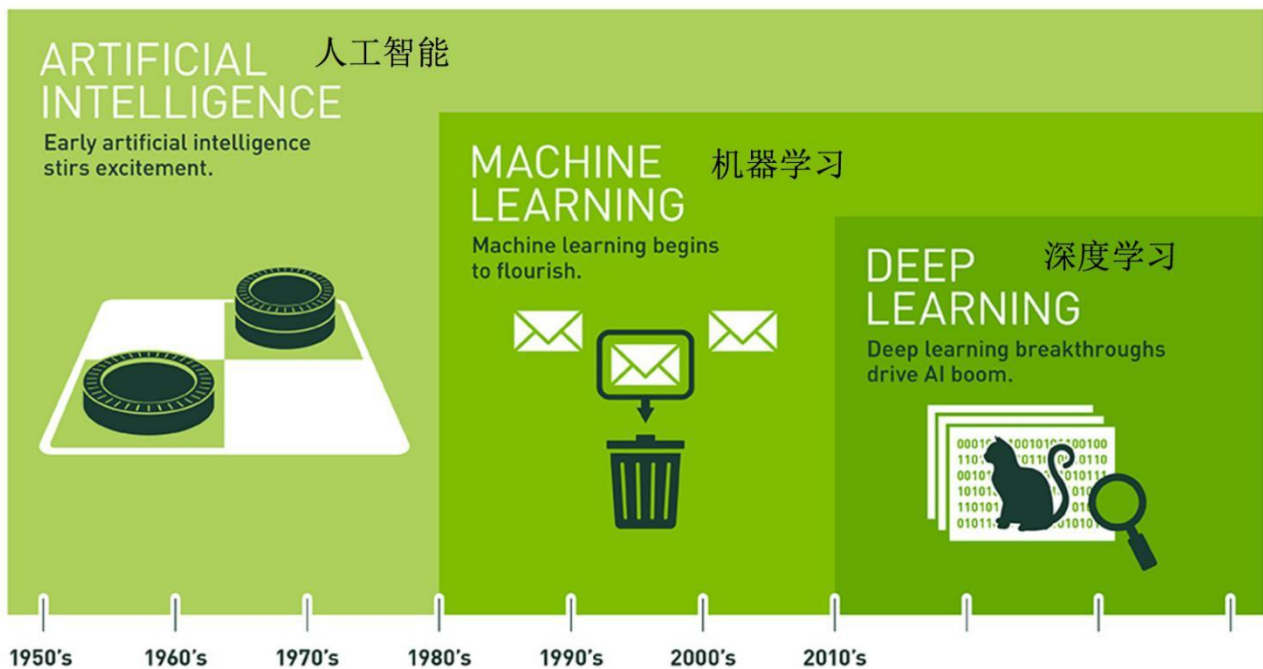


机器学习

一、什么是机器学习

1. 人工智能简介



<https://www.zhihu.com/question/57770020>

机器学习和人工智能，深度学习的关系

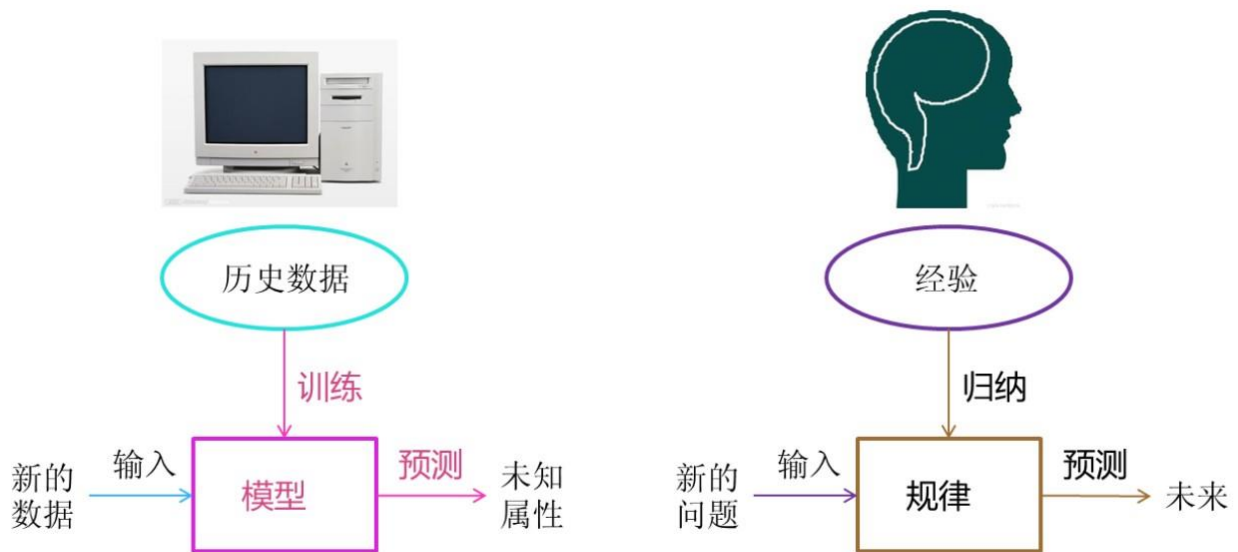
机器学习是人工智能的一个实现途径

深度学习是机器学习的一个方法发展而来

2. 什么是机器学习

什么是（监督式）机器学习？简单来说，它的定义如下：

- 机器学习系统通过学习如何组合输入信息来对从未见过的数据做出有用的预测。



我们人从大量的日常经验中归纳规律，当面临新的问题的时候，就可以利用以往总结的规律去分析现实状况，采取最佳策略。

3.人工智能，机器学习，深度学习

经典的机器学习：算法为王，时间和空间

$1 + 2 + 3 + 4 + 100 = ..$

$(1+100) * 100 / 2 = ..$

基于神经网络的机器学习：

浅层学习

深度学习（深度学习） 大量数据（几十 P）

强化学习

迁移学习

2 -> 4

1 -> 2

100 -> ?

