# 人工智能实验指导书

**一、实验内容和要求**

本课程有16学时的实验课，由4次课内实验课构成。

其主要内容和应达到的要求是：

第一个实验4学时：从学习的任务（Tasks）、范式（Paradigms）与模型（Models）这三个角度理解机器学习的一般方法。编程实现线性回归，逻辑斯蒂回归算法。

第二个实验4学时：理解支持向量机与核方法的基本设计思想及其与其他有监督学习算法的联系与区别，掌握几种基本的核函数实现方法。

**二、实验详情**

实验一：有监督学习

课时分配：4学时 教学方法：上机实验

教学要求：熟悉Jupyter Notebook：Web机器学习应用； 掌握Python 编程语言；编程实现线性回归，逻辑斯蒂回归，线性判别分析等算法。

教学内容：

1. 安装使用Jupyter Notebook 或者 Pycharm。
2. 编程实现线性回归任务。数据集：diabete数据集。根据特征值‘bmi’对糖尿病情作出预测。
3. 编程实现二元或多元逻辑斯蒂回归任务。数据集：Iris数据集（安德森鸢尾花卉数据集）。Iris 数据集是用来给鸢尾花做分类的数据集，每个样本包含了花萼长度、花萼宽度、花瓣长度以及花瓣宽度四个特征。学生需要构建一个逻辑斯蒂回归模型，通过一个样本中的一个或者多个特征，来判断该样本属于Setosa（山鸢尾）、Versicolour（变色鸢尾）还是Virginica（维吉尼亚鸢尾） 。

实验二：支持向量机与核方法

课时分配：4学时 教学方法：上机实验

教学要求：理解支持向量机与核方法的基本设计思想及其与其他有监督学习算法的联系与区别，掌握几种基本的核函数实现方法。

教学内容：

1．生成数据集。使用sklearn.datasets.make\_blobs函数为聚类任务生成数据集，输出数据集和相应的标签。特征数设置为2 (n\_features)，标签数设置为2 (centers);划分数据集，80%为训练集、其余为测试集。

2．数据可视化。(参考使用matplotlib)。以第一个样本特征为x轴，第二个样本特征为y轴，绘制散点图。(根据标签着色)。

3．搭建模型，svm.LinearSVC。

4．train and test。训练集注入模型，随后将训练好的模型用于测试集预测。测试结果衡量指标(metrics)：accuacy。

5．分析讨论。调整模型参数C，对不同结果进行分析。

6．数据集：iris。特征采用,数据集前两个特征值;划分数据集，前130为训练集，其余为测试集。

7．搭建模型。svm.SVC(kernel=‘ rbf ’)，选择核函数。

8．train and test。训练集注入模型，随后将训练好的模型用于测试集预测。测试结果衡量指标(metrics)：accuacy，precision and recall。

9．可视化。绘制支持向量机分类边界。

实验三：无监督聚类算法

课时分配：4学时 教学方法：上机实验

教学要求：理解无监督聚类算法的基本设计思想及其与有监督学习的联系与区别。掌握几种基本的聚类算法框架，包括dbscan和K-均值法。

教学内容：

1．加载数据集：Data\_for\_Cluster.npz，X为特征，labels\_true为标签。

2．搭建模型，k-means与dbscan。

3．训练模型，调参，得出分类结果。

4．结果分析及可视化。绘制散点图。(根据分类结果进行着色) ；从算法原理的角度分析两个算法优缺点，及适应的数据集特征。

5．评估标准，轮廓系数法（Silhouette Cofficient），用来评估聚类算法的效果。

实验四：手写数字识别——CNN的应用

课时分配：4学时 教学方法：上机实验

教学要求：理解神经网络的构造原理与求解方法，理解主流的CNN网络。

教学内容：

1．加载数据集：

(x\_train, y\_train), (x\_test, y\_test) = tf.keras.datasets.mnist.load\_data()

2．搭建CNN模型。优化器（optimizer）：adam；评估指标（metrics）：accuacy。

3．训练好模型后，将其用于测试。