深度学习基本概念（补充中）

书写人：李宴军

数据集：可用于训练模型的数据的集合，其中的数据分为有标签和无标签两种，但不论是有标签的还是无标签的都能用于机器学习。以MINST为例，MINIST中的图像分为训练图像和测试图像，其中训练图像带有标签（标记图像代表哪个数字）用于训练模型，测试图像没有标签由于测试模型识别手写体的准确率。数据集有各种各样的，可以通过互联网得到。

模型：模型就是数据从输入到输出所使用的函数，放在神经网络中就是偏置、权重、激活函数以及神经网络的层数。

损失函数：用于衡量神经网络判断结果的“失败程度”的函数，既衡量模型有多差的函数。对损失函数求导后导数为负的反向就是损失函数减小的方向，沿着这个方向神经网络的性能将会提高。常见的有均方误差、交叉熵误差等。

感知机：一种算法，类似于神经网络，但和神经网络有个极大的不同。感知机的激活函数是阶跃函数，并不连续。感知机的运行模式比较简单，但是感知机的权重需要自己配，但神经元较多、连接较复杂时就用不了。

激活函数：神经网络的层数大于等于2时，就要涉及到将上一层的输出经过一定处理后再加权成为下一层的输入。而这个“一定处理”就是激活函数。另外，激活函数都是非线性的，这一点至关重要。常见的激活函数有Sigmoid、ReLU等。

全连接层：上一层的每一个神经元和下一层的每一个神经元都有连接。上一层神经元的输出通过和权重相乘后在经过激活函数的作用作为输入流向下一层。一般最后一层比较特殊要接softmax函数。（可能还能接其他函数）

优化器：实现参数更新的算法部分，调用时要用传入参数学习率和动量。比较常见的优化器有：SGD，AdaGrad，Ddam。

超参数：用以调整更新模型的权重，优化模型，是模型有学习功能的基础，如：学习率和动量。

学习率：学习率用来控制每次参数更新时参数沿着梯度为负的方向更新的幅度大小。学习率过大有可能会造成过拟合：而学习率过小，损失函数收敛的慢，浪费运行时间。在网上查到学习率位于0.001~0.01比较合适。

动量：一个超参数其作用是在一定程度上帮助优化器走出“局部最优点”寻找“全局最优解”。

能够有效的防止过拟合。

one-hot表示：正确标签表示为1，其余标签表示为0。例如：有一个带标签的数据数字2，

它用1表示，那么其它的带标签的数据都用0来表示。因为softmax函数最后会

返回一个概率，经过处理概率最大的输出改为1，其余改为0。那么数据就能和标签对应上，以测试输入的正确性。

均方误差：一种损失函数，其值等于监督数据和输出数据的差的平方求和后除以2。

交叉熵误差：一种损失函数，对于单个输出其值等于对正确解标签对应的索引的神经输出取负对数，实际使用时都是对全体训练数据使用，只要取算术平均数就可以。调用函数nn.CrossEntropyLoss（）可以实现。

批处理：为了提高处理数据的速度，节省运行时间可以将图片打包处理。

梯度：对于几个自变量和一个因变量，将因变量对于各个自变量的偏导形成一个向量，这个向量就被称为梯度，而负梯度的方向就是神经网络中要学习的参数的改变方向，沿着这个方向，损失函数的值将会进一步减小并收敛。

梯度法（梯度下降）：实际上梯度法有梯度上升法和梯度下降法两种，但其实不过是将损失函数变号而已。而且也没什么人用梯度下降法。

计算图：将数学计算式（计算过程）以图片的形式表现出来，主要有直线和节点构成，直线传递数据，节点表达运算关系。通过计算图可以清晰的看出计算的每个步骤。这也为反向传播带来了便利。

反向传播：沿着与正方向相反的方向计算，在节点处乘上偏导数即可。一个比较有意思的事情是这里的反向传播居然满足链式法则。

正则化：

过拟合：

特征图：卷积层的输出输入数据又叫特征图，分输出特征图和输入特征图。

卷积层：相比全连接层，

卷积运算：

填充：

步幅：

通道数：

池化层：

池化：