

1.关于区块链及比特币及挖矿

https://blog.csdn.net/weixin_37504041/article/details/80005616

比特币：中本聪的白皮书标题是：《比特币：一种点对点的**电子现金系统**》（Bitcoin: A Peer-to-Peer Electronic Cash System）。你可以看出来，这个系统就叫比特币，比特币就是这个系统。**chain**只是系统的一部分，是系统的子集。**bitcoin**牛逼的地方不在于技术，这些技术都存在很多年了，**bitcoin**牛逼的地方在于**从经济的角度出来构建了一套系统**，而这套系统包含了这些技术。所以本质上**bitcoin**牛逼之处的确在于巧妙的结合了这些技术，但牛逼之处也在于**经济学与技术的完美结合**。而后来的区块链系统，大多偏离了**bitcoin**的设计思路。比如很多**blockchain**系统纯粹称为了**数据共享框架**了。

用一个跳广场舞的例子来解释一下：假设广场舞俱乐部为了鼓励大家多来跳舞，规定每来一次加2分，无故不来扣3分，领舞一次加1分……积分变动的情况由俱乐部的王阿姨来记录。积分可以在年底兑换花生油等奖品。其实，现实生活中，**银行的职责和王阿姨的职责是一样的——都是记账**（只不过王阿姨记的是积分，银行记的是钱）。发工资了，银行给你的账户上加2000块钱；买了一瓶洗发水减20元；存款一年加利息100元；转账就是在你账户上减1000，在对方账户上加1000……但银行与王阿姨最大的不同在于，王阿姨是免费为大家服务的，银行却要问我们收钱，办卡费、年费、转账费……看看银行有多少座高楼大厦、雇了多少员工，就知道银行收了我们多少钱了。如果某天我们不想要银行或王阿姨记账了，可以实现吗？当然可以了，比特币就是这样的。我们可以把使用比特币的人看成是一个广场舞俱乐部的成员。这个俱乐部的积分是这样规定的：1、每个人最初的积分（就是比特币）数额都是0；2、社群里的人共同维护一个账本，无论是否有变动，账本每10分钟更新一次，更新后要通知社群所有人；3、社群里的任何一个人都有权记账，只要记账，就能获得50个比特币的奖励（每四年减半，现在是12.5个比特币）；4、所有想记账的人都要算同一道数学题，第一个算对的人才资格记账；5、比特币的上限是2100万枚。

挖矿：中本聪就想出一个办法，出一道数学计算题，谁算得快，算的准，谁就是冠军，这个比特币就归谁。要解开这个计算题，就看谁的算力高（单位时间内的运算速度）谁就越有可能算出来，有时候也要讲一点运气。算力高低，就看你投入的计算机数量，cpu速度，内存等等。就好比，有一个空的游泳池，谁先填满谁就赢，当然是你能叫来越多人来挑水，就越快啊。比特币的诞生机制，挖矿本身指的是**通过算力去计算下一个哈希值的过程**。谁能第一个计算出来，并通知全网得到验证，谁就算挖到了这个区块，拥有这个区块的奖励和打包的矿工费。

