

人工智能基础第二章实验

王斐 Michael

SenseTime Edu

1 前言

这一章是人工智能基础的入门章节，即通过鸢尾花案例来介绍简单的感知器 (perceptron) 线性二元分类，并且引入支持向量机 (support vector machine) 模型的学习。在本章节的实验中，我们会通过不同的案例来，来学习和训练以下相关知识：

- 熟悉数据阵
- 向量运算和其几何性质
- 线性分类
- 梯度下降法 (Gradient Descent)
- 感知器线性二元线性分类
- 支持向量机 (support vector machine) 二元线性线性分类
- 支持向量机 (support vector machine) 二元非线性分类 (选修)

2 实验重点

同学们在进行本章的实验练习时，要留意下面几个要点：

1. 摆脱计算思维，所有计算能扔给计算机的就扔给计算机
2. 理解数学概念后，善于用计算机来直观得呈现相关概念
3. 培养对数据阵的直观感受训练自己对机器学习模型的流程化理解，即
 - 数据准备和可视化
 - 模型搭建和调用
 - 模型运行寻找规律参数
 - 模型预测和评估

提示：只要你会解一元二次方程组，那么你就可以顺利完成该实验，所以请耐心等待 (可以分三次，每次 30 分钟完成全部实验)

代码不可复制!

本实验内容所有代码需要同学们自己**输入**，因为如果可以复制的话，你可能不会自己输入，也不会养成良好的编程习惯。

3 实验 1: 数据阵、向量和及其性质

我们在课程中,学习了数据阵的定义,并且给出了下面的图形演示。

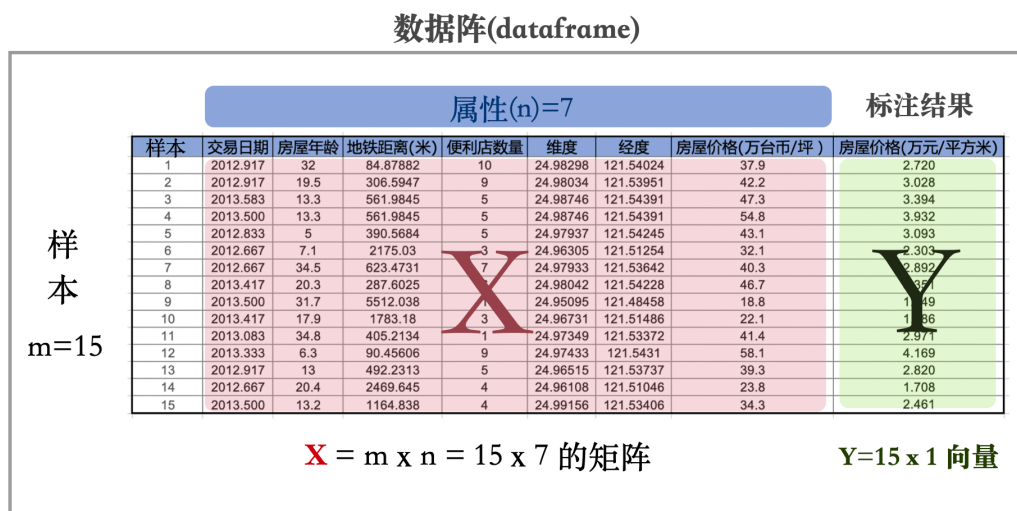


Figure 1: 数据阵可视化

下面我们就来学习 Python 语言环境下的数据阵操作。按照惯例我们首先加载具体的工具包。有些工具包是辅助我们实验的,有些被标记重点内容的,需要同学们掌握。

```
# 加载常用工具包, SenseStudy 目前测试可以加载下列工具包
# 注意, 非重点内容不需要掌握
import numpy as np #重点内容
import pandas as pd #重点内容
import matplotlib.pyplot as plt #重点内容
import seaborn as sns
from sklearn import datasets
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Perceptron
```

3.1 鸢尾花数据阵构造和读取

```
iris = datasets.load_iris() # 加载数据阵
print(iris.data[0: 5])
print(iris.target[0:5])
print(iris.feature_names)
iris.feature_names.append('Category')
iris_df = pd.DataFrame(np.column_stack((iris.data, iris.target)),
                      columns=iris.feature_names)
iris_df.Category = iris_df.Category.astype(int)
print(iris_df.shape) # 150 x 5 dataframe, 特征变量矩阵X 为 150 x 4, 即150 个样本,
                    # 4个属性, 还有1个结果向量
```

3.2 鸢尾花数据阵可视化

```
sns.set(style="whitegrid")
fig, axes = plt.subplots(1, 2, figsize=(12, 6))
sns.scatterplot(x=iris_df.iloc[:, 0],
```

```

        y=iris_df.iloc[:, 1],
        hue=iris_df.iloc[:, 4],
        ax=axes[0], s=80, palette="Set1")
sns.scatterplot(x=iris_df.iloc[:, 2],
                y=iris_df.iloc[:, 3],
                hue=iris_df.iloc[:, 4],
                ax=axes[1], s=80, palette="Set1")

```

3.3 鸢尾花种类选取

现在我们选取两种鸢尾花，并且只对其花瓣的长度和宽度进行可视化。

```

iris_df_binary = iris_df[iris_df.Category != 1] # 选取种类标记为 0 和 2 的鸢尾花
print(iris_df_binary.shape) # 查看新的维数
print(iris_df_binary.columns) # 查看属性名称
sns.scatterplot(x=iris_df_binary.iloc[:, 2],
                y=iris_df_binary.iloc[:, 3],
                hue=iris_df_binary.iloc[:, 4],
                s=80, palette="Set1") # 只对两种鸢尾花进行可视化

```