



東莞理工學院  
DONGGUAN UNIVERSITY OF TECHNOLOGY

# 人工智能概论

## 实验一：回归模型

丁烨，计算机科学与技术学院

[dingye@dgut.edu.cn](mailto:dingye@dgut.edu.cn)



# 准备实验环境



<https://websitesetup.org/wp-content/uploads/2020/04/Python-Cheat-Sheet.pdf>  
[https://perso.limsi.fr/pointal/\\_media/python:cours:mementopython3-english.pdf](https://perso.limsi.fr/pointal/_media/python:cours:mementopython3-english.pdf)

# 准备实验环境

- ❖ scikit-learn
- ❖ <https://scikit-learn.org/>
- ❖ 一个开源的科学计算及机器学习工具包
- ❖ 属于 SciPy 项目的一部分
- ❖ 包含了常见的、基础的机器学习算法
- ❖ 不支持深度学习
- ❖ 较难支持 GPU 加速



# 准备实验环境



- ❖ NumPy
- ❖ <https://www.numpy.org/>
- ❖ 针对数组运算提供大量的数学函数库
- ❖ 支持大规模的多维数组与矩阵运算
- ❖ NumPy 是 SciPy、Matplotlib 等扩展程序库的基础组件

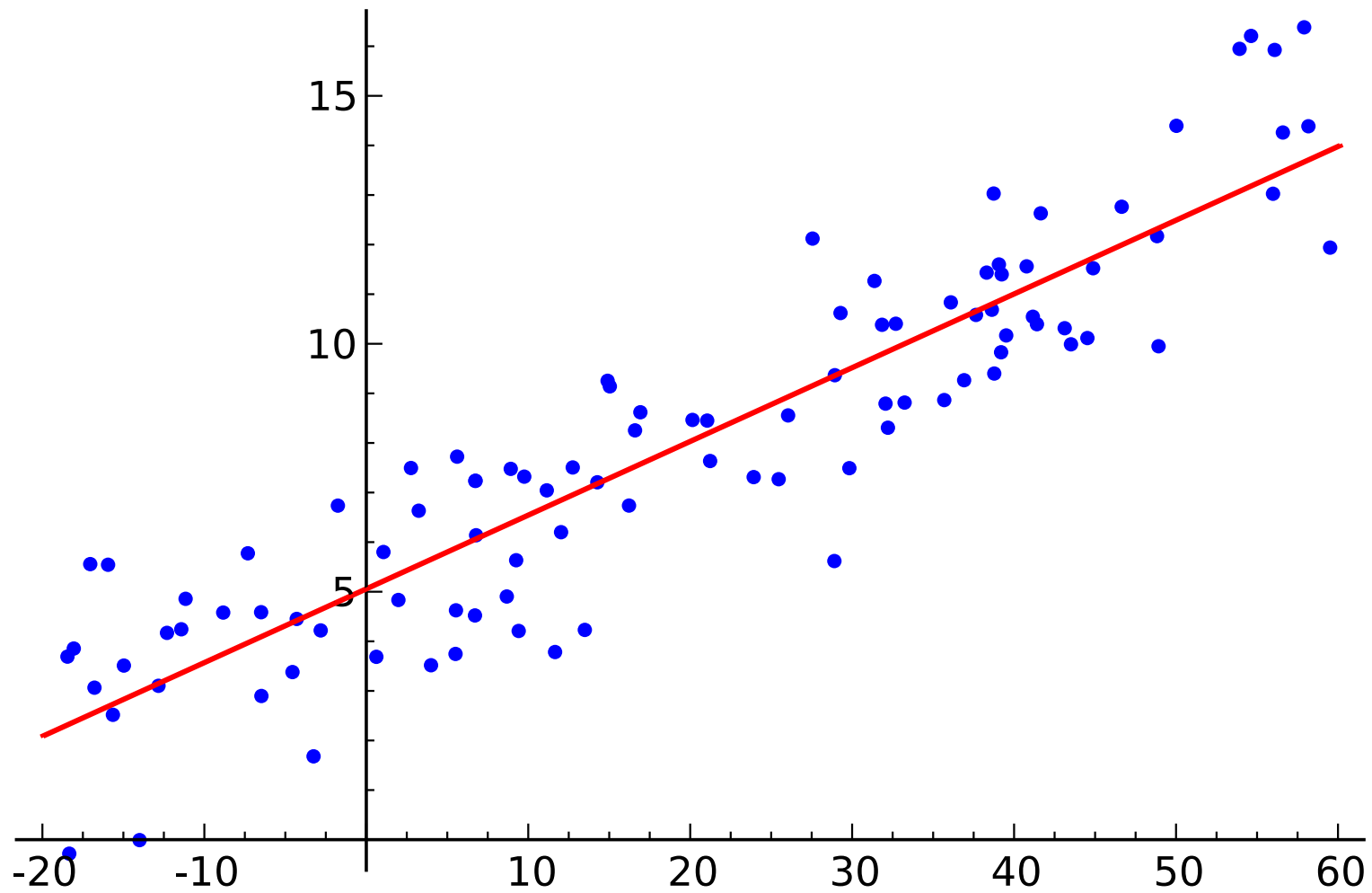
# 准备实验环境

- ❖ 使用 pip 安装 scikit-learn:
- ❖ `pip3 install --user -U scikit-learn`
  
- ❖ 如果安装不成功, 可尝试使用 apt 安装:
- ❖ `sudo apt install python3-sklearn`

# 线性回归

- ❖ 线性回归 (Linear Regression)
- ❖ 找到一个线性方程尽可能的表达原始样本数据的分布
- ❖ 我们先考虑一个最简单的场景，样本的特征维度只有一维
- ❖ 寻找： $f(x) = wx + b$
- ❖ 使得： $f(x) \cong y$
- ❖ 找到这个线性方程之后
- ❖ 对每一个样本来说，给定其特征  $x$  就能预测对应的  $f(x)$

