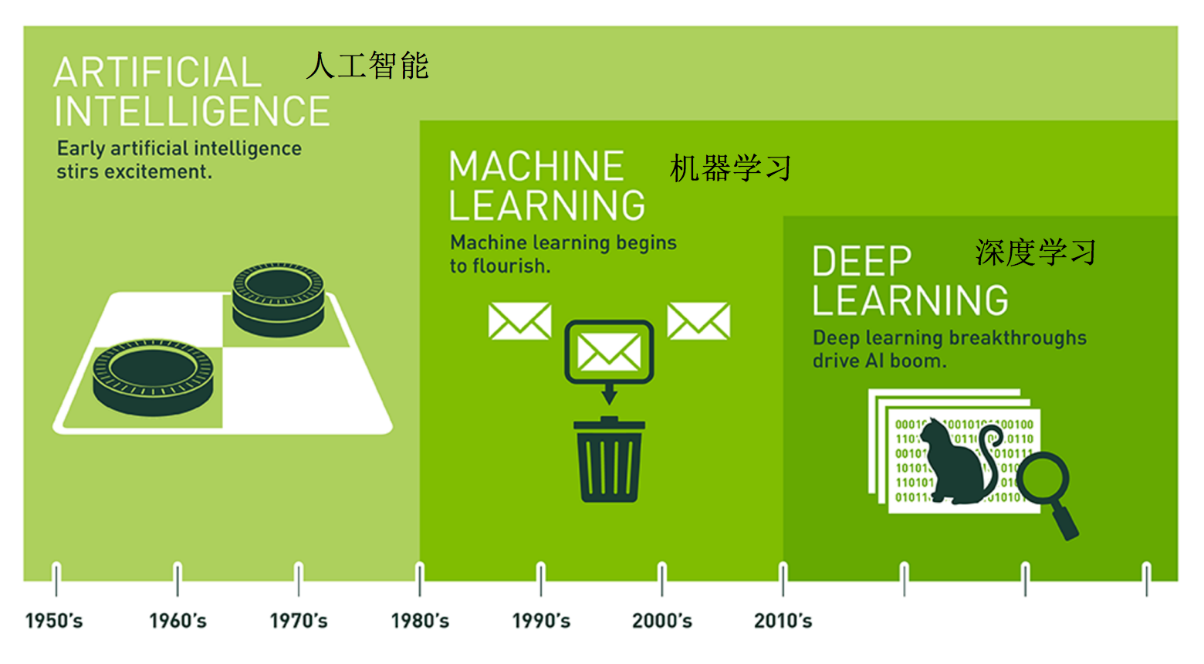
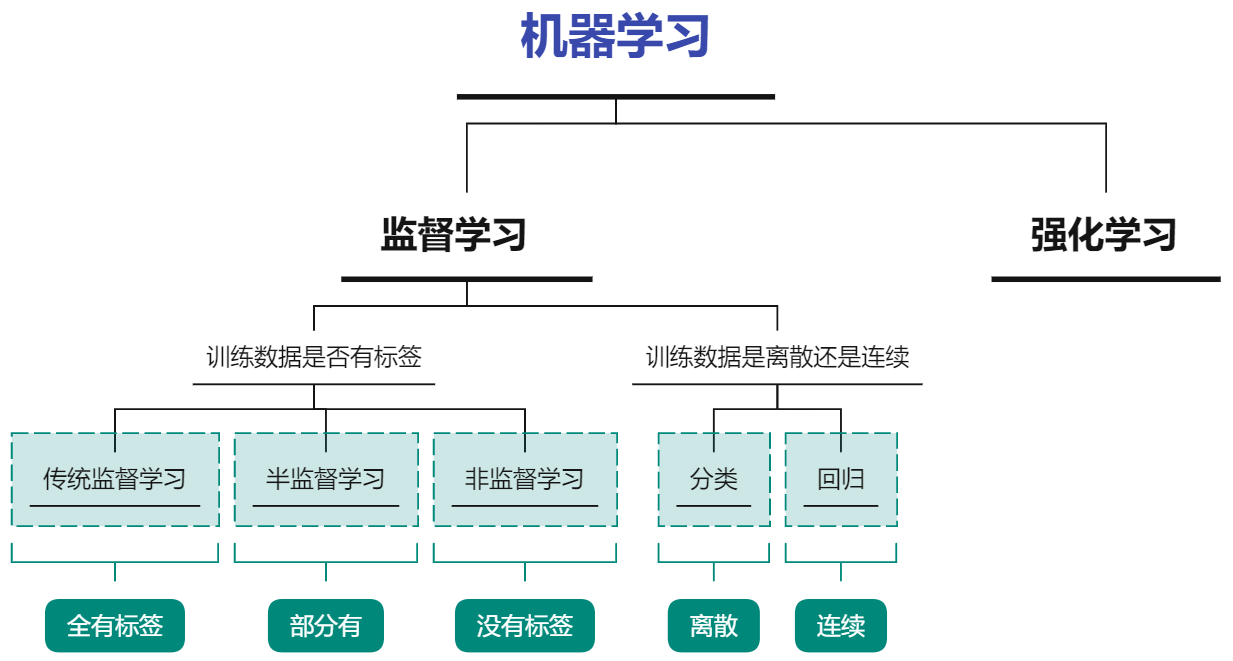
**课程学习感悟**

