角；与门的情况也可以线性地区分，即输出同时为1时在右上角，其它情况在左下角。但异或门呢？这种情况是无法作出线性区分的，也就是说，单层感知器无法实现异或门。

多层感知器吗？

先来分析输入和输出的情况。最左边两列列出了输入的4种情况。异或门的输出是最右边一列的情况，即两个输入相同时输出1，否则为0。我们在输入层和输出层之间加入两个单元的隐藏层，那么，它给输出层的输入应该是什么呢？答案如下图。你可以发现，对于隐藏层的a_1单元（上标2代表这是第2层）来说，它实际上是且门（都为1时才输出1）；对a_2单元来说，它的逻辑是(not x_1) and (not x_2)，即同时为0时才输出1。而从隐藏层到输出层，是逻辑或。前馈神经网络可以实现异或门！

图8:异或门

▲ 659 ▼

36 条评论 分享 收藏 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

于是我们建立如下的神经网络，但是其输出并非我们想要。为什么？

图9:线性激活函数下的前馈神经网络

这是因为上面感知器的激活函数是线性函数。这种情况下，神经网络的输出也只是输入的某种线性函数，只不过是通过网络的形式来进行加权。线性函数的线性组合仍然是线性函数。也就是说，即便是多层的感知器，激活函数为线性时也无法实现输入00和输入11时的输出比输入01和10时大，即非线性。

如果激活函数是线性函数，前馈神经网络将不会比单层感知器强大多少，不管它有多少层。

我们需要logistic函数的帮助。大部分神经网络会使用非线性激活函数，比如logistic函数（也被称为Sigmoid函数，这是就其形状而言的）。此外，由于训练神经网络要使用微积分，而要使用微积分，就得使用光滑函数（在反向传播时我们会详细讲解）。logistic函数也满足这一要求。

在logistic函数的作用下，线性函数的输出值将被转化到0到1之间。从下图右上角的logistic曲线可以看到，输入值越大，输出越接近1，输入为4时输出为0.99；输入值越小，输出越接近0，输入值为-4时，输出为0.01。

下图是单层感知器的激活函数为logistic函数时与门的示例。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

图10:激活函数为logitic函数时的与门单层感知器

综上，我们可以通过前馈神经网络和logistic函数解决异或门的问题：

图11:非线性神经网络实现异或门

反向传播（backpropagation）：训练神经网络

反向传播是使用数据来训练神经网络的算法，它是神经网络的梯度下降算法。

假设我们有一个训练集，其中含有输入向量和相应的目标输出向量。同时，假定我们的网络已经拥有一组权量（相当于我们知道每个神经元的激活函数），那么接下来，我们就需要使用以下算法来调整这些权量。

- 1、利用初始权量，在输入向量上运行前向传播，从而得到所有网络所有神经元的输出。
- 2、这样，每个输出层神经元都会得到一个误差，即输出值与实际值之差。
- 3、计算作为神经元权量的函数的误差的梯度，然后根据误差降低最快的方向调整权量。
- 4、将这些输出误差反向传播给隐藏层以便计算相应误差。
- 5、计算这些误差的梯度，并利用同样的方式调整隐藏层的权量。

不断迭代，直到网络收敛。

要理解上述过程，梯度下降算法是关键。

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

图12:单个参数的损失函数示例

在梯度下降算法中，我们需要损失函数（即上面提到的作为神经元权量的函数的误差）。损失函数衡量了模型的误差，即神经网络的输出值与实际值之间的差异，它的参数是模型的参数，即连接的权量。当损失函数达到全局最低时，得到最优的权量。在训练阶段，权量不断被更新，慢慢接近全局最低值。例如，权量是0.6时，它就需要向0.4变化，这样才能接近全局最低值。

上图描述的是只有一个参数的情况，即模型的误差取决于一个参数。然而，神经网络的误差取决于每一个权量，它的损失函数也会复杂得多。

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log h_{\theta}(x^{(i)})_k + (1 - y_k^{(i)}) \log (1 - h_{\theta}(x^{(i)})_k) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\theta_j^{(l)})^2$$

其中，i代表第几层，k代表第i层的第几个单元。可见，损失函数把整个神经网络的误差都加和了。后面的第二部分是正则化项，暂时不用理睬，不影响后面的理解。

有人可能会问，为什么神经网络的损失函数是上面的形式？这个问题等价于为什么logistic函数采用上面的形式？

在回归模型中，损失函数是平方误差函数（最小二乘法）：

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

损失函数衡量了模型的输出与实际值之间的误差：

$$Cost(h_{\theta}(x), y) = \frac{1}{2} (h_{\theta}(x) - y)^2$$

但对logistic函数而言，平方误差函数是一个非凸函数，利用梯度下降算法，它将无法保证找到损失函数的全局最优解。

图13:非凸函数

对logistic函数，我们需要其他的损失函数：

图14:

图15:

