

# 《人工智能-总览、应用与前沿》

**主讲老师：高彦杰**

## 课程信息

**人工智能技术和应用场景的介绍。常用人工智能的算法，主流的应用构建方法。主流机器学习框架介绍，针对机器学习场景能够更好的应用相关工具进行分析和处理。**

# 课程大纲

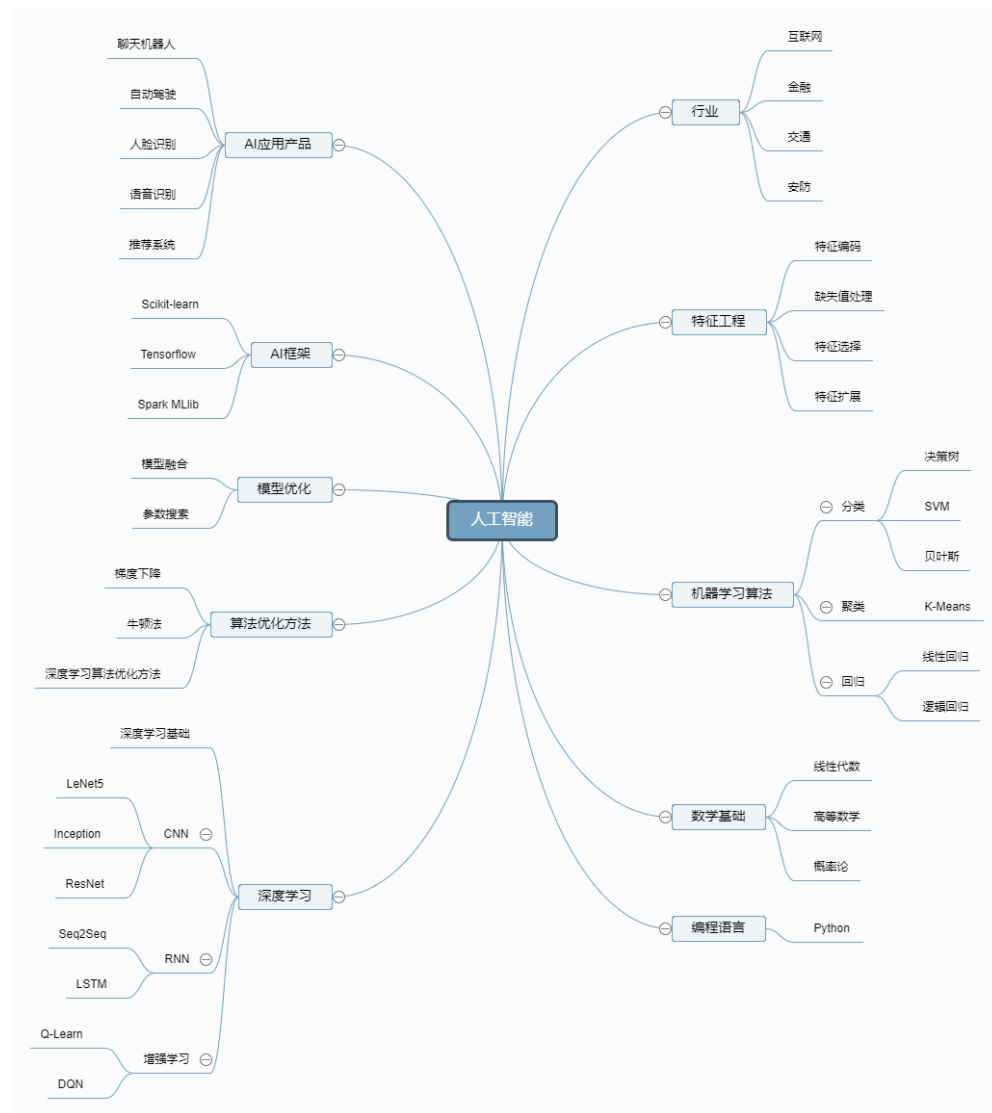
1. 前置，课程概览(★★★)
2. 人工智能简介，过去现在与未来(★)
3. 人工智能应用场景(★★)
4. 人工智能技术概览(★★)
5. 机器学习算法(★★★)
6. 主流机器学习库和计算框架介绍(★★)
7. 如何进行机器学习(★★★)