# 线性回归



# 线性回归

- ❖ 刚才我们考虑的场景比较简单，样本的特征维度只有一维
- ❖ 如果  $x$  具备  $d$  个特征： $x = (x_1, x_2, \dots, x_d)$
- ❖ 那么我们就需要寻找： $f(x) = w^T x + b$
- ❖ 使得： $f(x) \cong y$
- ❖ 这称为“多元线性回归（Multivariate Linear Regression）”



# 线性回归实验

```
lr.py ×  
1 import numpy  
2 from sklearn.linear_model import LinearRegression  
3  
4 X = numpy.array([[1, 1], [2, 2], [3, 3], [4, 4]])  
5 y = numpy.array([1, 2, 3, 4])  
6 m = LinearRegression().fit(X, y)  
7 print(m.score(X, y))  
8 print(m.predict(numpy.array([[5, 5]])))  
9
```


# 线性回归实验

```
-----  
~/Workspace/test » python3 lr.py  
1.0  
[5.]  
-----
```

# 线性回归实验

lr.py ×

```
1 import numpy
2 from sklearn.linear_model import LinearRegression
3
4 X = numpy.array([[1, 1], [2, 2], [3, 3], [4, 4]])
5 y = numpy.array([1, 2, 3, 4])
6 m = LinearRegression().fit(X, y)
7 print(m.score(X, y))
8 print(m.predict(numpy.array([[5, 6]])))
9
10
```



# 线性回归实验

```
-----  
~/Workspace/test » python3 lr.py  
1.0  
[5.5]  
-----
```

# 线性回归实验

```
lr.py ×  
1 import numpy  
2 from sklearn.linear_model import LinearRegression  
3  
4 X = numpy.array([[1, 1], [2, 2], [3, 3], [4, 4]])  
5 y = numpy.array([1, 2, 3, 4])  
6 m = LinearRegression().fit(X, y)  
7 print(m.score(X, y))  
8 print(m.coef_)  
9 print(m.predict(array([[5, 6]])))  
10  
11
```

# 线性回归实验

```
~/Workspace/test » python3 lr.py
```

```
1.0
```

```
[0.5 0.5]
```

```
[5.5]
```



# 线性模型特点

- ❖ 线性模型形式简单、易于建模
- ❖ 但却蕴涵着机器学习中一些重要的基本思想
- ❖ 许多功能更为强大的非线性模型（Non-linear Model）可在线性模型的基础上通过引入层级结构或高维映射而得
- ❖ 此外，由于  $w$  直观表达了各属性在预测中的重要性
- ❖ 因此线性模型有很好的可解释性（Comprehensibility）

# 作业



❖ UCI 葡萄酒质量数据集

❖ <https://archive.ics.uci.edu/ml/datasets/wine+quality>

<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	4898	<b>Area:</b>	Business
<b>Attribute Characteristics:</b>	Real	<b>Number of Attributes:</b>	12	<b>Date Donated</b>	2009-10-07
<b>Associated Tasks:</b>	Classification, Regression	<b>Missing Values?</b>	N/A	<b>Number of Web Hits:</b>	1992906

❖ 两套（红葡萄酒、白葡萄酒）的物理化学性质数据，已量化

❖ 这些葡萄酒对应的品质等级



# 作业

- ❖ 获取白葡萄酒数据: <https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv>
- ❖ 取后 100 条为测试数据, 其他为训练数据
- ❖ 用训练数据训练一个线性回归模型
- ❖ 在测试数据上测试训练好的线性回归模型
- ❖ 使用均方误差 (MSE) 计算预测准确度
  
- ❖ 提供完整的代码
- ❖ 提供完整的实验结果截图

# 作业

```
~/Workspace/test » python3 lr.py
```

```
[[ 7.      0.27  0.36  ...  0.45  8.8    6.   ]
```

```
 [ 6.3    0.3    0.34  ...  0.49  9.5    6.   ]
```

```
 [ 8.1    0.28  0.4    ...  0.44 10.1   6.   ]
```

```
...
```

```
 [ 6.5    0.24  0.19  ...  0.46  9.4    6.   ]
```

```
 [ 5.5    0.29  0.3    ...  0.38 12.8   7.   ]
```

```
 [ 6.     0.21  0.38  ...  0.32 11.8   6.   ]]
```

```
[5.63144651 6.10115806 5.85674592 ... 5.34351476 6.57567039 6.35192467]
```

```
[5. 5. 5. ... 6. 7. 6.]
```

```
0.5630446931574506
```

# 作业

- ❖ 在作业系统中下载并完成本实验课对应实验报告
- ❖ <https://hw.dgut.edu.cn/>
- ❖ 注意：所有标识为 \* 的地方都需要填写
- ❖ 截止日期：2024-04-01 23:59

课程名称：人工智能概论

实验名称	回归模型		
姓名	***	学号	***
实验地点	***	实验日期	***

## 四、实验作业及分析

### 4.1 实验过程

\*\*\* 请将详细实验过程的截图和相关说明填写在此处 \*\*\*

### 4.2 实验结果

\*\*\* 请将实验结果的截图和相关说明填写在此处 \*\*\*

