

## 线性回归

用一条直线拟合数据的方法称为线性回归，注意线性回归的说法经常会涉及到高维空间。

## 有监督学习

所谓有监督学习，就是先利用有标签的训练数据学习得到一个模型，然后使用这个模型对新样本进行预测。在本质上，监督学习的目标在于，构建一个由输入到输出的映射，该映射用模型来表示。

## 无监督学习

与监督学习相反，无监督学习的数据是没有标签的，我们把数据给机器，并让它试着从这些数据中找到某种结构。无监督学习使我们能够在不知道结果应该是什么的情况下处理问题。我们可以从不一定知道变量影响的数据中导出结构。

简单来说，无监督学习中的数据没有任何标签，拿到数据集，我们并不知道要拿它来做什么，也不知道每个数据点究竟是什么，而只是被告知这里有一个数据集，能否在其中找到某种有意义的结构。

## 神经网络

神经网络是一种模仿生物神经网络的结构和功能的数学模型或计算模型。由大量的人工神经元联结进行计算。大多数情况下人工神经网络能在外界信息的基础上改变内部结构，是一种自适应系统。

现代神经网络是一种非线性统计性数据建模工具，常用来对输入和输出间复杂的关系进行建模，或用来探索数据的模式。

## 神经元

神经元是构成神经网络的基本单元，其主要是模拟生物神经元的结构和特性，接受一组输入信号并产出输出。通过神经元，人工神经网络可以以数学模型模拟神经元活动，继而进行高效的计算以及其他处理。

神经元的机制是指输入信号经过加权和偏置处理后传递到下一层，激活下一层的神经元。

## 神经网络层

神经网络层是神经网络的基本单位，层之间是相互独立的，可以理解为神经元的集合。

## 激活层

进行非线性整合，模拟神经元的激活行为。之所以为什么要加入非线性因素，是因为线性模型表达不够。

## 激活函数

所谓激活函数(Activation Function)，就是在人工神经网络的神经元上运行的函数，负责将神经元的输入映射到输出端。一般神经网络是线性的，引入激活函数，是在神经网络中引入非线性，强化网络的学习能力。所以 激活函数的最大特点就是非线性。常见的激活函数有 sigmoid、tanh、softsign、ReLU、softplus、logistics 等。

## 全连接层

全连接层的每一个节点都与上一层所有的节点相连，从而把前边提取到的特

征综合起来。将图像展开，进行分类操作。

## 激活函数

激活函数是神经元的一个重要部件，它对输入信号进行非线性转换，并把非线性信号转换为线性信号，从而实现非线性映射。

## 权重

权重是一个相对的概念，是针对某一指标而言。某一指标的权重是指该指标在整体评价中的相对重要程度。

## 偏置 (bias)

偏置单元 (biasunit)，在有些资料里也称为偏置项 (bias term) 或者截距项 (intercept term)，它其实就是函数的截距，与线性方程  $y=wx+b$  中的  $b$  的意义是一致的。在  $y=wx+b$  中， $b$  表示函数在  $y$  轴上的截距，控制着函数偏离原点的距离，其实在神经网络中的偏置单元也是类似的作用。偏置可以用来避免输出为零的情况，并且能够加速某些操作，这让解决某个问题所需要的神经元数量也有所减少。

## 前向传播

简单理解就是将上一层的输出作为下一层的输入，并计算下一层的输出，一直到运算到输出层为止。

## 反向传播

反向传播算法定义为在已知分类的情况下，为给定的输入模式训练某些给定的前馈神经网络的算法。反向传播是在链式法则的基础上实现的。

## 链式法则

链式法则是微积分中的求导法则，用于求一个复合函数的导数，是在微积分的求导运算中一种常用的方法。复合函数的导数将是构成复合这有限个函数在相应点的导数的乘积，就像锁链一样一环套一环，故称链式法则。

## 梯度下降

梯度下降算法是一种优化算法，它通过不断调整参数来最小化损失函数。

## 损失函数

在训练模型中，用来衡量预测值和真实值之间的误差，就叫损失函数，数值越小表示的误差越小。

## 学习率

每一步梯度下降时向目标方向前行的长度。将输出误差反向传播给网络参数，以此来拟合样本的输出，本质上是最优化的一个过程，逐步趋向于最优解，但是每一次更新参数利用多少误差，就需要通过一个参数来确定，这个参数就是学习率。

