

神经网络和机器学习

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example.(是机器学习的一个子集)

需要:大量标记的数据
大量稳定的算力

函数

用固定数量的基函数来逼近原函数, 最成功的是 feedforward neural network(multilayer perceptron)

- Artificial Neuron(人工神经元)
 - 得到输入
 - 求加权和
 - 通过一个非线性函数(激活函数)
 - 输出

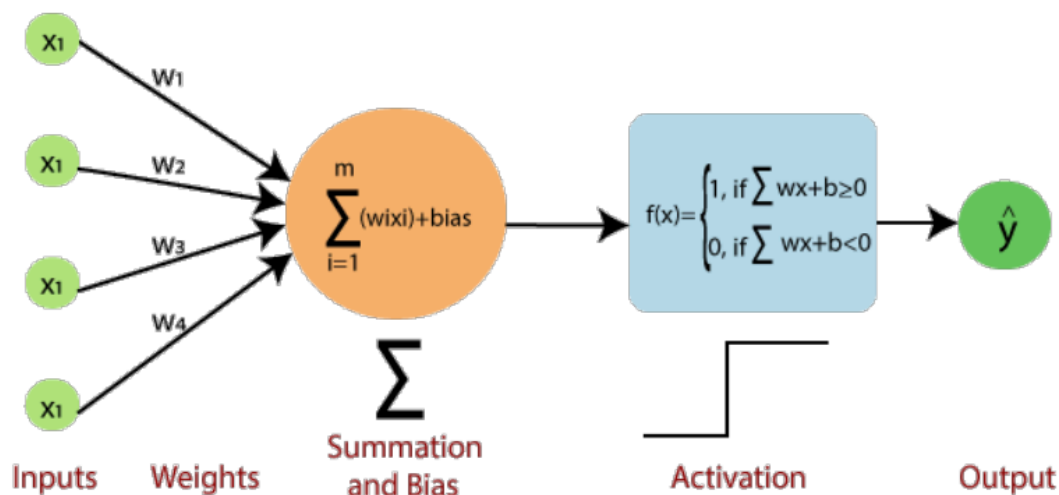


Figure 1: 一种简单感知器的结构

- 训练算法

ANN(FNN, 人工神经网络就是前馈神经网络)

- 包括一个输入层, 一个输出层, 多个隐藏层
- 每一个节点都有自己的权重和 threshold(超过这个值才会把自己的输出传递, 成为节点激活(activated))

- 关于决策边界(Decision Boundary)
 - 没有隐藏层, 线性分类(超平面)
 - 一个隐藏层, 凸区域边界(可以是开发或者封闭)
- 对于类别分类的输出, 结果必须是 0 到 1 之间而且和为 1 使用 softmax 激活函数

$$y_i = \sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

$$0 < y_i < 1$$

$$\sum y_i = 1$$

Figure 2: softmax 就是个指数占比函数

- 其他激活函数
 - sigmoid $s(z) = \frac{1}{1+e^{-z}}$
 - Tangent: $z = \tanh(x)$
 - ReLu: $r(x) = \max(0, x)$
- 分割数据集
 - Training
 - Validation(训练过程中)
 - Test(验收训练结果)

问题

Underfitting(欠拟合)

Overfitting(过拟合)

Gradient Descent(梯度下降)

直到 point of convergence

Learning Rate(学习率)

学习率高可以减少迭代次数同时也可能没法达到最优点,是一种 hyper-parameter(超参数)

Momentum

是梯度下降法的扩展,允许再某个方向上建立 inertia

Backpropagation(反向传播)

全称:backward propagation of errors(误差的反向传播)

CNN(卷积神经网络)

优点

- 相对于全连接,减少了参数量

缺点

- 比较难以设计和维护
- 相对速度较慢(受到隐藏层的个数影响)

RNN(循环神经网络)

因为要对时序数据做预测的时候产生(sequential)

RNN 具有与前馈网络相似的结构,除了各层还接收前一个实例预测的延时输入

优点

- 可以有效的对每一个数据对于之前的数据有关的建模
- 可以与卷积结合增加效率

缺点

- Gradient vanishing and exploding(梯度消失或者暴增)
- 训练难度大
- 用 ReLU 作为激活函数,难以处理长时序数据

Sequence to Sequence Models

- 包括两个 RNN
- 有一个 encoder 处理输入,一个 decoder 处理输出
- 该模型特别适用于输入数据长度等于输出数据长度的情况

Long Short-Term Memory(LSTM,长短期记忆)

增加了长期记忆单元来解决梯度消失的问题

LSTM 增加了三个门来决定哪一些输出应该被保留,哪些应该被忘记?

Input gate, Output gate, forget gate

输入门控制哪些数据应该放到记忆中

输出门控制哪些数据应该给到下一层

忘记门控制哪些数据应该被忘记

- lstm 被明确设计为避免长期依赖问题
- Remembering information for long periods of time is practically their default

behavior, not something they struggle to learn

GRU(Gated Recurrent Unit)

可以被视为 LSTM 的一种变种

Graph Neural Network