# 前置与学习方法

# 课程概览

- 1) 人工智能基础
- 2) 机器学习算法
- 3) 深度学习
- 4) 人工智能大数据框架应用
- 5) 企业级人工智能项目实战

# 知识点覆盖



# 课程概览

- 1) 人工智能基础
- 2) 机器学习算法
- 3) 深度学习
- 4) 人工智能大数据框架应用
- 5) 企业级人工智能项目实战

# 课程概览

## 1) 人工智能基础

人工智能简介

数学与数学分析基础

特征工程



# 课程概览

## 2) 机器学习算法

决策树与随机森林

分类算法

回归算法

聚类算法



# 课程概览

## 3) 深度学习

深度学习-基础

深度学习-高级



# 课程概览

## 4) 人工智能大数据框架应用

深度学习TensorFlow

大数据与机器学习-Spark MLlib

# 课程概览

## 5) 企业级人工智能项目实战

语音识别

自动驾驶

图像人脸识别

聊天机器人

# 知识体系



## 习“武” (AI) 心得

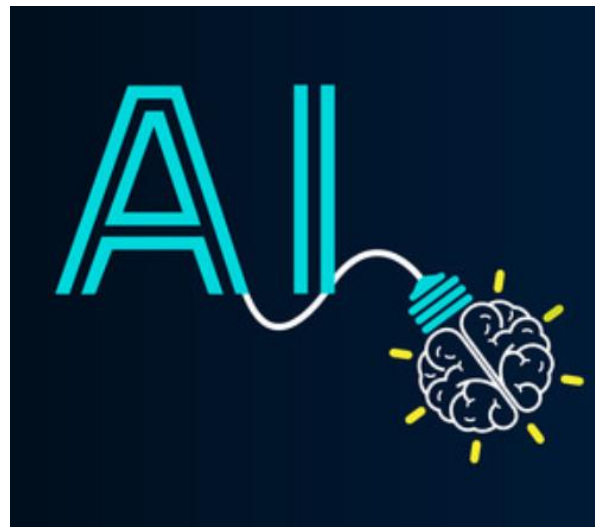
- 练拳不练把，等于胡乱打
- 招没绝，功夫有绝
- 冬练三九，夏练三伏
- 师傅领进门，成艺在自身
- 只要功夫深铁杵磨成针

人工智能



# 学习方法

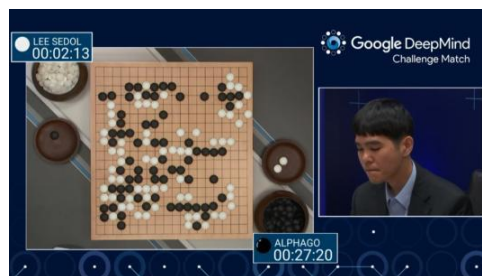
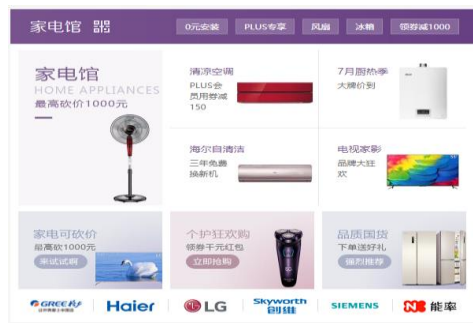
- 大纲为重点知识主线与脉络，形成知识体系
- 有余力情况下深度广度方向拓展学习
- 动手练习学习，加深理解
- 组内多交流与知识共享
- 打好通用技术基础，后期毕设和未来工作中在一个主要技术方向多沉淀