柳成林-20206722

在学习这门课之前，我对机器学习与深度学习的了解仅在于听说过的层面上。从名字上看，机器学习与深度学习都是感觉非常高大上的名词，所以我总是不敢去接触这些领域。



在课程上，我了解到，机器学习与深度学习都是人工智能的一部分，并且深度学习是机器学习的一个分支领域。想要学习深度学习，就要从机器学习开始。在最初的几次课程中，老师为我们介绍了机器学习的几个基本算法。包括回归、决策树、支持向量机和随机森林等等，我了解到了机器学习的整体体系。



机器学习可以分为监督学习和强化学习，由人工采集数据以及相应标签输入计算机的机器学习方法称为监督学习；即通过计算机与环境的互动逐渐强化自己的行为模式达到一个最佳的效果，称之为强化学习。

对监督学习根据训练样本是否有标签我们将其分为三类：

1.训练数据全部**都有对应标签**——传统监督学习

2.训练数据全部**都没有标签**——非监督学习

3.训练数据**一部分有标签**，一部分没有标签——半监督学习

另外，监督学习还有另一种分类方法，根据标签是否是离散的可以分为两类：一类是训练数据对应标签，全部是离散的值，我们称为**分类**；另一类是其训练数据对应标签为连续的值，我们称为**回归**。

在刚开始学习的时候，有一个让我影响深刻的点。因为我之前有参加过数学建模比赛，在数学建模的方法中也有一个叫回归的模型（即统计学中的回归），他和机器学习中的回归又有什么不同呢？

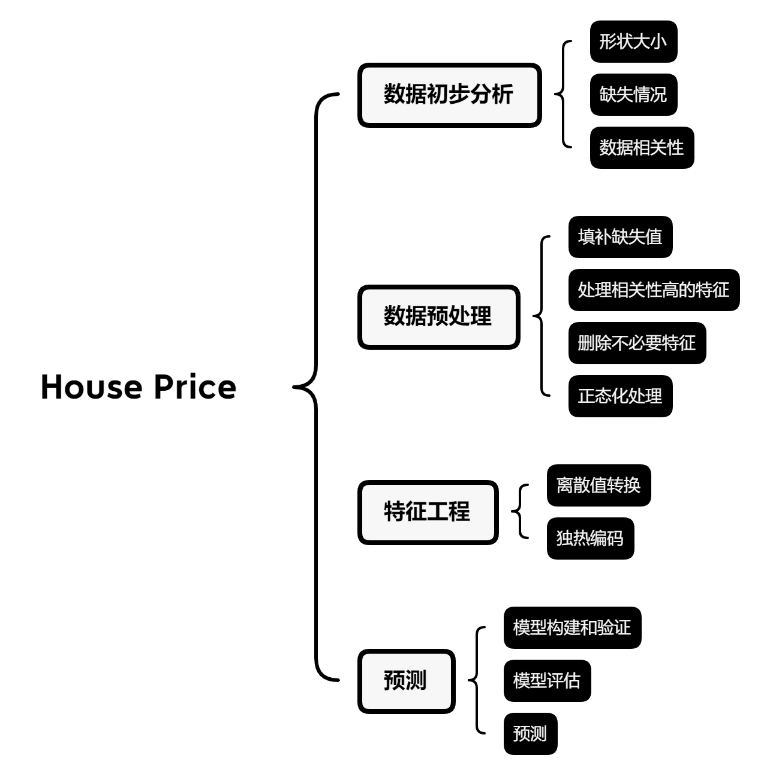
统计学中的回归分析主要强调的是解释变量与被解释变量之间的相关关系，并在假设检验的基础上进行参数估计，往往重视这个方程是否显著，各个参数是否显著。机器学习中的回归有点不同，把解释变量看成是一种输入，被解释变量视为一种输出，更加强调找到最优函数，没有假设，而是用训练数据来提取信息，再由测试数据验证结果，哪个模型的平均残差小，就选择哪个模型。

