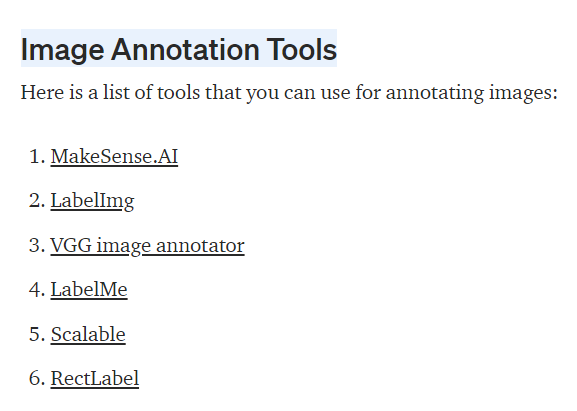
**Problem 1 (40 points) Suppose you are implementing a deep learning system for image classification task. Please describe how to build a deep learning system step-by-step. (a) (i) Describe how to annotate your dataset for image classification. What kind of tool will you use? (ii) How do you save each label as files? (iii) How would you split the dataset as training, validation, and test sets? Why? (15 points)**



Bounding boxes: Bounding boxes are the most commonly used type of annotation in computer vision. Bounding boxes are rectangular boxes used to define the location of the target object. They can be determined by the 𝑥 and 𝑦 axis coordinates in the upper-left corner and the 𝑥 and 𝑦 axis coordinates in the lower-right corner of the rectangle. Bounding boxes are generally used in object detection and localisation tasks.

1. Training sets are used to adjust the weights of the neural network during training;

2. Validation sets are not used to adjust weights, but to prevent overfitting.

3. The test sets are only used to test the model. To see how good the model is, it is to evaluate the generalization ability of the model.

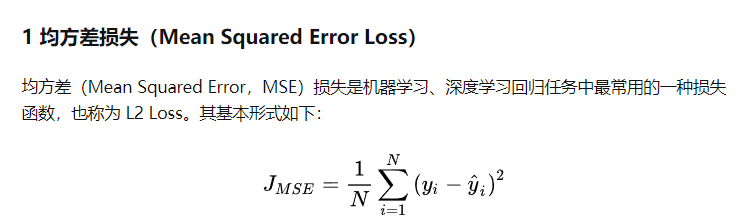
**(b) (i) Please choose a convolutional neural network (CNN) architecture (e.g., AlexNet, GoogLeNet, VGGNet, ResNet ....). Justify your answer why you chose the architecture. (ii) How can you be sure that your model is implemented correctly? Explain that in terms of loss. (10 points)**

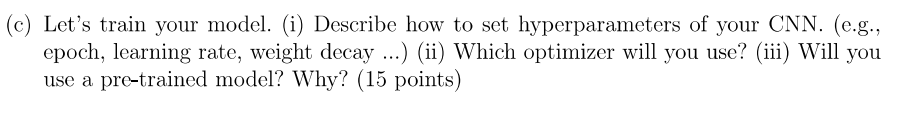
The advantages of AlexNet are:-Using the non-linear activation function ReLU

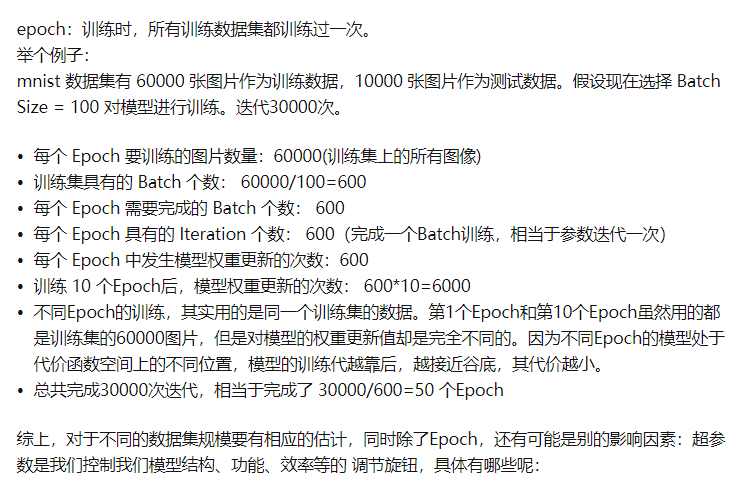
GoogleNet uses auxiliary classifiers (auxiliary classifiers), the output of a certain middle layer is used as a classification, and a small weight is added to the final classification result. This is equivalent to model fusion, and at the same time adds a back-propagation gradient signal to the network, and also provides additional regularization, which is of great benefit to the training of the entire network.