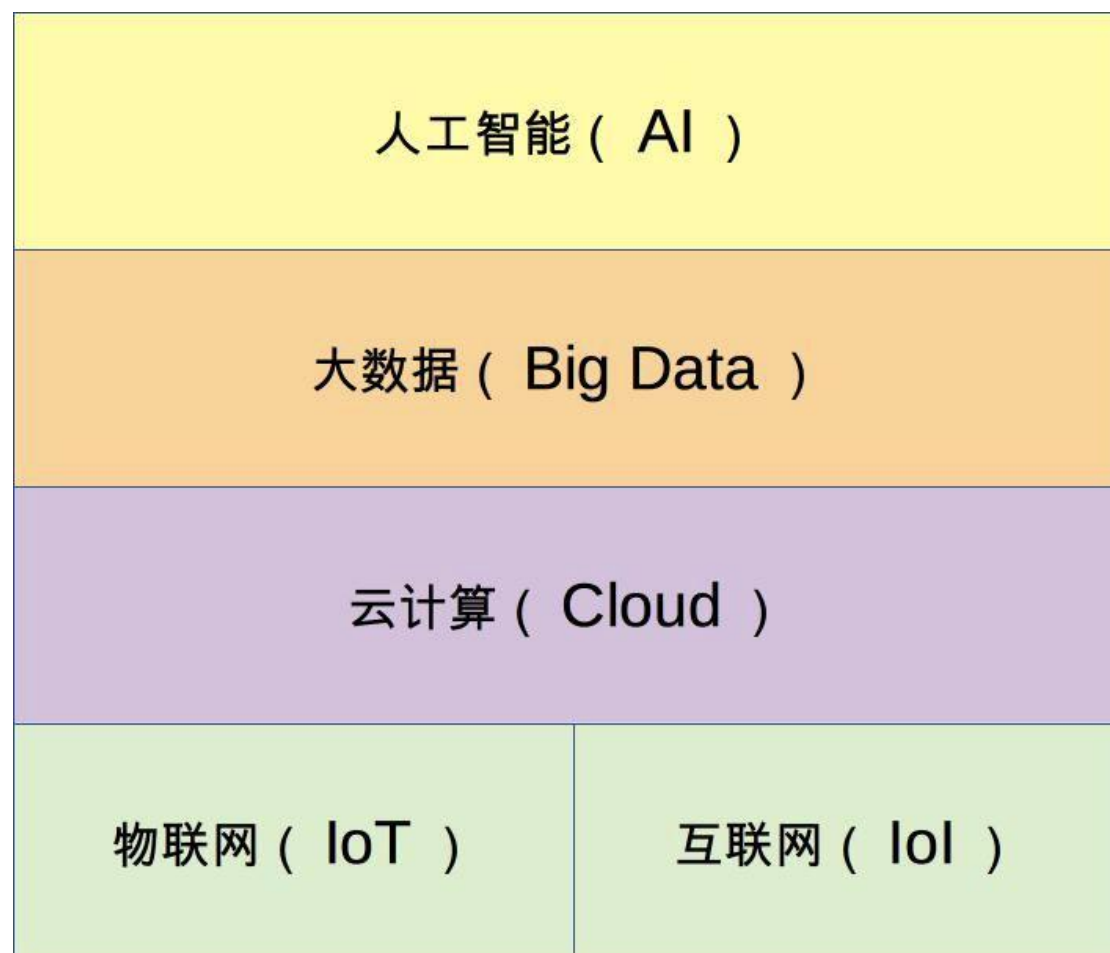
哈希值：哈希值可以称为哈希散列，哈希算法。是密码学里经典的技术，是根据文件大小、类型、时间、创作者等用计算机计算出来的，很容易就会出现变化，例如一个区块中储存的数据，就算只改了一个字母，他的哈希值也会发生变化。就是因为容易发生变化，所以没有人捉住下一个区块的哈希值是多少，也没有更改它的软件，所以哈希值的安全性非常高，伪造哈希值的难度非常高。而哈希值的主要作用是检验区块中的信息是不是有效。

区块链：区块链（Blockchain）是指通过去中心化和去信任的方式集体维护一个可靠数据库的技术方案。通俗一点说，**区块链技术就指一种全民参与记账的方式**。所有的系统背后都有一个数据库，你可以把数据库看成是就是一个大账本。那么谁来记这个账本就变得很重要。目前就是谁的系统谁来记账，微信的账本就是腾讯在记，淘宝的账本就是阿里在记。但现在区块链系统中，系统中的每个人都可以有机会参与记账。在一定时间段内如果有任何数据变

化，系统中每个人都可以来进行记账，系统会评判这段时间内记账最快最好的人，把他记录的内容写到账本，并将这段时间内账本内容发给系统内所有的其他人进行备份。这样系统中的每个人都了一本完整的账本。这种方式，我们就称它为区块链技术。

区块链解决的问题：区块链最重要的是解决了中介信用问题。在过去，两个互不认识 and 信任的人要达成协作是难的，必须要依靠第三方。比如支付行为，在过去任何一种转账，必须要有银行或者支付宝这样的机构存在。但是通过区块链技术，比特币是人类第一次实现在没有任何中介机构参与的情况下，完成双方可以互信的转账行为。这是区块链的重大突破。

