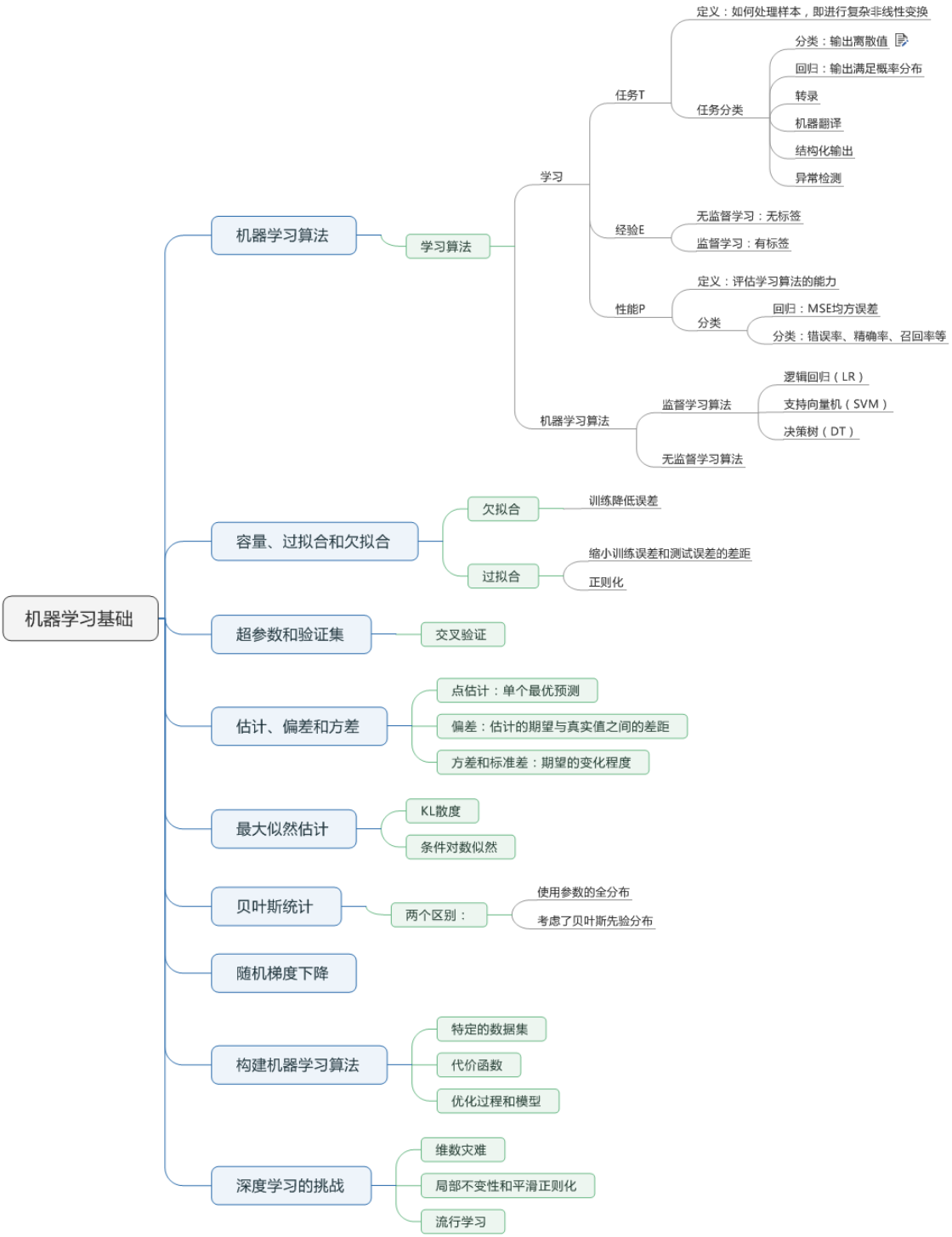
《深度学习》花书学习

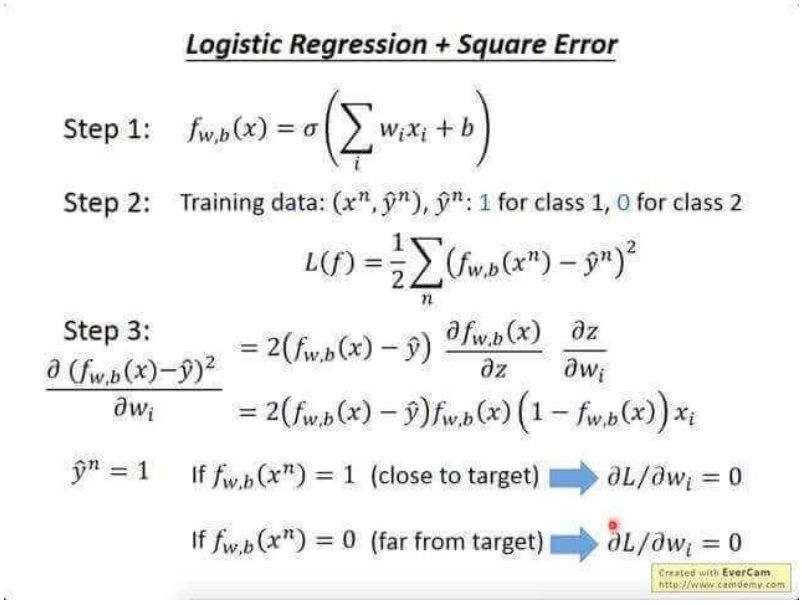
第一章 Week02 总结

1.1 重点知识



1.2 分析为什么平方损失函数不适用于分类问题

如果把平方损失函数用在逻辑回归上，那么就是下图这样的过程。



最后两行的意思说，如果真是标签是 1，你的预测值越接近 0，梯度越小。这样的目标函数显然是无法进行二元分类的。

1.3 论文阅读

深度学习可以让拥有多处理层的计算模型来学习具有多层次抽象的数据表示。这些方法在许多方面都带来了显著的改善，包括最先进的语音识别、视觉对象识别、对象检测和许多其它领域，例如药物发现和基因组学等。深度学习能够发现大数据中的复杂结构。它是利用 BP 算法来完成这个发现过程的。BP 算法能够指导机器如何从前一层获取误差而改变本层的内部参数，这些内部参数可以用于计算表示。深度卷积网络在处理图像、视频、语音和音频方面带来了突破，而递归网络在处理序列数据，比如文本和语音方面表现出了闪亮的一面。

深度学习就是一种特征学习方法，把原始数据通过一些简单的但是非线性的模型转变成为更高层次的，更加抽象的表达。通过足够多的转换的组合，非常复杂的函数也可以被学习。深度学习的核心方面是，上述各层的特征都不是利用人工工程来设计的，而是使用一种通用的学习过程从数据中学到的。

1、监督学习

机器学习中，不论是否是深层，最常见的形式是监督学习。在实际应用中，大部分从业者都使用一种称作随机梯度下降的算法（SGD）。它包含了提供一些输入向量样本，计算输出和误差，计算这些样本的平均梯度，然后相应的调整权值。通过提供小的样本集合来重复这个过程用以训练网络，直到目标函数停止增长。