如果 $y=1$ ，而 $h(x)=1$ 的话，则 $cost=0$ ，即图14曲线上（1， 0）的位置；如果 $h(x)=0$ ，则 $cost$ 函数输出值趋向于无穷大，这意味着当我们认为某种情况不会发生，例如用户不会成为付费用户，而事实上是可能的时，这种损失是不可估量的。同理，如果 $y=0$ ，而 $h(x)=0$ 的话，则 $cost=0$ ，即图15曲线上（0， 0）的位置；如果 $h(x)=1$ ，则 $cost$ 趋向于无穷大。

可以把两种情况写在一个表达式中：

$$Cost(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x))$$

也即：

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right]$$

神经网络衡量整个网络的误差，所以在形式上看起来会复杂些，但实质是一样的。

那么，接下来我们要如何更新权重才能使得损失函数接近全局最低值呢？

在神经网络中，我们使用反向传播算法来更新神经元之间连接的权重。反向传播的意思是指，我们从输出层开始计算，计算输出层与实际值的误差，然后返回到前面的隐藏层计算其误差，以此类推，不断迭代，直到网络收敛。

在完成第一步的前向传播后，我们会得到神经网络的输出值，利用损失函数可以计算出输出值与实际值之间的误差。损失函数的偏导数可以告诉我们参数往哪个方向更新可以得到更小的损失函数值。导数的几何意义为曲线的斜率，例如，对于图12简单的损失函数而言，参数在0.5时斜率为正，参数值需要减少才能得到更小的损失函数值；而如果参数值为0.1，斜率为负，参数值需要增加才能得到更小的损失函数值。导数可以告诉我们全局最低的方向！

为了得到最小的 $J(\theta)$ ，我们需要不断更新参数 θ ，即神经网络连接的权重。每次更新的变化值由下面的公式决定：

$$\Delta \theta = -\alpha \frac{\partial}{\partial \theta} J(\theta)$$

▲ 659 ▼

● 36 条评论

➦ 分享

★ 收藏

♥ 感谢

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

其中， $\frac{\partial}{\partial \theta_{ij}} J(\theta)$ 为损失函数的偏导数，前面为负是因为斜率为正时，我们需要减少参数的值才能“下山”，而斜率为负时，我们需要增加参数的值才能“下山”，代表学习速率，代表下山的步子有多大，如果步子太小，要走很远才能走到山底，如果步子太大，容易越过谷底，导致无法收敛； $J(\theta)$ 为神经元输出的误差。

从隐藏层到输出层的两个神经元之间的权量一旦更新，就可以更新隐藏层神经元的输出，从而得到隐藏层神经元新的误差值（更新后的输出值和之前前向传播时得到的输出值），进而根据上述公式更新隐藏层和上一层两个神经元之间的权量。依此类推，直到更新完整个神经网络的权量。

利用损失函数继续更新整个网络的权量，直到神经网络输出值的误差达到最小。

所有神经元之间的权量确定后，神经网络就完成训练了。

综上，让机器人看到2绝非易事。

参考资料：

Machine Learning from Coursera by Andrew Ng

Data Science from Scratch by Joel Grus

A Deep Learning Tutorial: From Perceptrons to Deep Networks BY IVAN VASILEV

维基百科

编辑于 2017-05-13

▲ 159 ▼

● 22 条评论

🔗 分享

★ 收藏

❤ 感谢

收起 ^



知乎用户

318 人赞同了该回答

首先把题主给的链接看了一下，大致内容如下：

你需要挑选芒果，你不知道什么样的芒果最好吃，所以你就尝遍了所有的芒果，然后自己总结出个大深黄色的比较好吃，以后再去买的时候，就可以直接挑选这种。

那什么是机器学习呢，就是让你让机器“尝”一遍所有芒果，当然，也假设它知道哪些好吃，让机器去总结一套规律（个大深黄色），这就是机器学习。具体操作，就是你描述给机器每一个芒果的特征（颜色，大小，软硬……），描述给机器其输出（味道如何，是否好吃），剩下的就等机器去学习出一套规则。

等等，那机器是怎么学习到这个规则（个大深黄色的好吃）的？没错，是通过机器学习算法。而题主所问的神经网络，恰好就是一种机器学习算法！近些年来，由于深度学习概念的兴起，神经网络又成为了机器学习领域最热门的研究方法。

我相信当明确神经网络是一种机器学习算法之后，就可以很轻易的给一个外行讲清楚神经网络到底是用来做什么的。可是神经网络为什么可以完成这个功能呢？

神经网络就像一个刚开始学习东西的小孩子，开始认东西，作为一个大人（监督者），第一天，他看见一只京巴狗，你告诉他这是狗；第二天他看见一只波斯猫，他开心地说，这是狗，纠正他，这是猫；第三天，他看见一只蝴蝶犬，他又迷惑了，你告诉他这是狗……直到有一天，他可以分清任何一只猫或者狗。

其实神经网络最初得名，就是其在模拟人的大脑，把每一个节点当作一个神经元，这些“神经元”组成的网络就是神经网络。而由于计算机出色的计算能力和细节把握能力，在大数据的基础上，神经网络往往有比人更出色的表现。