云计算和区块链：云计算通常定义为通过互联网来提供动态易扩展且经常是虚拟化的资源，但是提供云计算平台的往往是一个中心化机构。而区块链组成的网络一般是没有特定的机构，所以区块链更接近分布式计算系统的定义，属于分布式计算的一种。不过，区块链是能够实现云存储的，不同于目前中心化提供云存储空间，区块链有一些提供去中心化的云存储方案。这样的项目包括 Storj, Sia, Maidsafe。



<https://zhuanlan.zhihu.com/p/29661821>

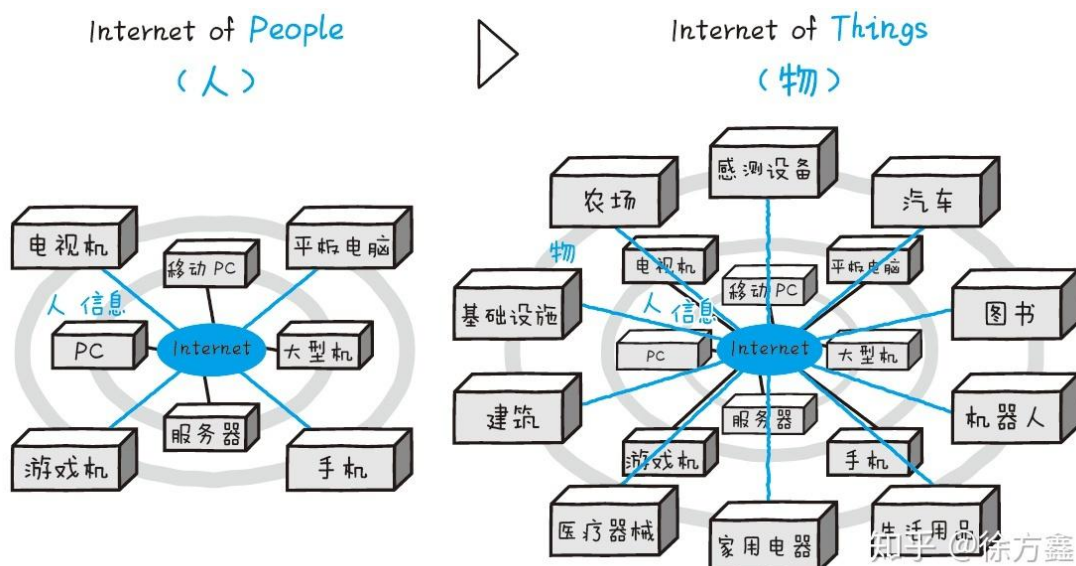
2.物联网 (IoT, Internet of Things)、互联网 (Iol, Internet of Information)：

<https://www.zhihu.com/question/19751763>

这两张网是用来将所有事物和信息联系起来，为何要联系起来呢？因为将事物和信息联系起来后，数据才有了关联，数据有了关联才能产生更大的价值。例如一辆车的位置数据没有太大价值，但几千辆车的位置数据关联起来，就可以用来判断路面拥堵情况，也可以用于交通