## Transformer 模型

Transformer 是 Google 的研究者于 2017 年在《AttentionIsAllYouNeed》一文中提出的一种用于 seq2seq 任务的模型，它没有 RNN 的循环结构或 CNN 的卷积结构，在机器翻译等任务中取得了一定提升。目前也在 CV 领域中有着广泛的应用。

## 注意力机制

注意力机制(Attention Mechanism)是人们在机器学习模型中嵌入的一种特殊结构,用来自动学习和计算输入数据对输出数据的贡献大小。它实际上就是将人的感知方式、注意力的行为应用在机器上,让机器学会去感知数据中的重要和不重要的部分。

## 前馈神经网络

各神经元分别属于不同的层,层内无连接,而是层与层之间的连接。整个网络无反馈,信号从输入层向输出层单向传播。

### 输入层

输入层主要用于获取输入信息。在神经网络中,输入层接收信号并将其传递给隐藏层。隐藏层通过神经元进行特征提取,每个隐藏层神经元与输入层神经元之间的连接具有不同的权重,并且每个神经元还包含一个偏置值。输入层的信号经过加权和偏置处理后传递到隐藏层,激活隐藏层的神经元。最后,输出层根据隐藏层的输出、各自的权重和偏置值计算并生成最终结果。

### 输出层

输出层负责对接隐藏层并输出最后的结果。

### 隐藏层

除输入层、输出层以外的层叫做隐藏层,隐藏层是其他几个功能层的总称。主要进行特征提取,调整权重让隐藏层的神经单元对某种模式形成反应。

## 卷积神经网络（CNN）

卷积神经网络(CNN)是含有卷积层的神经网络，

### 卷积核

是在卷积神经网络（CNN）中用于提取图像特征的小矩阵。

### 下采样

下采样就是图像缩小的过程。把底层图生成特征图的过程。

### 上采样

简单的理解就是把图片进行放大了，把特征图反推出原图的过程。

### 步长

卷积核每次滑动时移动的像素数。

### 感受野

感受野是指在卷积过程中卷积核关注的区域，故浅层的卷积核关注的是细节特征，例如颜色、角点、直线曲线等；高层的卷积核关注的是全局特征，例如汽车、飞机、老虎等。

## 循环神经网络

循环神经网络(RNN)的来源是为了刻画一个序列当前的输出与之前信息的关系。从网络结构上，循环神经网络会记忆之前的信息，并利用之前的信息影响后面结点的输出。即：循环神经网络的隐藏层之间的结点是有连接的，隐藏层的输入不仅包括输入层的输出，还包括上一时刻隐藏层的输出。

## 长短时记忆神经网络

长短时记忆网络 (LSTM) 是一种能够捕捉长时依赖的特殊循环神经网络。LSTM 相对于基础的 RNN 网络来说，记忆能力更强，更擅长处理较长的序列信号数据。LSTM 有三个门，输入门、输出门、遗忘门。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。门能够有选择性的决定让哪些信息通过。长短时记忆神经网络属于循环神经网络 RNN。

## 过拟合

模型在训练数据上表现非常好，但在未见过的测试数据上表现较差。这通常是由于模型学习到了训练数据中的噪声和细节导致的。过拟合现象通常与数据量不足，模型过于复杂、参数过多，或模型对训练数据中的噪声和异常值敏感等因素有关。

## 欠拟合

模型在训练数据和测试数据上都表现不佳。欠拟合产生的原因有：使用的模型复杂度过低、使用的特征量过少或特征选择不当、学习率设置不合理等因素有关。

## 梯度消失与梯度爆炸

梯度消失 (Vanishing Gradient) 是指在深度神经网络中，梯度在反向传播过程中逐渐变小，导致浅层的权重更新非常缓慢或几乎不更新，从而使模型难以训练。梯度爆炸 (Exploding Gradient) 是指在深度神经网络中，梯度在反向传播过程中逐渐变大，导致权重更新过大，使模型变得不稳定，甚至无法收敛。

梯度消失与梯度爆炸都可能与链式法则的累积效应有关，由于梯度是多个导数的乘积，因此当这些导数都小于 1 或大于 1 时，在深层网络中的梯度会变得非常小或非常大，进而导致梯度消失或梯度爆炸。