第1讲 分词

1、简述快速词典的作用？

2、简述最大正向匹配分词的工作过程。

3、简述最少分词的工作过程。

4、词网格分词与N-gram的关系。

5、为什么N-gram中大多使用Bigram和Trigram模型

6、简述Bigram分词的工作过程。

7、简述Bigram分词中Viterbi求解算法。

8、会给定词网格，利用Bigram模型进行计算各个路径评价值，并选择最佳路径。。

第2讲 词性标注

1、简述HMM进行词性标注的主要过程

2、会给定网格，利用HMM模型计算各个路径评价值，并选择最佳路径。

第3讲 命名实体识别

1、中文命名实体常见的类型包括哪些？

2、时间、日期、数值等这类仿词常使用哪种识别方法。

3、假设将商品名视作一种命名实体，设计一个用于识别商品名的命名实体系统，举例说明该系统的工作过程和具体工作算法。例如：

第4讲 文本分类

1、学会词向量空间表示、布尔向量空间表示、词频向量空间表示、TF-IDF向量空间表示。

2、学会TF-IDF手工计算。例如给定几篇文档，计算其中某个词的TF-IDF值。

3、学会使用Naïve Bayes进行文本分类。

4、简述在Naïve Bayes中常用的数据平滑方法工作过程？

第5讲 文本聚类

1、学会k-Means文本聚类的工作过程。

2、学会层次文本聚类的工作过程。

3、如何使用向量空间模型进行两个词义相似度计算。

第7讲 垃圾邮件和情感分析

1、垃圾邮件过滤的常用特征，并设计一个基于分类模型的垃圾邮件过滤模型，指出工作过程。

2、如何在有监督数据集下度量一个情感词的褒贬程度。

3、如何在无监督数据集下度量一个情感词的褒贬程度。

4、基于加权的语句级情感分析工作过程。

5、基于贝叶斯的情感分析工作过程。

6、设计一种构建情感词典的技术方案。

7、舆情分析方面当前常见的研究问题有哪些？

第8讲 个性化推荐

1、学会采用改进余弦相似度的基于项协同过滤计算。

2、什么是通用推荐、什么是个性化推荐。

3、协同过滤中的显式评分和隐式评分获取方法。

第9讲 组合推荐方法

1、用户模型通常包括的内容有哪些？

2、基于内容推荐的工作原理。

3、组合推荐的三种策略（整体式混合策略、并行式混合策略、流水线式混合策略）的工作原理。