当然了，也可以把神经网络当作一个黑箱子，只要告诉它输入，输出，他可以学到输入与输出的函数关系。神经网络的理论基础之一是三层的神经网络可以逼近任意的函数，所以理论上，只要数据量够大，“箱子容量”够大（神经元数量），神经网络就可以学到你要的东西。

第一次回答问题，什么都想说，可能比较乱，有不对的地方，欢迎大家批评指正。

发布于 2014-02-13

▲ 318 ▼

● 36 条评论

🔗 分享

★ 收藏

❤ 感谢



沈晟

心正

▲ 659 ▼

● 36 条评论

🔗 分享

★ 收藏

❤ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验

相关问题

30 人赞同了该回答

还好问题只是怎样尽量简单的解释什么是神经网络，否则还真不敢回答。学术圈比较喜欢用一些名词术语，很容易对理解造成更多困惑。我只说一下个人的粗浅理解，希望对初入门的同学有帮助。可能存在疏漏和误导，只能请大家充分利用自己的判断力。

先放轻松态度，应该更便于理解神经网络。

神经网络+深度学习中的很多理论方法已经存在很多年（大于10），并不是突然之间像魔术那样出现而不可思议。最近开始吸引焦点，主要是因为可用的计算能力（CPU、GPU）的快速发展，让海量矩阵乘法运算更容易被实验、测试和验证。例如，TensorFlow的出现，最大的意义，就是AI for everybody，让「旧时王谢堂前燕，飞入寻常百姓家」。

如果尝试强行一句话解释深度学习神经网络，我会这么总结：

「通过堆积大量且多样计算方法的函数计算式的计算单元(有关名词：node/activation function/ReLU/...)，并尝试为这些计算单元建立大量的相互连接，同时为了有序有效我们通过分层使连接有次序并避免需要尝试的连接数成幂指数上升，之后制定一个策略帮助整个神经网络的计算评价尝试的结果更好更可用或更坏更不可用(有关名词：lose/cost/objective funtion/softmax/cross entropy)，在提供大量数据训练神经网络之后，神经网络通过计算结果的评价策略，决定以什么样的次序连接哪些计算单元的计算结果更值得被使用(有关名词：强化连接)，通过大量循环不同的连接方式找到评价最好的一些的连接计算单元的次序和权值，就此建立模型(计算式组合)可用于解决更多同类问题。」

堆积大量计算单元（神经元）的最大的好处是什么呢。就是减轻设计（或者说supervise）的压力。可以更少考虑应该怎样去针对一个很具体的feature细节（carefully engineered function）去解决，而是通过训练来强化或弱化它们直接的连接，以一种自然选择的方式淘汰和筛选，留下可用的解题模型。

那么上面说的大量是多大呢。一个昨天看的一个论文，是用了65万个。不过虽然训练时需要很大的计算量，但是训练的结果在使用时，就只需要很小的运算量了。就像迷宫有力地图，再走一次时，就不会走那么多弯路了。

其实通过尝试组合大量不同的算法，找到最适合解决问题的算法组合，训练+强化理论，是机器学习一直都在做的。

那么深度学习神经网络的核心突破是什么，我认为主要是两点：

1、分层。这个最大的好处，是将计算单元之间需要尝试的连接（组合）数量，控制可控的范围。

2、一些被发现很适合作为深度学习组件的计算单元函数，例如ReLU（When in doubt, ReLU）

最后推荐一个我觉得很好的入门资源：[github.com/jtoy/awesome...](https://github.com/jtoy/awesome-deep-learning-on-tensorflow)

编辑于 2016-10-19

▲ 30

▼



● 3 条评论

➦ 分享

★ 收藏

♥ 感谢

收起 ^

 周筠  儿童教育 话题的优秀回答者

622 人赞同了该回答

早年读电信系在职硕士时，选修了一门《人工神经网络》课程，老师讲得云山雾罩，我们听得如梦似幻。虽然最后也考了80多分，但同学们相互间吐槽，都觉得这门课把人给学神经了！别人怎么样我不知道，我反正是稀里糊涂的，没弄清楚概念。

今年，审读吴军老师的《数学之美》第二版，读到这两段话时，我笑了，原来我并不孤独——

【有不少专业术语乍一听很唬人，“人工神经网络”就属于这一类，至少我第一次听到这个词就被唬住了。你想啊，在大家的印象当中，人们对人脑的结构都还根本没有搞清楚，这就冒出来一个“人工的”神经网络，似乎是在用计算机来模拟人脑。想到人脑的结构那么复杂，大家的第一反应一定是人工神经网络肯定非常高深。如果我们有幸遇到一个好心同时又善于表达的科学家或教授，他愿意花一两个小时的时间，深入浅出地把人工神经网络的底细告诉你，你便会发现，“哦，原来是这么回事”。如果我们不幸遇到一个爱卖弄的人，他会很郑重地告诉你“我在使用人工神经网络”或者“我研究的课题是人工神经网络”，然后就没有下文了，如此，你除了对他肃然起敬外，不由得可能还会感到自卑。当然还有好心却不善于表达的人试图将这个概念给你讲清楚，但是他用了一些更懂的名词，讲得云

