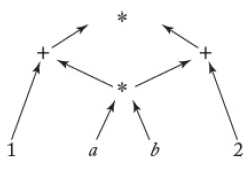
1. 训练神经网络
2. 计算图抽象

计算图抽象能让我们方便的构建任意的网络结构，计算输出并对任意标量损失函数的参数求导。

计算图是任意数学计算的图的表示。计算图为有向无环图DAG，图中的节点表示数学运算或者变量，边表示节点之间的中间值。计算公式(a\*b+1)\*(a\*b+2)的计算图如下：



神经网络可以

1. 文本数据的特征

本章介绍了处理文本数据时可用特征的几个来源。第7章讨论了一些具体NLP问题上选择的特征。第8章介绍了将特征转化成可以输入神经网络的输入变量

1. NLP分类问题

根据分类对象的不同，自然语言的分类问题可分为几大类。

1. Word

这类问题中，我们将对单个词语，比如dog、magnificent等完成一些任务，这是一个有生命的物体么？这个词是什么语言中的？词拼写正确么等等。这类任务比较少见，因为单个词很少独自出现，大多数情况下，词的含义和其使用的上下文有关。

1. Texts

在这类问题中，我们可能面对一段文字，一个短语、一个句子、一个段落或者一个文档，需要解决的问题类似于这是否是垃圾邮件？这是关于政治的还是体育的？作者是谁？等。这类问题比较常见，我们将这类问题统称为文档分类问题。

1. Paired texts

对于给出的一对单词或者长文本，解决单词A和B是否近义词？文档A和B是否为同一作者？从句子B能推出句子A么？这类问题

1. Word in context

对于一段文字中的一个特定词语，我们进行分类。比如，在I want to book a flight中的book是一个名词还是一个动词或者形容词？词apple是一种水果还是一个公司？在I read a book on london中on是一个正确的介词么？这个词是一个人名、地名还是组织名？这些问题经常包含于一个大的任务中，比如将文档分成句子，找到文档中的命名实体，找到所有包含给定实体的文档。

1. Relation between two words

对于一个大文档上下文中的两个词语或短语，他们之间的关系是什么？在一个句子中词A是动词B的主语么？等等

这些分类问题可以扩展到我们感兴趣的相关分类决策的结构化问题，一个问题的决策影响其他问题的决策。

1. NLP问题的特征

下面我们讨论一些上面问题的通用特征。由于单词和字母都是离散的，我们的特征通常采用指示或者数量表示。指示的特征的值为0或1，数量是某个事物出现的次数。

1. 直接观察的属性
2. 单个单词的特征

在上下文之外观察一个单词，信息的主要来源包括构成单词的字母和字母顺序，单词的长度，单词的拼写（首字母是否大写？所有字母是否大写？是否包含连字符？是否包含数字？等等），词的前缀和后缀是什么？

我们还关注这个词的外部信息：这个词在一个大文本集合中出现的次数，这个词是否在美国常用人名中出现？

1. 文本的特征

当我们考虑一个句子、一个段落或者一个文档，可观察的特征就是字母的数量和顺序以及文本中的单词。

Bag of words

句子和文档的一个常用特征提取方法是bag of words BOW方法。在这个方法中，我们观察文本中单词的柱状图，将每一个单词数量看作一个特征。

我们还可以单词和字母的数量，比如一个句子包含字母的数量或单词的数量。

Weighting

当使用bow方法时，经常使用TF-IDF权重。除了单词之外，我们还可以考虑连续的单词对或者三个单词，这叫做ngrams。

1. 上下文中的单词特征

当考虑一个句子或文档中的单词时，单词的直接观察特征是单词在句子中的位置，还有在它前后的单词。离目标单词越近的单词能提供更多信息。

Windows

通常用以目标此为中心的窗口作为上下文，比如目标词前后k个单词，k通常为2、5或10。比如，考虑句子the brown fox jumped over the lazy dog，目标词为jumped。K取2时，将产生特征集合{word=brown, word=fox, word=over, word=the}。窗口方法是bow方法的一个版本，但是限定在窗口内。

窗口大小固定在bow模型基础上，进一步考虑了单词的相对位置。比如，上面的例子将产生特征集合{word-2 = brown, word-1 = fox, word+1 = over, word+2 = the}。

基于窗口的特征转化为向量将在第8章讨论，第14和16章我们将介绍biRNN的结构，biRNN实现了灵活的、可调整的、可训练的窗口。

Position

除了单词的上下文，我们或许还关系单子在句子中的绝对位置。比如目标单词是句子的第五个单词，目标单词出现在前10个单词中等等。

1. 单词之间的关系特征

当考虑上下文中的两个单词时，除了每个单词的位置和窗口，还可以看两个单词之间的距离，以及他们之间出现的单词的特征。

1. 推断的语言特征
2. 核心特征和联合特征
3. Ngram特征
4. 分布特征