它被称为随机的是因为小的样本集对于全体样本的平均梯度来说会有噪声估计。这个简单过程通常会找到一组不错的权值，同其他精心设计的优化技术相比，它的速度让人惊奇。训练结束之后，系统会通过不同的数据样本——测试集来显示系统的性能。这用于测试机器的泛化能力——对于未训练过的新样本的识别能力。

2、反向传播来训练多层神经网络

反向传播算法的核心思想是：目标函数对于某层输入的导数（或者梯度）可以通过向后传播对该层输出（或者下一层输入）的导数求得。反向传播算法可以被重复的用于传播梯度通过多层神经网络的每一层：从该多层神经网络的最顶层的输出（也就是改网络产生预测的那一层）一直到该多层神经网络的最底层（也就是被接受外部输入的那一层），一旦这些关于（目标函数对）每层输入的导数求解完，我们就可以求解每一层上面的（目标函数对）权值的梯度了。

3、卷积神经网络

卷积神经网络使用 4 个关键的想法来利用自然信号的属性：局部连接、权值共享、池化以及多网络层的使用。

一个典型的卷积神经网络结构是由一系列的过程组成的。最初的几个阶段是由卷积层和池化层组成，卷积层的单元被组织在特征图中，在特征图中，每一个单元通过一组叫做滤波器的权值被连接到上一层的特征图的一个局部块，然后这个局部加权和被传给一个非线性函数，比如 ReLU。在一个特征图中的全部单元享用相同的过滤器，不同层的特征图使用不同的过滤器。使用这种结构处于两方面的原因。首先，在数组数据中，比如图像数据，一个值的附近的值经常是高度相关的，可以形成比较容易被探测到的有区分性的局部特征。其次，不同位置局部统计特征不太相关的，也就是说，在一个地方出现的某个特征，也可能出现在别的地方，所以不同位置的单元可以共享权值以及可以探测相同的样本。