所获，于是

▲ 659

▼

● 36 条评论

➦ 分享

★ 收藏

♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区

违法和不良信息举报：010-82716601

联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

大家可别以为我是在说笑话，这些都是我的亲身经历。首先，我没有遇到过一两小时给我讲懂的好心人，其次我遇到了一批在我前面卖弄的人，作为年轻人，总是希望把自己不明白的东西搞懂，于是我决定去旁听一门课。不过，我听了大约两三次便不再去了，因为除了浪费时间，似乎我并没得到什么收获。好在我自己做研究暂时用不到它，也就不再关心了。后来在美国读博士期间，我喜欢在睡觉前躺着看书，没事儿就捧着几本关于人工神经网络的教科书在床上看，居然也看懂了。然后再用它做了两三个项目，算是学会了。到这时回过头来再看“人工神经网络”，其实并不复杂，入门也不难，只是我走了弯路。——吴军】

君子报仇，三十年不晚。大学毕业三十年后，吴老师在数学之美（第二版）里，单独用一章讲**Google大脑和人工神经网络**，一抒当年的郁闷，也顺便帮我解了气:)

我把这一章的内容贴一部分在这里，希望对大家理解神经网络的概念有帮助：

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 ·
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

发布于 2015-01-05

▲ 622 ▼ 62 条评论 分享 收藏 感谢 收起 ^



许铁-巡洋舰科技
微信公众号请关注chaoscruiser ,铁哥个人微信号ironcruiser

17 人赞同了该回答

作者：许铁-巡洋舰科技

链接：[知乎专栏](#)

来源：知乎

著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。

人工神经网络的第一个里程碑是感知机perceptron，这个名字其实有点误导，因为它根本上是做决策的。一个感知机其实是对神经元最基本概念的模拟，都未必有多少网络概念，他就是一个自动做决策的机器。

比如说你要决定今天出不出去看电影，你要考虑3个因素，一个是女朋友在不在，一个是电影好不好看，另一个是今天有没有工作，这三个因素每个人的权重都不同，有的人看重女朋友，有的人看重工作，所以权重就不等，最后每个人根据自己的权重做出0或1，去或不去，to be or not to be的决策。那么你怎么做呢？你把三个要素按照它们需要的权重加和在一起，在把这个分数送到一个叫sigmoid的门面前得到去或不去的决定，工作原理如上图。

比单层感知机更复杂的多层感知机-或者我们常说的深度网络，是进行数据处理和模式识别的利器。深度神经网络之所以能够处理这些数据类型，主要是因为这些数据本身具有的复杂结构很适合被NN识别，而人类不需要预先设计识别这些结构的函数而是任由网络学习，D-CNN 深度卷积网络能够同时看到一个图像从细节到抽象的结构，所以能够抓住一些我们人类都说不出细节。

DCNN 深度卷积网络，信号在多级网络里一级级传递，从而使得从微观到宏观的特征都得到分辨。每一层神经元之间没有相互连接。



下载知乎客户端
与世界分享知识、经验和见解

相关问题

▲ 659 ▼ 36 条评论 分享 收藏 感谢 收起 ^

而RNN- 反馈式神经网络（每一层神经元之间有相互连接）则适合处理sequence序列类的数据，发现序列内部的时间结构。

RNN 之所以具有这个能力，就是因为网络内部通过自反馈，具有之前各个时间点的输入信息，因此它可以从无限久远的历史里推测系统的未来，RNN与之前的卷积网络相比最大的特点是它包含了动力学特性，如果说卷积网络是任意函数逼近器，那么RNN就是任意程序逼近器。犹如包含某种工作记忆。用一个比喻来说，就是RNN犹如一个宽阔的池塘宁静的水面，当你投入一个石子，激起的涟漪会在水池里不停反射传播，这是对石头进入那一时刻信息的保存，如果之后在落入一个石头，那么它再度激起的涟漪会和之前的水波叠加作用，形成更复杂的相互作用和纹样。

RNN示意图，同层神经元之间有相互连接，从而使得历史信息在网络里向回声一般交替传递

RNN 具有相互连接的反馈式神经网络也是最接近生物组织神经网络的人工神经网络，具有更多的未来潜力，只是它的训练比feed forward network更复杂。

人工神经网络的训练就如同生物神经网络的学习，都是一个不停试错并减少错误的原理，不过人工神经网络的方法更加简单化，比如gradient descent，就是说在参数空间里寻找使得错误减少最快的方法改进。

人工神经网络对于生物神经网络的奇妙之处在于，它反过来启迪了生物神经网络是在干什么的，就好像费曼的那句话“你要想真正理解一个东西，就造出一台”。

发布于 2016-11-29

▲ 17 ▼

1 条评论

分享

★ 收藏

♥ 感谢

收起 ^



vczh

专业造轮子，拉黑抢前排。gaclib.net

720 人赞同了该回答

- 1+1是多少？10！傻逼，是2
- 1+2是多少？37！傻逼，是3
- 1+3是多少？-5！傻逼，是4
-
- 1+1是多少？5！傻逼，是2
- 1+2是多少？14！傻逼，是3
- 1+3是多少？25！傻逼，是4
-
- 1+1是多少？2！傻逼，是2
- 1+2是多少？5！傻逼，是3
- 1+3是多少？8！傻逼，是4
-
- 1+1是多少？2！傻逼，是2
- 1+2是多少？3！傻逼，是3
- 1+3是多少？4！傻逼，是4
-
- 1+1是多少？2！傻逼，是2
- 1+2是多少？3！傻逼，是3
- 1+3是多少？4！傻逼，是4

▲ 659 ▼

36 条评论

分享

★ 收藏

♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

.....
=====3年后=====

100+200是多少? 300!