**机器学习实战**

在了解了机器学习的大体框架之后，我使用一些机器学习模型进行了练习。我选取了Kaggle上的房价预测题目进行训练。

题目背景：通过数据集中爱荷华州埃姆斯市住宅的方方面面79个解释变量，综合分析房屋的各类特征对房屋价格的影响，并对test数据集中房屋的最终价格进行预测。

我所做的工作如下：



在进行数据预处理后，我选取了一些基本的回归模型对房价进行预测，其结果如下：

Lasso模型回归评价:



Elastic Net Regression模型回归评价:



Kernel Ridge Regression模型回归评价:



Gradient Boosting Regression模型回归评价:



Ridge模型：

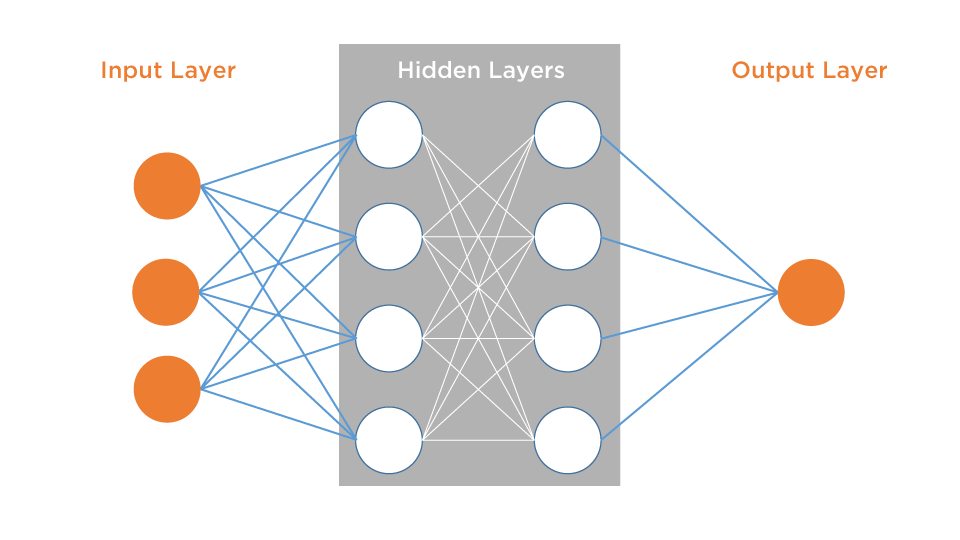


之后选取Lasso、ENR、GBR三种模型做了一个加权平均,使用加权平均模型进行了回归分析。得到的结果在官网排名200/6000，说明模型还算不错。

**深度学习学习过程与实战**

在课堂上，我了解了一些深度学习的内容，例如各种神经网络方法，图像识别方法。

深度学习的各种方法都是在人工神经网络ANN上进行演化，除了输入层和输出层外，在深度ANN的可见层之间还存在成百上千个用于输入和输出的“隐藏层”。每个隐藏层的结果矩阵即是下一层的输入矩阵。因此，只由最后一层的输出矩阵来提供结果。



在实战中，我选择了气温预测，因为在之前的数学建模中有遇到过类似的场景。我采用CNN与LSTM结合进行多特征的时间序列预测，LSTM即长短期记忆网络，基于RNN进行改进，可以学习长期依赖信息。在本项目中通过CNN提取特征，由于CNN权值共享的特性，能降低模型复杂度，降低计算量。具体内容详见PPT。