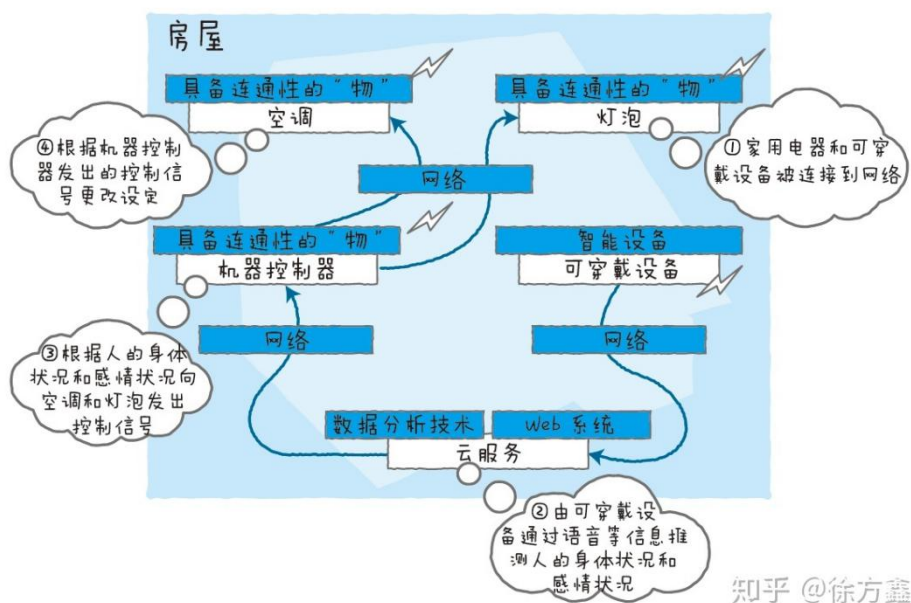
调度。

”物联网“如果按英文而言是 Internet of Things，所以本身的定义上还是集中于 Internet，是一种物与物之间的连接方式。不过就目前物联网的发展趋势而言，其已经不是一个单纯的 Internet，而进一步成为一种抽象意义的系统，或者说生态，我们目前更多的不是在关注其网络的连接，而是利用物联网构成了一种什么样的系统。



往小的说，传统的 internet 是关于人的连接网络，主要是人与人时间传递信息，而发展到物联网之后，就不仅仅是人与人之间的连接了，更多的是人与物，物与物之间的连接关系，形形色色的“物”都可以通过 internet 相连。

物联网本身是从连接的方式来定义的，但是目前随着物联网的发展，物联网更多的是与智能结合在一起。物联网目前应该是包含了传感器，网络连接，终端设备三个角色的具有智能的反馈型网络。还是借用《图解物联网》给的例子来说，目前我非常认同的观点是，物联网必须和智能化连接在一起的，那么下图是描述了一个场景：



那么大致是 4 个步骤：连接：家用电器和可穿戴设备要连接到网络。传感：所有的设备要采集信息，反馈到云服务器，反馈的内容包含人的身体情况，传感器数据等等。分析：云服务器采用数据分析技术（也就是 AI 技术了），分析人的身体情况之类，接着向空调或者灯泡发出控制信号，调温度之类。控制：最后是根据这些控制信息，对电器进行具体的控制。

青岛工业互联网：工业互联网的本质，就是通过开放的、全球化的通信网络平台，把设备、生产线、员工、工厂、仓库、供应商、产品和客户紧密地连接起来，共享工业生产全流程的各种要素资源，使其数字化、网络化、自动化、智能化，从而实现效率提升和成本降低。

“工业互联网，就是把人、数据和机器连接起来”。

<https://www.zhihu.com/question/59039920/answer/683176380>

3.关于云计算：<https://www.zhihu.com/question/20449565/answer/220276973>

物联网和互联网产生大量的数据，这些数据肯定要找地方集中存储和处理，这就必须要有云计算了。如果没有云计算，一台冰箱产生的数据都要部署独立一台后台服务器来接收，成本和便利性无法接受。云计算的作用就在于将海量数据集中存储和处理。

云计算产生的背景：传统的应用正在变得越来越复杂：需要支持更多的用户、需要更强的计算能力、需要更加稳定安全等等而为了支撑这些不断增长的需求，企业不得不去购买各类硬件设备（服务器，存储，带宽等等）和软件（数据库，中间件等等），另外还需要组建一个完整的运维团队来支持这些设备或软件的正常运作，这些维护工作就包括安装、配置、测试、运行、升级以及保证系统的安全等。便会发现支持这些应用的开销变得非常巨大，而且它们的费用会随着你应用的数量或规模的增加而不断提高。这也是为什么即使是在那些拥有很出色 IT 部门的大企业中，那些用户仍在不断抱怨他们所使用的系统难以满足他们的需求。而对于那些中小规模的企业，甚至个人创业者来说，创造软件产品的运维成本就更加难以承受了基于此，云计算应运而生：更大量、更快速、更强大

定义：云计算是将我们传统的 IT 工作转为以网络为依托的云平台运行，NIST（美国国家标准与技术研究院）在 2011 年下半年公布了云计算定义的最终稿，给出了云计算模式所具备的 5 个基本特征（按需自助服务、广泛的网络访问、资源共享、快速的可伸缩性和可度量的服务）、3 种服务模式（SaaS（软件即服务）、PaaS（平台即服务）和 IaaS（基础设施即服务））和 4 种部署方式（私有云、社区云、公有云和混合云）