发布于 2016-06-21

▲ 720

▼

● 102 条评论

✈ 分享

★ 收藏

♥ 感谢

鄙人 lee philip

李景 神经网络/自然语言处理

57 人赞同了该回答

补充一下@贾伟的答案

训练过程的那段，写得比较像强化学习，神经网络用到的方法大部分是有监督学习。

我觉得故事应该是这样。。。

你想让神经网络（简称神经）辨认猫和狗（算法类型为分类）。于是给它看一百万只狗，和一百万只猫，每只狗的脑门上都贴着“我是狗”的字条，猫则贴“我是猫”（使用标注数据）。神经同时看五十万只禽兽，分四批看，并作出判断（实现时用并行处理，使用mini-batch）。每个看到的猫或狗，如果判断错了，就修正自己的判断标准（优化）。每看完这二百万只禽兽算一轮，每学十轮就看一些之前完全没见过的猫猫狗狗做出判断（推广性测试）。当神经刻苦学习猫狗知识的时候，你正在睡觉，一觉睡醒，发现神经一小时前判断猫和狗比较准（次最优解），现在学太多反而判断不准了（过拟合），于是你让神经使用一小时前的标准（防止过拟合技巧之early-stopping）。这样训练过程就结束了。

编辑于 2014-02-15

▲ 57


▼

● 9 条评论

✈ 分享

★ 收藏

♥ 感谢

 胡昌振

学企管不做企管

30 人赞同了该回答

人工神经网络算法的作用机理还是比较难理解，现在以一个例子来说明其原理。这个例子是关于人的识别技术的，在门禁系统，逃犯识别，各种验证码破译，银行预留印鉴签名比对,机器人设计等领域都有比较好的应用前景，当然也可以用来做客户数据的挖掘工作，比如建立一个能筛选满足某种要求的客户群的模型。

机器识别人和我们人类识别人的机理大体相似，看到一个人也就是识别对象以后，我们首先提取其关键的外部特征比如身高，体形，面部特征，声音等等。根据这些信息大脑迅速在内部寻找相关的记忆区间，有这个人的信息的话，这个人就是熟人，否则就是陌生人。

人工神经网络就是这种机理。假设上图中X(1)代表我们为电脑输入的人的面部特征，X(2)代表人的身高特征X(3)代表人的体形特征X(4)代表人的声音特征W(1)W(2)W(3)W(4)分别代表四种特征的链接权重，这个权重非常重要，也是人工神经网络起作用的核心变量。

现在我们随便找一个人阿猫站在电脑面前，电脑根据预设变量提取这个人的信息，阿猫面部怎么样，身高多少，体形胖瘦，声音有什么特征，链接权重初始值是随机的，假设每一个W均是0.25,这时候电脑按这个公式自动计算,Y=X(1)*W(1)+X(2)*W(2)+X(3)*W(3)+X(4)*W(4)得出一个结果Y,这个Y要和一个阈值（设为Q）进行比较，如果Y>Q,那么电脑就判定这个人是阿猫,否则判定不是阿猫.由于第一次计算电脑没有经验,所以结果是随机的.一般我们设定是正确的,因为我们输入的就是阿猫的身体数据啊.

现在还是阿猫站在电脑面前,不过阿猫怕被电脑认出来,所以换了一件衣服,这个行为会影响阿猫的体形,也就是X(3)变了,那么最后计算的Y值也就变了,它和Q比较的结果随即发生变化,这时候电脑的判断失误,它的结论是这个人不是阿猫.但是我们告诉它这个人就是阿猫,电脑就会追溯自己的判断过程,到底是哪一步出错了,结果发现原来阿猫体形X(3)这个体征的变化导致了其判断失误,很显然,体形X(3)欺骗了它,这个属性在人的识别中不是那么重要,电脑自动修改其权重W(3),第一次我对你是0.25的相信,现在我降低信任值,我0.10的相信你.修改了这个权重就意味着电脑通过学习认为体形在判断一个人是否是自己认识的人的时候并不是那么重要.这就是机器学习的一个循环.我们可以要求阿猫再穿一双高跟鞋改变一下身高这个属性,让电脑再一次进行学习,通过变换所有可能变换的外部特征,轮换让电脑学习记忆,它就会记住阿猫这个人比较关键的特征,也就是没有经过修改的特征.也就是电脑通过学习会总结出识别阿猫甚至任何一个人所依赖的关键特征.经过阿猫的训练电脑,电脑已经非常聪明了,这时你在让阿猫换身衣服或者换双鞋站在电脑前面,电脑都可以迅速的判断这个人就是阿猫.因为电脑已经不主要依据这些特征识别人了,通过改变衣服,身高骗不了它.当然,有时候如果电脑赖以判断的阿猫能为力,你

