一、选择题（共30分，多选10分每题，单选5分每题）

1. 简而言之，数据（有效的）就是（）。

A 数字 B 信息 C 记录 D 符号

1. 在二维散点图中，进行聚类时，之所以可以形成相对确定的分类是因为（）。

A 数据量足够大 B 数据足够分散 C 数据信息足够强 D 数据离散度大

1. 大数据操作主要的两个基本流程是（）。（多选）

A Map B Distribute C Reduce D Collect

1. 机器学习过程主要包括哪几个主要步骤（）。（多选）

A 数据采集 B 数据清洗处理 C 训练应用 D 数据库创建

1. 主观题（70分）
2. 在机器学习过程中，针对 对应产品质量未标记 数据，相对 已标记数据 有哪些特点？（30分）
3. 根据自己对大数据机器学习的学习了解，谈谈该部分如何和自己擅长/熟悉领域的嵌入应用？（40分）