4.关于大数据

海量数据上传到云计算平台后，自然而然的就需要对数据进行深入分析和挖掘了，这就是大数据的目的。将几千辆车的位置信息综合起来分析出某条路的拥堵状况；将某个城市几百万人的健康状况综合分析，也许就可以得出某个工厂周围某种疾病的发病率比较高的结论。。。。。。这些都是大数据做的事情。

定义：大数据本身是一个抽象的概念。从一般意义上讲，大数据是指无法在有限时间内用常规软件工具对其进行获取、存储、管理和处理的数据集合。大数据技术可以定义为一种软件实用程序，旨在分析、处理和提取来自极其复杂的大型数据集的信息，而传统数据处理软件永远无法处理这些信息。

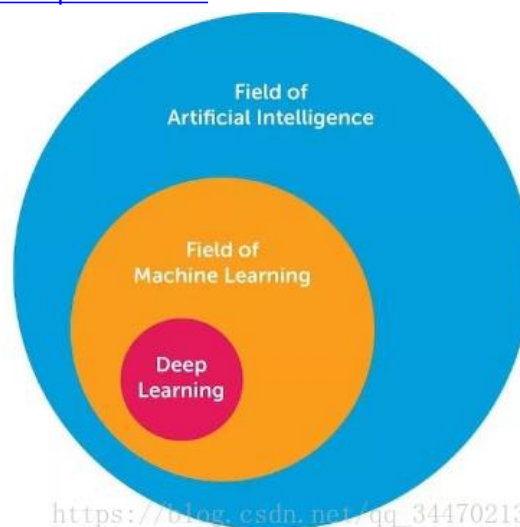
目前，业界对大数据还没有一个统一的定义，但是大家普遍认为，大数据具备 Volume、Velocity、Variety 和 Value 四个特征，简称“4V”，即数据体量巨大、数据速度快、数据类型繁多和数据价值密度低

5.关于机器学习、深度学习、人工智能

大数据是基于海量数据进行分析从而发现一些隐藏的规律、现象、原理等，而人工智能在大数据的基础上更进一步，人工智能会分析数据，然后根据分析结果做出行动，例如无人驾驶，自动医学诊断。

区分人工智能、机器学习、深度学习：

<https://zhuanlan.zhihu.com/p/26442277>



人工智能（英语：Artificial Intelligence, AI）：通过计算机程序实现人类智能。

详细：是指由人工制造出来的系统所表现出来的智能。通常人工智能是指通过普通电脑实现的智能。人工智能的研究可以分为几个技术问题。其分支领域主要集中在解决具体问题，其中之一是，如何使用各种不同的工具完成特定的应用程序。AI的核心问题包括推理、知识、规划、学习、交流、感知、移动和操作物体的能力等。目前有大量的工具应用了人工智能，其中包括搜索和数学优化、逻辑推演。而基于仿生学、认知心理学，以及基于概率论和经济学的算法等等也在逐步探索当中。

机器学习（英语：Machine Learning）：人工智能的一个分支。从数据中自动分析获得规律（模型），并利用规律对未知数据进行预测。

详细：是人工智能的一个分支。人工智能的研究是从以“推理”为重点到以“知识”为重点，再到以“学习”为重点，一条自然、清晰的脉络。显然，机器学习是实现人工智能的一个途径，即以机器学习为手段解决人工智能中的问题。机器学习在近30多年已发展为一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、计算复杂性理论等多门学科。机器学习理论主要是设计和分析一些让计算机可以自动“学习”的算法。**机器学习算法是一类从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法。**因为学习算法中涉及了大量的统计学理论，机器学习与推断统计学联系尤为密切，也被称为统计学习理论。算法设计

方面，机器学习理论关注可以实现的，行之有效的学习算法。很多推论问题属于无程序可循难度，所以部分的机器学习研究是开发容易处理的近似算法。

机器学习有下面几种定义：

- 1.机器学习是一门人工智能的科学，该领域的主要研究对象是人工智能，特别是如何在经验。
- 2 学习中改善具体算法的性能。
- 3.机器学习是对能通过经验自动改进的计算机算法的研究。
- 4.机器学习是用数据或以往的经验，以此优化计算机程序的性能标准。

应用：机器学习已广泛应用于数据挖掘、计算机视觉、自然语言处理、生物特征识别、搜索引擎、医学诊断、检测信用卡欺诈、证券市场分析、DNA 序列测序、语音和手写识别、战略游戏和机器人等领域。