▲ 659

▼

● 36 条评论

✈ 分享

★ 收藏


♥ 感谢

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区

违法和不良信息举报：010-82716601

联系我们 © 2017 知乎



下载知乎客户端

与世界分享知识、经验

相关问题

这个人声音好熟悉,体形好熟悉,----都像自己一个朋友,就是脸长的不像.你不敢贸然上去搭讪吧(否定的判断).因为我们判定一个人是否是自己的朋友的时候依靠的关键的特征就是面部特征,而他恰恰就是改变了这一特征.当然也存在我们把一个拥有和我们朋友足够多相似特征的人判定为我们的朋友,这就是认错人的现象了.这些问题电脑也会出现.

不过这个算法还是有比较积极的意义的,实现了一定程度上的智能化.

发布于 2015-01-06

▲ 30

▼

💬 10 条评论

➦ 分享

★ 收藏

♥ 感谢

收起 ^

 知乎用户

63 人赞同了该回答

最近学习不确定优化攒了一些笔记，连着说一说也算总结一下。

零基础慎读。

结论：神经网络是一个参数数量突破天际的回归分析。

%突然发现在做算法的眼睛里神经网络就是模式识别。。

Part1 这些名字很帅的统计和优化方法是“粗暴的”

通常来说，现在解决不确定优化问题。都会使用随机模拟、神经网络和遗传算法联立的混合智能算法来求解。比如使目标变量的数学期望达到最优值的期望值模型， 概率最优的可靠度模型（机会约束）， 还有我校Liu老师常提的随即相关机会，就是你前几个优化组合一下。

其中随机模拟是一个不确定问题不可少的输入，遗传算法用来调神经网络的参数，神经网络用来逼近现实。互相配合，解决问题。在我看来是及其简单粗暴，不得已而为之的。只是用它们最干脆的思路，优化了最粗暴的方法。

比方说，你能写出一个事物期望值的显示表达，那么对参变量求导就可以找到极大值点。当然有的时候不是很方便求出全局导数，那么梯度递降、可行方向等规划学方法和数值方法也大有可为。离散的最优值问题在凸区域上的解也是很方便就可以求出的，此所谓整数优化。

这样一看这个混合算法真是毫无必要呀，那我们为什么还要用它们呢？原因是因为这是通用方法。

我来举一个例子。整数优化，我们选取适当的函数把两种变量聚为两类。

图中有圈匿的是取机器

▲ 659

▼

💬 36 条评论

➦ 分享

★ 收藏

♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区

违法和不良信息举报：010-82716601

联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

有九个参数。(均值协方差)

优化目标是对几千个数聚类之后聚类结果中训练集被识别的个数的两倍与分出的非训练集的元素个数的差值。我们知道这个值兼顾一二类错误的重要性，又是一个明确的目标，当然它越大越好。这个时候常规的优化算法就很难找出这九个参数的最优值了。

结果：

毕竟经过一次聚类算法之后，中心点的位置是迭代的结果而完全不确定的。如果能在聚类之前就知道中心点的位置，那么还聚类干嘛呢？这个时候聚类这个操作就像黑盒子，是传统优化问题极度依赖，却表达不出的优化过程。输入是几千个点，输出只是我之前说的评价函数，参数有九个，这时候不太好把他们看做神经元，因为我并没有显式的给出其激活函数，或者这里的激活函数实际上是相互耦合的，或者我没有将每一个输入都看做是均等的，所以不太好做一个典型的例子。

我选取6个染色体（开始是对训练集主成分分析的参数加上白噪声），迭代一万次之后，其9个参数依旧没有收敛。而且对于参数极度敏感，gauss核的均值稍做变化，聚类结果可能就会均分整个样本集，从而导致失败。

在这种情况下，我觉得，可能没有比遗传算法更好的调参数方法了。虽然其毫无道理，但由于其根基是遍历，只是加了稍许随机性，合理的选择解，来使得收敛变快。所以不失为一个好方法。也许有人分析其收敛速度，但我觉得并不太必要。因为它并不好看。当我们开始考虑用遍历求解一个问题的时候，我猜我们对其难度，对其解的形式，对其收敛性也不会有太好的判断。这样我们评估遗传算法的复杂度时，肯定也是要因地制宜的。有些绝大部分信息我们都都知道的问题，也许遗传算法也不慢；有些从头到尾都在拍脑袋的问题，遗传算法也能用。

Part2 神经网络就是一个“超多参数的回归分析”

那么神经网络到底是什么呢？【扯了那么多遗传算法还没有说正题】正如遗传算法是遍历所有参数选最优的过程，神经网络大致也就是遍历全部函数选最优的过程。

盗图：

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

我们知道一些输入和输出的数据，想找出其中的数量关系。但又不是那么容易看出来，比如我们不像开普勒那样一眼就能看出行星运动时半径的立方和时间平方成反比。有时我们也不太在意具体的显式表达，这个时候我们就说：你们看好了，这是一张已经训练好的神经网络，你想要知道你想求的函数值，那么你把输入丢进去，然后把输出抱走就好。虽然有的时候【绝大部分时候】，单论神经网络，我们是可以把我们想求得函数展开成为显式表达的，但是太难看了我们就不这么做。