ResNet（残差网络）, deep residual network, prevents over-fitting：深度残差网络，防止过拟合

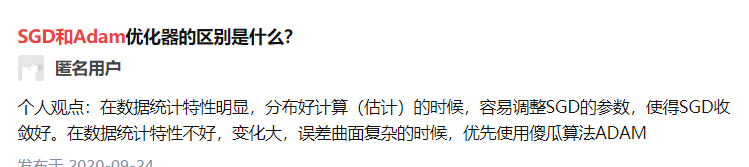
机器学习中的监督学习本质上是给定一系列训练样本 ，尝试学习 的映射关系，使得给定一个 ，即便这个 不在训练样本中，也能够输出 ，尽量与真实的 接近。损失函数是用来估量模型的输出 与真实值 之间的差距，给模型的优化指引方向。(Supervised learning in machine learning is essentially given a series of training samples, trying to learn the mapping relationship, so that given a, even if this is not in the training sample, it can still output, as close to the real as possible. The loss function is used to estimate the gap between the output of the model and the true value, and to guide the optimization of the model.)

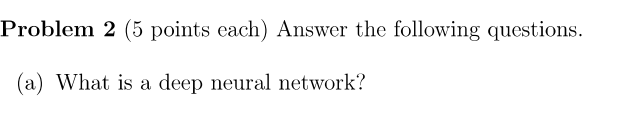




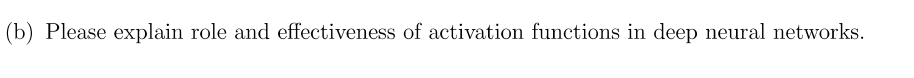


你的想法没错，预训练模型就是别人用同样的代码训练出来的模型，你当然可以自己训练，也许比下载下来的的预训练模型效果更好 ，但是很花时间，一般为了节省时间，我们都会选择基于预训练模型finetune





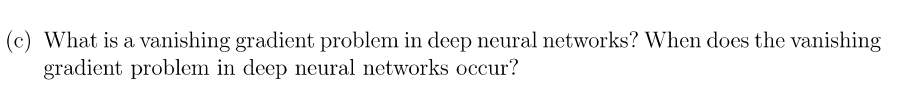
深度学习是机器学习的一种，而机器学习是实现人工智能的必经路径。深度学习的概念源于人工神经网络的研究，含多个隐藏层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。研究深度学习的动机在于建立模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本等。（Deep learning is a type of machine learning, and machine learning is the necessary path to realize artificial intelligence. The concept of deep learning originates from the research of artificial neural networks, and a multi-layer perceptron with multiple hidden layers is a deep learning structure. Deep learning forms a more abstract high-level representation attribute category or feature by combining low-level features to discover distributed feature representations of data. The motivation for studying deep learning is to build a neural network that simulates the human brain for analysis and learning. It mimics the mechanism of the human brain to interpret data, such as images, sounds, and texts.）



**激活函数的作用？**

**引入非线性因素。**

在我们面对线性可分的数据集的时候，简单的用线性分类器即可解决分类问题。但是现实生活中的数据往往不是线性可分的，面对这样的数据，一般有两个方法：引入非线性函数、线性变换。（When we are faced with a linearly separable data set, simply use a linear classifier to solve the classification problem. However, data in real life is often not linearly separable. In the face of such data, there are generally two methods: introducing nonlinear functions and linear transformations.）



在神经网络中，当前面隐藏层的学习速率低于后面隐藏层的学习速率，即随着隐藏层数目的增加，分类准确率反而下降了。这种现象叫做消失的梯度问题。（In a neural network, the learning speed of the hidden layer decreases, and now the learning speed of the hidden layer, followed by the increase of the hidden layer, decreases the speed of the gradual decrease.）