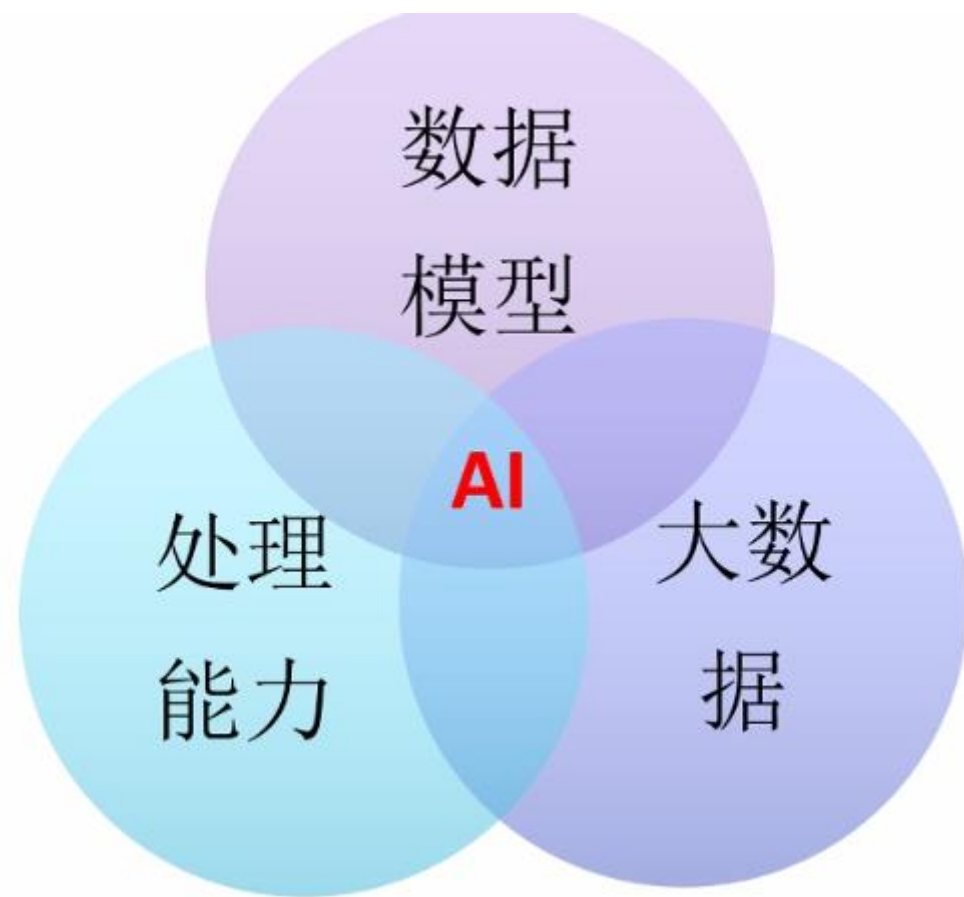
# 人工智能简介



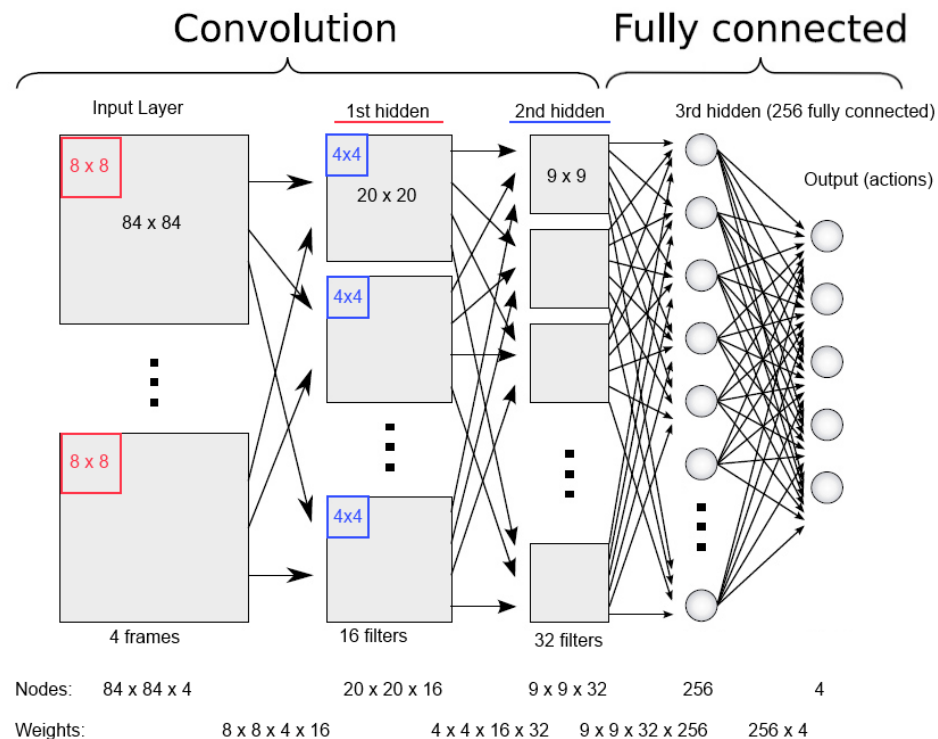
# 人工智能的加速发展



## 人工智能核心点



© 2006 The Authors  
Journal compilation © 2006 Blackwell Publishing Ltd



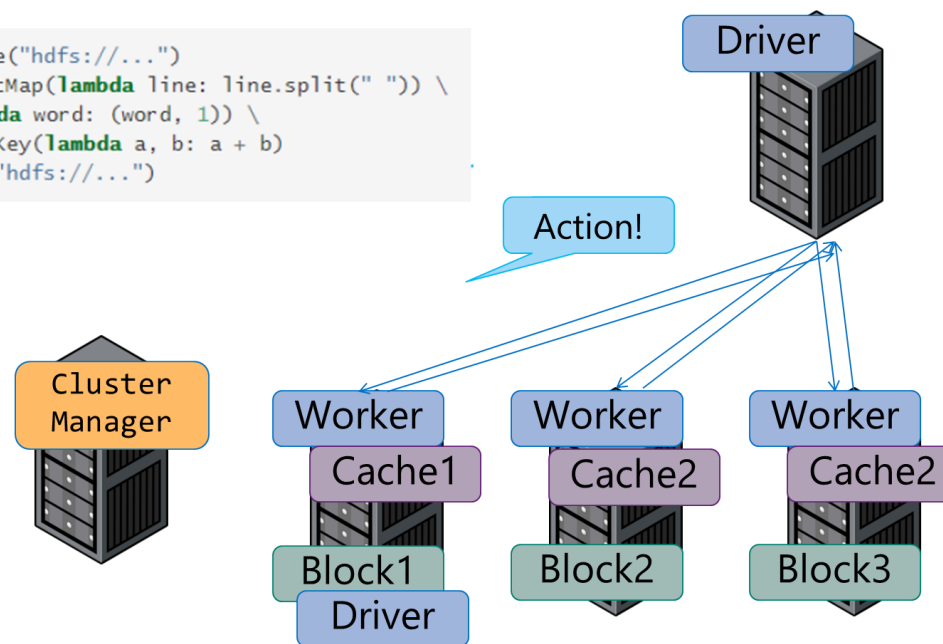