当然也许你会说这不就是拟合吗？Indeed . 一般来说我们用来评价神经网络的函数也就是用和拟合一样的最小二乘和。神经网络的参数（权重）也就可以看做一般回归分析的参数。好，我们来看一下一般的线性回归。 $ax + b$ ，一共两个参数。我们把他们调到最优，可惜 r^2 还是只有0.3，不work。这个时候我们就想，会不会是 $ax^2 + bx + c$ 呢？一共三个参数， r^2 到了0.5，然后你心里想，那我把原函数泰勒展开:-D，假设原函数足够光滑，我们还是可以得到一个足够好的解的。

那么假设我们有一张神经网络，有6个输入神经元，15个隐层神经元，3个输出神经元，我们大概有多少个参数呢？【这里跳过了神经网络中权重和下一层的关系描述，包括激活函数，因为我觉得这毕竟不是神经网络的教材】

每个隐层都有一个linear的偏差，还有对应六个输入的权重，这一层有105个参数，然后三个输出神经元每个对16个权重参数，共有48个参数，加起来一共153个参数，(⊙ o ⊙),你用153阶泰勒展开逼近函数！老师，他作弊QAQ！【当然未必全都要变化

道理就是这个道理，简单粗暴。接下来就是些技巧了，借着计算机的东风，这些本来没什么内涵的方法都放出了其绚丽的光芒。把朴素的枚举法发挥到了极致。至于之后的选取光滑激活函数，比如常用的sigmoid，做好梯度递降，或者直接简单粗暴的按照下图方法进行训练网络，都是可以的。

- Step1 随机初始产生权重向量w
- Step2 通过交叉变异更新w
- Step3 算误差函数
- Step4 算适应度，然后轮盘赌，
- Step5 重复2到5
- Step6 给出最优权重向量

妈妈再也不用担心我不会训练神经网络了，【手动再见】，激活函数怎么选貌似不是很要紧呀。【你再想想是这样吗？】

为了满足某些强迫症，这里给出一个定理：具有一个或多个隐层的前向神经网络可以以任意精度逼近任意连续函数。[手动不规范引用 2016]再看看我之前的和泰勒展开的比较，想想魏尔施特拉斯定理，是不是也觉得可以接受了呢？

我知道知乎这样的长文，最后一般要贴个例子。

借宝定老师书上的例子：

相关问题



下载知乎客户端
与世界分享知识、经验和见解

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

至于神经网络被一些人鼓吹的神奇特效，比如这里说的可以减小白噪声。我只能表示我还需要学习一个。毕竟我没有试过用153个参数去进行回归分析，能不能消除白噪声呀~

对了，这里还要提一句，一般用153个参数回归分析，总要用154个输入对吧~
但你看神经网络，其实只要随机生成一个权重，然后调几圈就可以了【现实中确实可以不让这些参数都变化，但我不是做算法或者计算机的，所以不太清楚计算机业界怎么做。我只是做优化的时候碰到了一些这样的问题，自学了一下来吹牛而已】，感觉帅帅的呢~
【不过这倒并不是黑点，只是说有些人乱用这些现代学习方法，拿来批判统计学；现代的这些学习方法，都是在大数据科学的要求下应运而生，也不是说其粗暴就是坏事，毕竟解决了很多新的以往无法解决的问题】不过搞机器学习的一天到晚过拟合过拟合。。我觉得这参数取多点还是有道理的。。不知道有没有暴力增加神经元防止过拟合的方法。

前面的高赞有一点我不是很能理解，我觉得神经网络的多层隐层的含义可能确实有。。但是想分析其具体含义只能是看到啥说啥了。所以这样将神经网络说的和玄学一样感觉不太好。

Appendix：
聚类原理

citation:
Liu B, Theory and Practice of Uncertain Programming,
3rd ed., <http://orsc.edu.cn/>

pretty gooc 659 36 条评论 分享 收藏 感谢

相关问题

有个答主举了个例子拿20个样本进去跑:-D

所以数学背景还是很重要

编辑于 2016-02-03

▲ 63

▼

12 条评论

分享

★ 收藏

♥ 感谢

收起 ^

5756

苗加加

16 人赞同了该回答

简单：

神经网络就是通过模拟神经元的运动，获得一个可以变形为任意形状的任何维函数
函数的形状取决于你告诉他哪些点应该是这个函数应该尽量经过的

关于形象和有趣。。实在是不擅长。。

发布于 2014-02-13

▲ 16


▼

2 条评论

分享

★ 收藏

♥ 感谢



毛毛林

选择自己真正喜欢的

14 人赞同了该回答

曾经做过一个计算机识别美女的程序，效果渣到可以把凤姐当作美女。不过还是说说吧
找二十个女性图片正面照，做成标准格式（此处不展开）编程测量眼睛大小，宽度，鼻子大小，高度，宽度等等你认为可以衡量美女的特点作为自变量。然后自己给这二十个女性的美貌打分0至100。把自变量和因变量一一对应输入计算机，让计算机学习
之后只要把未判定的图片给计算机，计算机就会根据你的审美打分了（一秒钟在一百个头像中找到凤姐发送约吗我会乱说？）