100 -> 200

二、机器学习基本类型

1、有监督学习

举例：从小长辈告诉我们什么是猫，什么是狗。从没见过哈士奇，我们能准备判别出来它是狗。

根据已知的输入（特征）和输出（标签），建立联系他们的模型，根据该模型对未知输出的输入提出判别。

1) 回归：以无限连续域的形式表示输出

2) 分类：以有限离散域的形式表示输出

2、无监督学习

举例：从小从来没有见过猫和狗，现在去到一个动物园，里面看到：

大多数耳朵竖起来，喜欢摇尾巴的 A 类总聚在一起，大多数身上有条纹的，喜欢爬树的比较高冷 B 类聚在一起。

走在街上，再次看到，我们可以叫它 A 或者 B。

在一组没有已知输出的输入中，根据数据的内部特征和联系，找到某种规则，进行线性的划分。

聚类，pca 降维

3、半监督学习

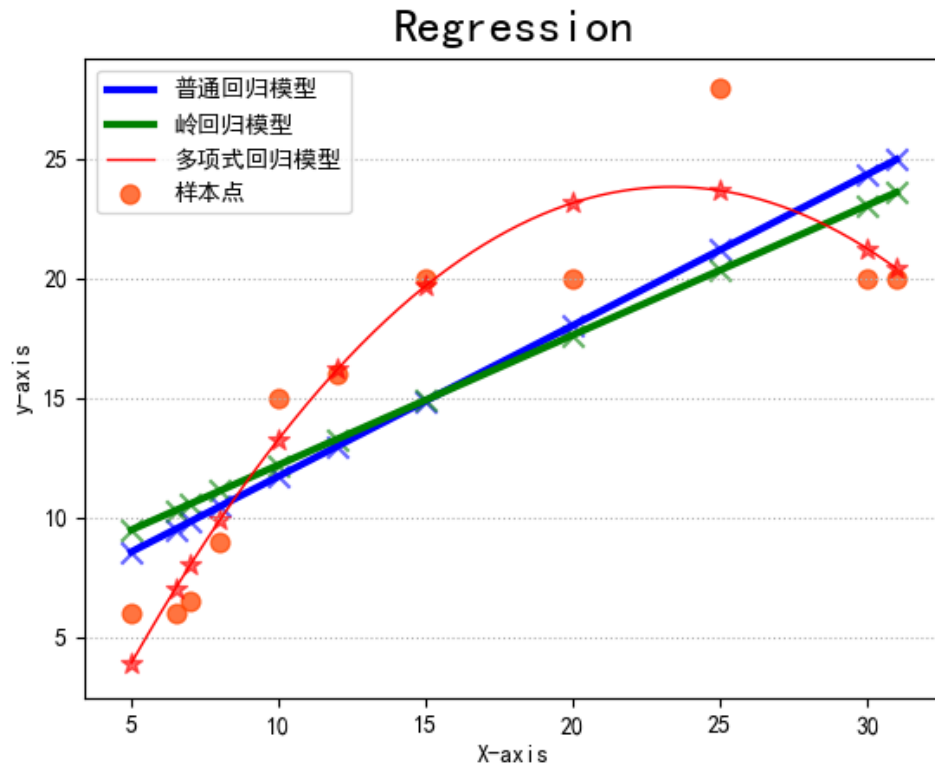
从一个相对有限的已知结构中利用有监督学习的方法，构建基本模型，通过未知输入和已知输入的对比，判断其输出。

4、机器学习的基本过程

数据采集 --> 数据清洗 --> 数据预处理 --> 选择模型(算法) --> 模型训练 --> 模型测试 --> 模型优化 --> 模型应用

数据决定了模型的上限，算法决定了模型的下限。

三、机器学习回归算法



1、线性回归

$\text{loss} = J(k, b)$

示例代码: 00-line_regression.py

2、岭回归

岭回归的 R 平方会稍低于普通回归，回归系数的显著性明显高于普通回归。

主要解决一些病态数据具有较大的价值。

$\text{loss} = J(k, b) + \text{正则函数}(\text{样本权重}) * \text{正则强度}$

示例: 01-ridge.py

3、多项式回归

x, y 是已知, 线性回归是特殊的多项式回归

$y = kx + b$

```
y = kx^2 + b
y = k1x + k2x^2 + k3x^3 + ... + knx^n
示例代码: 02-poly.py
import sklearn.linear_model as le
import sklearn.preprocessing as sp
from sklearn import pipeline
多项式回归模型 = pipeline.make_pipeline(sp.PolynomialFeatures(最高次
幂),le.LinearRegression())
```

4、保存模型

```
from sklearn.externals import joblib
joblib.dump(模型,"模型名称.m") -->保存模型
加载后的模型 = joblib.load("模型名称.m")
示例代码: 03-joblib.py
```