卷积层的作用是探测上一层特征的局部连接，然而池化层的作用是在语义上把相似的特征合并起来，这是因为形成一个主题的特征的相对位置不太一样。一般地，池化单元计算特征图中的一个局部块的最大值，相邻的池化单元通过移动一行或者一列来从小块上读取数据，因为这样做就减少的表达的维度以及对数据的平移不变性。两三个这种的卷积、非线性变换以及池化被串起来，后面再加上一个更多卷积和全连接层。在卷积

神经网络上进行反向传播算法和在一般的深度网络上是一样的，可以让所有的在过滤器中的权值得到训练。

4、使用深度卷积网络进行图像理解

21 世纪开始，卷积神经网络就被成功的大量用于检测、分割、物体识别以及图像的各个领域。这些应用都是使用了大量的有标签的数据，比如交通信号识别，生物信息分割，面部探测，文本、行人以及自然图形中的人的身体部分的探测。近年来，卷积神经网络的一个重大成功应用是人脸识别。

尽管卷积神经网络应用的很成功，但是它被计算机视觉以及机器学习团队开始重视是在 2012 年的 ImageNet 竞赛。在该竞赛中，深度卷积神经网络被用在上百万张网络图片数据集，这个数据集包含了 1000 个不同的类。该结果达到了前所未有的好，几乎比当时最好的方法降低了一半的错误率。这个成功来自有效地利用了 GPU、ReLU、一个新的被称为 dropout 的正则技术，以及通过分解现有样本产生更多训练样本的技术。这个成功给计算机视觉带来一个革命。如今，卷积神经网络用于几乎全部的识别和探测任务中。最近一个更好的成果是，利用卷积神经网络结合回馈神经网络用来产生图像标题。

5、分布式特征表示与语言处理

深度学习理论表明深度网络具有两个不同的巨大的优势。这些优势来源于网络中各节点的权值，并取决于具有合理结构的底层生成数据的分布。首先，学习分布式特征表示能够泛化适应新学习到的特征值的组合（比如，n 元特征就有 2^n 种可能的组合）。其次，深度网络中组合表示层带来了另一个指数级的优势潜能（指数级的深度）。

层神经网络中的隐层利用网络中输入的数据进行特征学习，使之更加容易预测目标输出。下面是一个很好的示范例子，比如将本地文本的内容作为输入，训练多层神经网络来预测句子中下一个单词。内容中的每个单词表示为网络中的 N 分之一的向量，也就是说，每个组成部分中有一个值为 1 其余的全为 0。在第一层中，每个单词创建不同的激活状态，或单词向量。在语言模型中，网络中其余层学习并转化输入的单词向量为输出单词向量来预测句子中下一个单词，可以通过预测词汇表中的单词作为文本句子中下一个单词出现的概率。网络学习了包含许多激活节点的、并且可以解释为词的独立特征的单词向量，正如第一次示范的文本学习分层表征文字符号的例子。这些语义特征在输入中并没有明确的表征。而是在利用“微规则”（‘micro-rules’,本文中直译为：微规则）学习过程中被发掘，并作为一个分解输入与输出符号之间关系结构的好的方式。当句子是来自大量的真实文本并且个别的微规则不可靠的情况下，学习单词向量也一样能表现得很好。利用训练好的模型预测新的事例时，一些概念比较相似的词容易混淆，比如星期二（Tuesday）和星期三（Wednesday），瑞典（Sweden）和挪威（Norway）。这样的

表示方式被称为分布式特征表示，因为他们的元素之间并不互相排斥，并且他们的构造信息对应于观测到的数据的变化。这些单词向量是通过学习得到的特征构造的，这些特征不是由专家决定的，而是由神经网络自动发掘的。从文本中学习得单词向量表示现在广泛应用于自然语言中。

6、递归神经网络

由于先进的架构和训练方式，RNNs 被发现可以很好的预测文本中下一个字符或者句子中下一个单词，并且可以应用于更加复杂的任务。例如在某时刻阅读英语句子中的单词后，将会训练一个英语的“编码器”网络，使得隐式单元的最终状态向量能够很好地表征句子所要表达的意思或思想。这种“思想向量”（thought vector）可以作为联合训练一个法语“编码器”网络的初始化隐式状态（或者额外的输入），其输出为法语翻译首单词的概率分布。如果从分布中选择一个特殊的首单词作为编码网络的输入，将会输出翻译的句子中第二个单词的概率分布，并直到停止选择为止。总体而言，这一过程是根据英语句子的概率分布而产生的法语词汇序列。这种简单的机器翻译方法的表现甚至可以和最先进的（state-of-the-art）的方法相媲美，同时也引起了人们对于理解句子是否需要像使用推理规则操作内部符号表示质疑。这与日常推理中同时涉及到根据合理结论类推的观点是匹配的。