

神经网络常见术语之——Batch与Epoch

Batch

Batch大小是一个超参数，用于定义在更新内部模型参数之前要处理的样本数。将批处理视为循环迭代一个或多个样本并进行预测。在批处理结束时，将预测与预期输出变量进行比较，并计算误差。从该错误中，更新算法用于改进模型，例如沿误差梯度向下移动。训练数据集可以分为一个或多个Batch。当所有训练样本用于创建一个Batch时，学习算法称为批量梯度下降。当批量是一个样本的大小时，学习算法称为随机梯度下降。当批量大小超过一个样本且小于训练数据集的大小时，学习算法称为小批量梯度下降。

Epoch

Epoch数是一个超参数，它定义了学习算法在整个训练数据集中的工作次数。一个Epoch意味着训练数据集中的每个样本都有机会更新内部模型参数。Epoch由一个或多个Batch组成。例如，如上所述，具有一批的Epoch称为批量梯度下降学习算法。您可以将for循环放在每个需要遍历训练数据集的epoch上，在这个for循环中是另一个嵌套的for循环，它遍历每批样本，其中一个批次具有指定的“批量大小”样本数。

Epochs 数量传统上很大，通常是数百或数千，允许学习算法运行直到模型的误差被充分地最小化了。您可能会看到文献和教程设置为10,100,500,1000和更大的时期数量的示例。通常创建线图，其显示沿x轴的时间以及模型在y轴上的误差或技能。这些图有时被称为学习曲线。这些图可以帮助诊断模型是否已经过度学习，学习不足或者是否适合训练数据集。

Batch和Epoch有什么区别？

Batch大小是在更新模型之前处理的多个样本。Epoch数是通过训练数据集的完整传递次数。批处理的大小必须大于或等于1且小于或等于训练数据集中的样本数。可以将epoch设置为1和无穷大之间的整数值。您可以根据需要运行算法，甚至可以使用除固定数量的epoch之外的其他条件来停止算法，例如模型错误随时间的变化（或缺乏变化）。它们都是整数值，并且它们都是学习算法的超参数，例如学习过程的参数，而不是学习过程找到的内部模型参数。您必须为学习算法指定batch大小和epoch数。如何配置这些参数没有固定的规则。您必须尝试不同的值，看看哪种方法最适合您的问题。

工作实例

假设您有一个包含200个样本（数据行）的数据集，并且您选择的Batch大小为5和1,000个Epoch。

这意味着数据集将分为40个Batch，每个Batch有5个样本。每批五个样品后，模型权重将更新。

这也意味着一个epoch将涉及40个Batch或40个模型更新。

有1000个Epoch，模型将暴露或传递整个数据集1,000次。在整个培训过程中，总共有40,000Batch。