**深度学习笔记**

**Task0**

1. **线性回归**
2. 线性回归根据输入数据预测输出，输入与输出呈线性关系
3. 深度模型的基本要素包括：

数据集（训练集、验证集、测试集）

损失函数

优化函数

训练模型

深度模型训练的基本步骤：

读取数据

设置超参数

初始化模型参数

前向传播计算输出

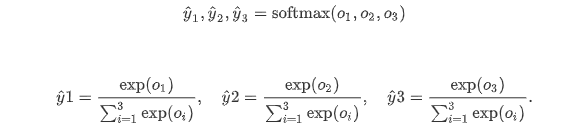
计算loss

反向传播计算梯度

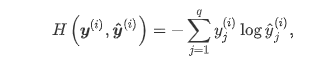
优化器更新参数

1. **Softmax与分类模型**

1、softmax计算



2、交叉熵计算

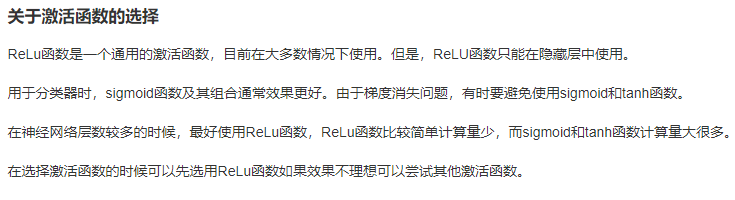


1. **多层感知机**

1、多层感知机为至少含有一层隐藏层的多层网络，如果全连接层之间不加激活函数层的话，则多个线性层叠加与只有一层的表达能力相同。激活函数对线性层的输出进行非线性化

2、各种激活函数

Relu、sigmoid、tanh

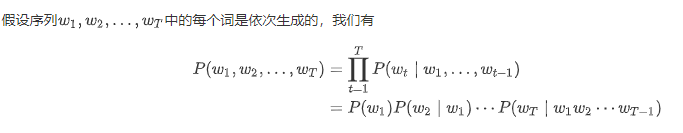


**Task1**

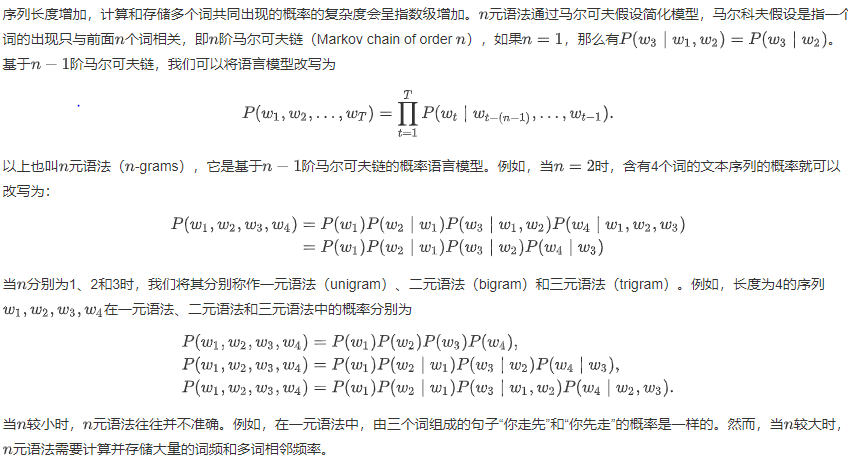
1. **文本预处理**

文本预处理基本步骤：

1. 读入文本
2. 进行特定级别的分词
3. 建立字典，为每个词映射到一个唯一的索引
4. 将文本从词的序列转为索引的序列，方便模型输出
5. **语言模型**
6. 语言模型的目的是为了验证一段词的序列是否合理



1. n元语法



n元语法存在的问题：

参数空间过大；

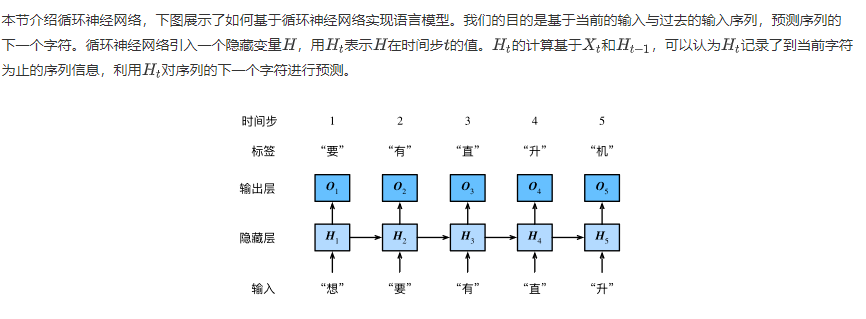
数据稀疏，因为很多词的词频太小导致出现的概率太小，很多概率接近0

1. 时序数据采样

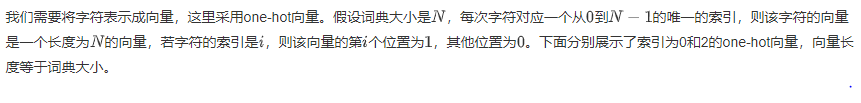
随机采样

相邻采样

1. **循环神经网络基础**
   1. 概念

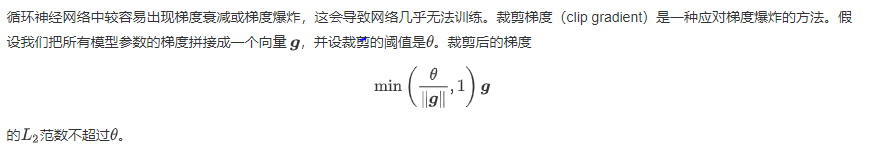


2、one-hot向量

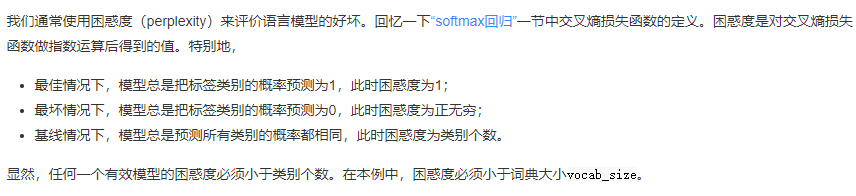


3、梯度裁剪

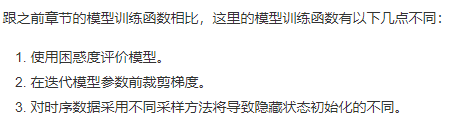
由于循环神经网络容易出现梯度消失和梯度爆炸的问题，为应对梯度爆炸的问题，采用梯度裁剪的方法



4、困惑度



5、模型训练



对于随机采样，要在每个batch开始时对隐藏层参数初始化，因为相邻的两个batch的数据不是相邻的，所以前一个batch的隐藏状态不能用于下一个batch

对于相邻采样，由于相邻的两个batch的数据也是相邻的，因此只需要在每个epoch开始时对隐藏层参数进行初始化即可