常使用的算法：

https://blog.csdn.net/sinat_23338865/article/details/81805538

监督学习：决策树、朴素贝叶斯分类、最小二乘法、逻辑回归(Logistic Regression)、支持向量机 (Support Vector Machine, SVM)

无监督学习：聚类算法、主成分分析 (PCA)

深度学习 (英语：Deep Learning) 是机器学习拉出的分支，它试图使用包含复杂结构或由多重非线性变换构成的多个处理层对数据进行高层抽象的算法。

详细：深度学习是机器学习中一种基于对数据进行表征学习的方法。观测值 (例如一幅图像) 可以使用多种方式来表示，如每个像素强度值的向量，或者更抽象地表示成一系列边、特定形状的区域等。而使用某些特定的表示方法更容易从实例中学习任务 (例如，人脸识别或面部表情识别)。深度学习的好处是用非监督式或半监督式的特征学习和分层特征提取高效算法来替代手工获取特征。

一些著名的深度学习库：**tensorflow** 最初由 Google 机器智能研究机构的 Google Brain 团队的研究人员和工程师开发。该系统旨在促进对机器学习的研究，同时也让机器学习研究原型过渡到生产系统更加高效容易。

Caffe 是一个知名的、被普遍使用的机器视觉库，其将 **Matlab** 的快速卷积网接口迁移到了 C 和 C++ 中。**Caffe** 不面向其他深度学习应用，比如文本、声音或时序数据。如同其他框架一样，**Caffe** 选择 Python 作为 API。

Keras 一个高层神经网络 API，**Keras** 由纯 Python 编写而成并基 **Tensorflow** 或 **Theano**。

深度学习算法分类：

https://blog.csdn.net/qg_34470213/article/details/79592484

深度学习网络与“典型”的前馈多层网络之间是有一些区别的，如下：

深度学习网络比之前的网络有更多的神经元

深度学习网络具有更复杂的连接层的方式

深度学习网络需要用强大的计算能力来训练

深度学习网络能够进行自动特征提取

因此深度学习可以被定义为在以下四个基本网络框架中拥有大量参数和层的神经网络：

卷积神经网络 (Convolutional Neural Networks)：基本上就是用共享权重在空间中进行扩展的标准神经网络。设计 CNN 主要是为了通过内部卷积来识别图片，内部卷积可以看到待识别物体的边。

循环神经网络 (Recurrent Neural Networks)：基本上是在时间上进行扩展的标准神经网络，因为边进入下一个时间步，而不是在同一时间步进入下一个层。设计 RNN 主要是为了识别序列，例如语音信号或者文本。它里面的循环意味着网络中存在短暂的记忆。（扩展：LSTM 长短时记忆网络）

递归神经网络 (Recursive Neural Networks)：更类似于分层网络，其中输入序列没有真正的时间面，而是输入必须以树状方式分层处理。

激活函数、优化函数分类：

https://blog.csdn.net/qg_36931982/article/details/89412759

一般说来，训练深度学习网络的方式主要有四种：监督、无监督、半监督和强化学习：

<https://blog.csdn.net/gdengden/article/details/84196883>

监督学习 (Supervised Learning)：监督学习是使用已知正确答案的示例来训练网络的。想象一下，我们可以训练一个网络，让其从照片库中（其中包含你父母的照片）识别出你父母的照片。以下就是我们在这个假设场景中所要采取的步骤。（即一个完整的训练过程）

(1) 数据集的创建和分类：首先，我们要浏览你的照片（数据集），确定所有有你父母的照片，并对其进行标注，从而开始此过程。然后我们将把整堆照片分成两堆。我们将使用第一堆来训练网络（训练数据），而通过第二堆来查看模型在选择我们父母照片操作上的准确程度（验证数据）。

等到数据集准备就绪后，我们就会将照片提供给模型。在数学上，我们的目标就是在深度网络中找到一个函数，这个函数的输入是一张照片，而当你的父母不在照片中时，其输出为 0，否则输出为 1。

此步骤通常称为分类任务（categorization task）。在这种情况下，我们进行的通常是一个结果为 yes or no 的训练，但事实是，监督学习也可以用于输出一组值，而不仅仅是 0 或 1。例如，我们可以训练一个网络，用它来输出一个人偿还信用卡贷款的概率，那么在这种情况下，输出值就是 0 到 100 之间的任意值。这些任务我们称之为回归。