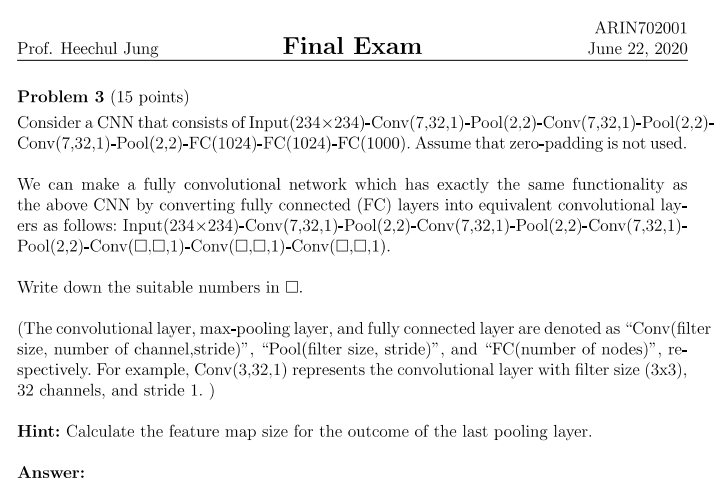
1.神经网络中大多时候都会选择softmax、tanh函数，而这两个函数的导数分别小于0.25和1，当神经网络的隐含层层增多的时候在求解loss函数关于参数的导数的时候，上一层函数关于下一层的导数的求导<1,多个求导的连乘,此时会发生梯度消失。

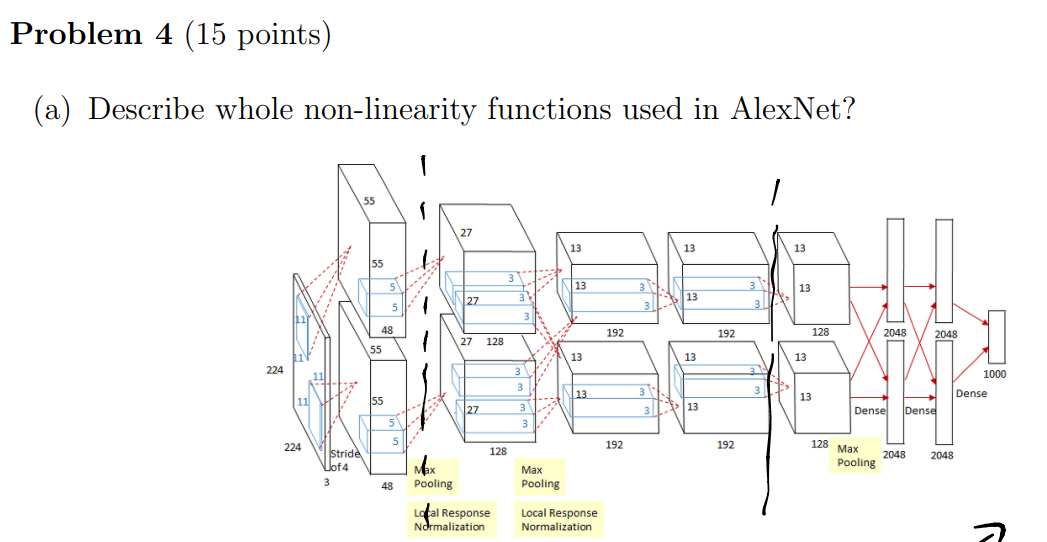
2.权重初始值的问题，在训练模型的时候，越靠近输出层的隐含层的权重变化比较大，越靠近输入层的隐含层的权重变化比较慢而且越接近初始化的权值，当初始化的权值很大的时候，在进行求导的时候会出现梯度消失

防止这种情况的发生主要是选择合适的激活函数，（1. The softmax and tanh functions are mostly selected in neural networks, and the derivatives of these two functions are less than 0.25 and 1, respectively. When the hidden layer of the neural network increases, when solving the derivative of the loss function with respect to the parameters, go up The derivation of the derivative of one layer of function with respect to the next layer is less than 1, and the continuous multiplication of multiple derivations will cause the gradient to disappear at this time.

2. The problem of the initial value of the weight. When training the model, the weight of the hidden layer closer to the output layer changes more, and the weight of the hidden layer closer to the input layer changes slowly and closer to the initialized weight. When the initial weight is very large, the gradient will disappear when deriving

To prevent this from happening is to choose a suitable activation function,）







GG 最大的特点就是它在之前的网络模型上，通过比较彻底地采用 3x3 尺寸的卷积核来堆叠神经网络，从而加深整个神经网络的层级。

VGG在AlexNet基础上做了改进，整个网络全部使用了同样大小的3\*3卷积核尺寸和2\*2最大池化尺寸，网络结果简洁。一系列VGG模型的结构图：

VGG论文给出了一个非常振奋人心的结论：卷积神经网络的深度增加和小卷积核的使用对网络的最终分类识别效果有很大的作用。记得在AlexNet论文中，也做了最后指出了网络深度的对最终的分类结果有很大的作用。这篇论文则更加直接的论证了这一结论。

