

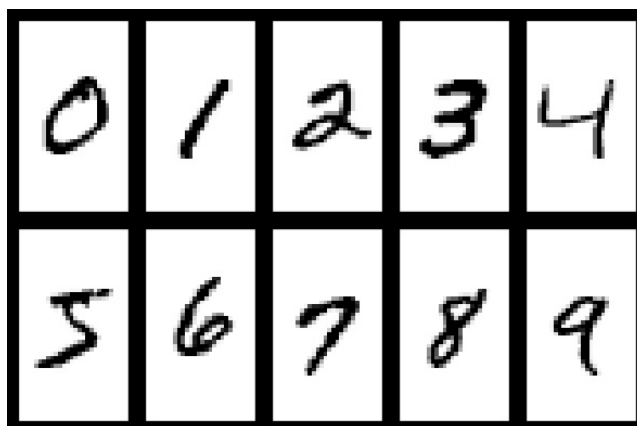
深层神经网络

前面一章我们简要介绍了神经网络的一些基本知识，同时也是示范了如何用神经网络构建一个复杂的非线性二分类器，更多的情况神经网络适合使用在更加复杂的情况，比如图像分类的问题，下面我们用深度学习的入门级数据集 MNIST 手写体分类来说明一下更深层神经网络的优良表现。

MNIST 数据集

mnist 数据集是一个非常出名的数据集，基本上很多网络都将其作为一个测试的标准，其来自美国国家标准与技术研究所, National Institute of Standards and Technology (NIST)。训练集 (training set) 由来自 250 个不同人手写的数字构成, 其中 50% 是高中学生, 50% 来自人口普查局 (the Census Bureau) 的工作人员，一共有 60000 张图片。测试集(test set) 也是同样比例的手写数字数据，一共有 10000 张图片。

每张图片大小是 28 x 28 的灰度图，如下



所以我们的任务就是给出一张图片，我们希望区别出其到底属于 0 到 9 这 10 个数字中的哪一个。

多分类问题

前面我们讲过二分类问题，现在处理的问题更加复杂，是一个 10 分类问题，统称为多分类问题，对于多分类问题而言，我们的 loss 函数使用一个更加复杂的函数，叫交叉熵。

softmax

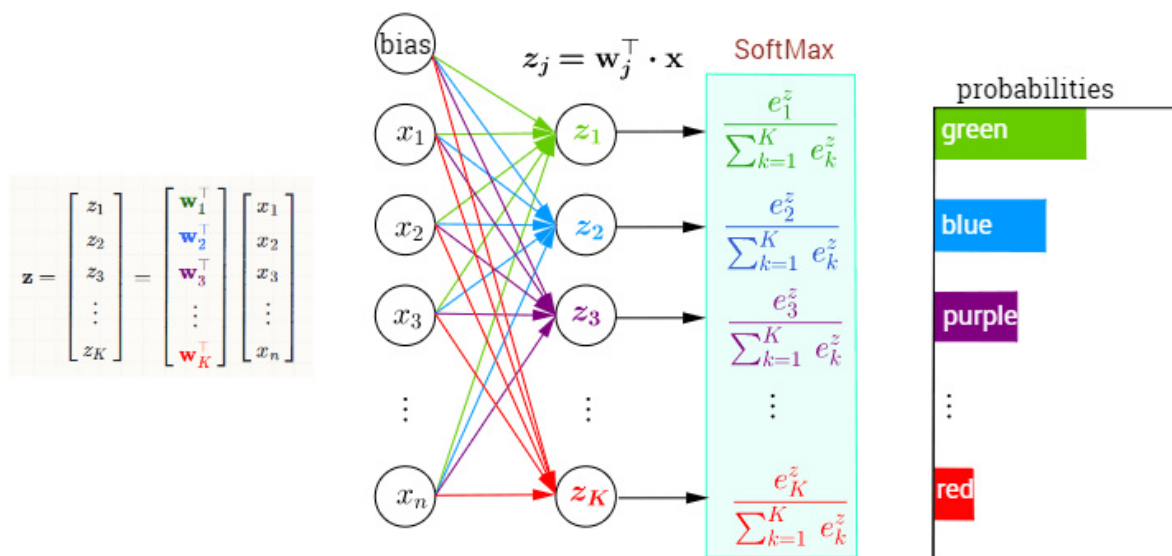
提到交叉熵，我们先讲一下 softmax 函数，前面我们见过了 sigmoid 函数，如下

$$s(x) = \frac{1}{1+e^{-x}}$$

可以将任何一个值转换到 0 ~ 1 之间，当然对于一个二分类问题，这样就足够了，因为对于二分类问题，如果不属于第一类，那么必定属于第二类，所以只需要用一个值来表示其属于其中一类概率，但是对于多分类问题，这样并不行，需要知道其属于每一类的概率，这个时候就需要 softmax 函数了。

softmax 函数示例如下

Multi-Class Classification with NN and SoftMax Function



对于网络的输出 z_1, z_2, \dots, z_k ，我们首先对他们每个都取指数变成 $e^{z_1}, e^{z_2}, \dots, e^{z_k}$ ，那么每一项都除以他们的求和，也就是

$$z_i \rightarrow \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (1)$$

如果对经过 softmax 函数的所有项求和就等于 1，所以他们每一项都分别表示属于其中某一类的概率。

交叉熵

交叉熵衡量两个分布相似性的一种度量方式，前面讲的二分类问题的 loss 函数就是交叉熵的一种特殊情况，交叉熵的一般公式为

$$cross_entropy(p, q) = E_p[-\log q] = -\frac{1}{m} \sum_x p(x) \log q(x) \quad (2)$$

对于二分类问题我们可以写成

$$-\frac{1}{m} \sum_{i=1}^m (y^i \log \text{sigmoid}(x^i) + (1 - y^i) \log(1 - \text{sigmoid}(x^i))) \quad (3)$$

这就是我们之前讲的二分类问题的 loss，当时我们并没有解释原因，只是给出了公式，然后解释了其合理性，现在我们给出了公式去证明这样取 loss 函数是合理的

交叉熵是信息理论里面的内容，这里不再具体展开，更多的内容，可以看到下面的[链接](#)