(2) 训练：为了继续该过程，模型可通过以下规则（激活函数）对每张照片进行预测，从而决定是否点亮工作中的特定节点。这个模型每次从左到右在一个层上操作——现在我们将更复杂的网络忽略掉。当网络为网络中的每个节点计算好这一点后，我们将到达亮起（或未亮起）的最右边的节点（输出节点）。

既然我们已经知道有你父母的照片是哪些图片，那么我们就可以告诉模型它的预测是对还是错。然后我们会将这些信息反馈（feed back）给网络。

该算法使用的这种反馈，就是一个量化“真实答案与模型预测有多少偏差”的函数的结果。这个函数被称为成本函数（cost function），也称为目标函数（objective function），效用函数（utility function）或适应度函数（fitness function）。然后，该函数的结果用于修改一个称为反向传播（backpropagation）过程中节点之间的连接强度和偏差，因为信息从结果节点“向后”传播。

我们会为每个图片都重复一遍此操作，而在每种情况下，算法都在尽量最小化成本函数。其实，我们有多种数学技术可以用来验证这个模型是正确还是错误的，但我们常用的是一个非常常见的方法，我们称之为梯度下降（gradient descent）。Algobeans 上有一个“门外

汉”理论可以很好地解释它是如何工作的。迈克尔尼尔森 (Michael Nielsen) 用数学知识完善了这个方法, 其中包括微积分和线性代数。

(3) **验证:** 一旦我们处理了第一个堆栈中的所有照片, 我们就应该准备去测试该模型。我们应充分利用好第二堆照片, 并使用它们来验证训练有素的模型是否可以准确地挑选出含有你父母在内的照片。我们通常会通过调整和模型相关的各种事物 (超参数) 来重复步骤 2 和 3, 诸如里面有多少个节点, 有多少层, 哪些数学函数用于决定节点是否亮起, 如何在反向传播阶段积极有效地训练权重, 等等。而你可以通过浏览 Quora 上的相关介绍来理解这一点, 它会给你一个很好的解释。

(4) **使用:** 最后, 一旦你有了一个准确的模型, 你就可以将该模型部署到你的应用程序中。你可以将模型定义为 API 调用, 例如 `ParentsInPicture(photo)`, 并且你可以从软件中调用该方法, 从而导致模型进行推理并给出相应的结果。

无监督学习: (Unsupervised Learning) 无监督学习适用于你具有数据集但无标签的情况。无监督学习采用输入集, 并尝试查找数据中的模式。比如, 将其组织成群 (聚类) 或查找异常值 (异常检测)。例如: 想像一下, 如果你是一个 T 恤制造商, 拥有一堆人的身体测量值。那么你可能就会想要有一个聚类算法, 以便将这些测量组合成一组集群, 从而决定你生产的 XS, S, M, L 和 XL 号衬衫该有多大。

如果你是一家安全初创企业的首席技术官 (CTO), 你希望找出计算机之间网络连接历史中的异常: 网络流量看起来不正常, 这可能会帮助你通过下载员工们的所有 CRM 历史记录来找到那名该为此事负责的员工, 因为他们可能即将退出或有人正在将异常大量的钱转移到一个新的银行账户。如果你对这种事情感兴趣的话, 那么我相信你会很喜欢这种对无监督异常检测算法的调查。

假设一下, 你是 Google Brain 团队中的一员, 你想知道 YouTube 视频中有什么。谷歌通过人工智能在视频网站中找到猫的真实故事, 唤起了大众对 AI 的热忱。在诸如这篇论文中, Google Brain 团队与斯坦福大学研究人员 Quoc Le 和吴恩达一起描述了一种将 YouTube 视频分为多种类别的算法, 其中一种包含了猫类别。当然他们并没有真正开始寻找猫, 但算法自动将包含猫的视频 (以及 ImageNet 中定义的 22000 个对象类别中的数千个其他对象) 组合在一起, 而不需要任何明确的训练数据。

你将在文献中阅读到的一些无监督的学习技术包括:

自编码 (Autoencoding)、主成分分析 (Principal components analysis)、随机森林 (Random forests)、K 均值聚类 (K-means clustering)

