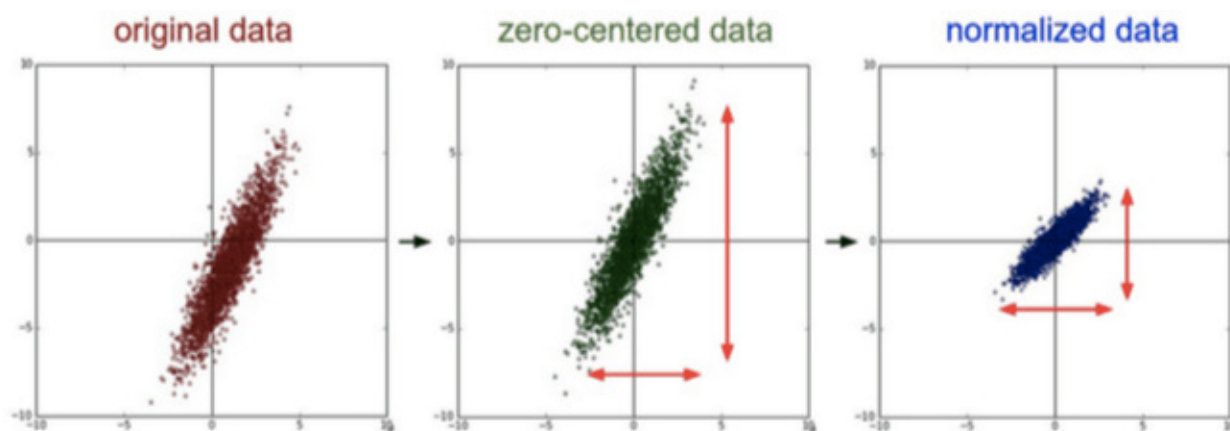


# 标准化

在我们正式进入模型的构建和训练之前，我们会先讲一讲数据预处理和批标准化，因为模型训练并不容易，特别是一些非常复杂的模型，并不能非常好的训练得到收敛的结果，所以对数据增加一些预处理，同时使用批标准化能够得到非常好的收敛结果，这也是卷积网络能够训练到非常深的层的一个重要原因。

## 数据预处理

目前数据预处理最常见的方法就是中心化和标准化，中心化相当于修正数据的中心位置，实现方法非常简单，就是在每个特征维度上减去对应的均值，最后得到 0 均值的特征。标准化也非常简单，在数据变成 0 均值之后，为了使得不同的特征维度有着相同的规模，可以除以标准差近似为一个标准正态分布，也可以依据最大值和最小值将其转化为  $-1 \sim 1$  之间，下面是一个简单的图示



这两种方法非常的常见，如果你还记得，前面我们在神经网络的部分就已经使用了这个方法实现了数据标准化，至于另外一些方法，比如 PCA 或者 白噪声已经用得非常少了。

## Batch Normalization

前面在数据预处理的时候，我们尽量输入特征不相关且满足一个标准的正态分布，这样模型的表现一般也较好。但是对于很深的网路结构，网路的非线性层会使得输出的结果变得相关，且不再满足一个标准的  $N(0, 1)$  的分布，甚至输出的中心已经发生了偏移，这对于模型的训练，特别是深层的模型训练非常的困难。

所以在 2015 年一篇论文提出了这个方法，批标准化，简而言之，就是对于每一层网络的输出，对其做一个归一化，使其服从标准的正态分布，这样后一层网络的输入也是一个标准的正态分布，所以能够比较好的进行训练，加快收敛速度。

batch normalization 的实现非常简单，对于给定的一个 batch 的数据  $B = \{x_1, x_2, \dots, x_m\}$  算法的公式如下

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (2)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3)$$

$$y_i = \gamma \hat{x}_i + \beta \quad (4)$$

第一行和第二行是计算出一个 batch 中数据的均值和方差，接着使用第三个公式对 batch 中的每个数据点做标准化， $\epsilon$  是为了计算稳定引入的一个小的常数，通常取  $10^{-5}$ ，最后利用权重修正得到最后的输出结果。