那么如何才能避开凤姐找到范冰冰呢？

- 1，增加学习样本量，上万为佳，越大越好
- 2，换个算法，这也是神经网络最核心的部分，或选择合适的层数，做非线性变换等

发布于 2015-01-05

▲ 14

▼

8 条评论

分享

★ 收藏

♥ 感谢



罗杰

15 人赞同了该回答

从函数上来看，我觉得神经网络是回归方程的级联叠加

发布于 2014-05-28

▲ 15


▼

2 条评论

分享

★ 收藏

♥ 感谢



百得

对世界好奇

7 人赞同了该回答

建议提炼或者修改一下问题。

我无法回答。

L S 各位的回答都不是简单形象又有趣的，对于题目来说都是费力不讨好的。

题主对于“神经网络是什么”这个命题举出那个选芒果的例子，我认为又是文不对题的，后者应该是问题归集和算法思路，虽然是研究神经网络当中的必要基础之一，但是它不是神经网络的特异性问题，跟“神经网络”以及“是什么”还是有非常大的距离。例如数学也是研究神经网络当中的必要基础之一……

发布于 2014-05-07

▲ 7

▼

添加评论

分享

★ 收藏

♥ 感谢



知乎

▲ 659

▼

36 条评论

分享

★ 收藏

♥ 感谢

收起 ^

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

10 人赞同了该回答

补充：

之前的回答更多的是看了一个简单的神经网络的代码，理解的不够深刻，最近看了更多的神经网络，输出可能是有多个结果(通过输入和输出中间的矩阵)的，就有可能不是降维的过程。

再给一个答案吧：

神经网络都应该是反向神经网络(如有其他的请指正)，神经网络是模拟人类的神经元。比如说：小时候不知道手放在打开盖的暖壶上，是不是烫的。输入有：暖壶的形状，暖壶的重量（判断有没有水），不考虑其他因素。输出：不烫/烫。我们第一次的时候不知道是不是烫的，妈妈告诉我们，是烫的，我们不相信。自己试发现是烫的。

当自己试验过一次后，就记住了。这个记忆就是连接输入和输出的矩阵。如果有多条数据，通过结果反向推到输入，调整矩阵，得到了神经网络的模型。

矩阵如果太抽象，把他理解成记忆的线条，就像哈利波特中记忆丝一样。

如有不对，欢迎指正

以前的答案：

写的答案太长了，直逼本质。神经网络算法，是对数据降维的过程。比如一个杯子有好多性状，再比如说就3个，有把，圆柱形，上面开口。我们看到一个物体，分析有没有这3个性状。得出是不是1个杯子。

3个性状就3个维度，是不是个杯子，就1维。

模拟人类的模式识别，需要第一个答案那样大型的神经网络。算是人工智能的一部分。如果计算机能把人类的各个部分，都弄出来，人工智能也算走出来了。

先决条件是首先要认识到人类的各个部分。模式识别算是一个部分。

编辑于 2017-04-06


▲ 10 ▼

● 2 条评论

➦ 分享

★ 收藏

♥ 感谢

 江何

看到社会阴暗的一面 才是看到社会真实的一面

7 人赞同了该回答

1 大多数情况下，神经网络的目的就是把多维的输入，经过处理以一维进行输出。比如输入一个28*28的图像，输出一个表示该图像的类别的数字。

2 单层神经网络只能应付线性可分的情景； 多层神经网络则可以任意拟合复杂的函数。

3 神经网络需要先进行训练，再进行识别。相比其他的分类算法，神经网络的特点就是训练时间长而识别时间短。

发布于 2015-09-05

▲ 7 ▼

● 1 条评论

➦ 分享

★ 收藏

♥ 感谢

 知乎用户

11 人赞同了该回答

用来逼近目标函数的模型。
本质就这样，听不懂也没必要懂了

发布于 2015-01-05


▲ 11 ▼

● 添加评论

➦ 分享

★ 收藏

♥ 感谢

 陈丹

理工男+伪文青

12 人赞同了该回答

▲ 659 ▼

● 36 条评论

➦ 分享

★ 收藏

♥ 感谢

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 · 侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

我也来试试讲一讲：
[神经网络：文艺女vs理工男 - Candela - 知乎专栏](#)

相关 Live 推荐

刘看山 · 知乎指南 · 知乎协议 · 应用 ·
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
联系我们 © 2017 知乎

发布于 2015-10-13

▲ 12 ▼

● 1 条评论

🚩 分享

★ 收藏

♥ 感谢

更多

10 个回答被折叠（为什么？）

▲ 659 ▼

● 36 条评论

🚩 分享

★ 收藏

♥ 感谢

收起 ^