半监督学习 (Semi-supervised Learning): 半监督学习在训练阶段结合了大量未标记的数据和少量标签数据。与使用所有标签数据的模型相比, 使用训练集的训练模型在训练时可以更为准确, 而且训练成本更低。举个例子来说明, 我们的朋友 Delip Rao 在 AI 咨询公司 Joostware 工作, 他构建了一个使用半监督学习的解决方案, 每个类中只需使用 30 个标签, 就可以达到与使用监督学习训练的模型相同的准确度, 而在这个监督学习模型中, 每个类中需要 1360 个左右的标签。因此, 这个半监督学习方案使得他们的客户能够非常快地将其预测功能从 20 个类别扩展到 110 个类别。

为什么使用未标记数据有时可以帮助模型更准确, 关于这一点的体会就是: 即使你不知道答案, 但你也可以通过学习来知晓, 有关可能的值是多少以及特定值出现的频率。

强化学习 (Reinforcement Learning): 强化学习是针对你再次没有标注数据集的情况而言的, 但你还是有办法来区分是否越来越接近目标 (回报函数 (reward function))。经典的儿童游戏——“hotter or colder”。(Huckle Buckle Beanstalk 的一个变体) 是这个概念

的一个很好的例证。你的任务是找到一个隐藏的目标物件，然后你的朋友会喊出你是否越来越 **hotter**（更接近）或 **colder**（远离）目标物件。“**Hotter/colder**”就是回报函数，而算法的目标就是最大化回报函数。你可以把回报函数当做是一种延迟和稀疏的标签数据形式：而不是在每个数据点中获得特定的“**right/wrong**”答案，你会得到一个延迟的反应，而它只会提示你是否在朝着目标方向前进。

DeepMind 在 *Nature* 上发表了一篇文章，描述了一个将强化学习与深度学习结合起来的系统，该系统学会如何去玩一套 Atari 视频游戏，一些取得了巨大成功（如 **Breakout**），而另一些就没那么幸运了（如 **Montezuma's Revenge**（蒙特祖玛的复仇））。

计算机视觉：

OpenCV（Open source Computer Vision Library，开放源代码计算机视觉库）是一个**用于图像处理、分析、机器视觉方面的开源函数库**。

无论你是做科学研究，还是商业应用，**opencv**都可以作为你理想的工具库，因为，对于这两者，它完全是免费的。该库采用 C 及 C++语言编写，可以在 windows, linux, mac OSX 系统上面运行。该库的所有代码都经过优化，计算效率很高，因为，它更专注于设计成为一种用于实时系统的开源库。**opencv**采用 C 语言进行优化，而且，在多核机器上面，其运行速度会更快。它的一个目标是提供友好的机器视觉接口函数，从而使得复杂的机器视觉产品可以加速面世。该库包含了横跨工业产品检测、医学图像处理、安防、用户界面、摄像头标定、三维成像、机器视觉等领域的超过 500 个接口函数。

同时，**由于计算机视觉与机器学习密不可分，该库也包含了比较常用的一些机器学习算法**。或许，很多人知道，图像识别、机器视觉在安防领域有所应用。但，很少有人知道，在航拍图片、街道图片（例如 **google street view**）中，要严重依赖于机器视觉的摄像头标定、图像融合等技术。近年来，在入侵检测、特定目标跟踪、目标检测、人脸检测、人脸识别、人脸跟踪等领域，**opencv**可谓大显身手，而这些，仅仅是其应用的冰山一角。

一个典型的计算机视觉算法，应该包含以下一些步骤：

https://blog.csdn.net/carson2005/article/details/6979806?utm_medium=distribute.pc_relevant_t0.none-task-blog-BlogCommendFromMachineLearnPai2-1.control&dist_request_id=1328690.11451.16165880174186737&depth_1-utm_source=distribute.pc_relevant_t0.none-task-bl

- (1)数据获取（对 OpenCV 来说，就是图片）；
- (2)预处理；
- (3)特征提取；
- (4)特征选择；
- (5)分类器设计与训练；
- (6)分类判别；

OpenCV 对这六个部分，分别（记住这个词）提供了 API。下面我分别就这六个部分对一些常见问题进行必要的解释。

一个完整的项目流程