## 大数据和深度学习使得人工智能取得突破进展

# 计算能力的提升加速人工智能的发展



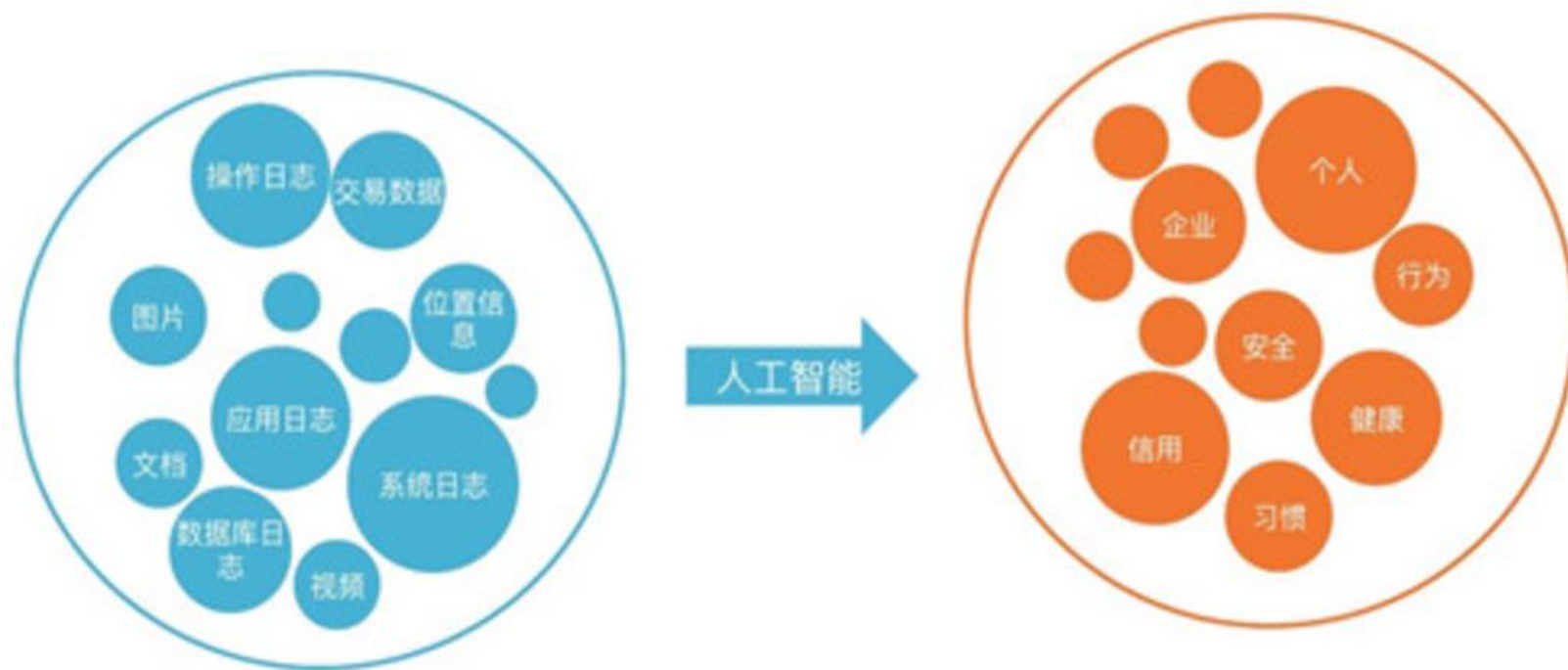
## Spark架构

```
text_file = sc.textFile("hdfs://...")
counts = text_file.flatMap(lambda line: line.split(" ")) \
    .map(lambda word: (word, 1)) \
    .reduceByKey(lambda a, b: a + b)
counts.saveAsTextFile("hdfs://...")
```



单机数据分析向分布式计算发展

# 大数据的理解与分析需要AI



人工智能可学习大量大数据，高效分析挖掘出数据价值。

# AI 历史

## 三次浪潮

- ✓ 1956~1976 专家系统
- ✓ 1976~2006 人工神经网络，深度学习尚未突破
- ✓ 2006~至今 基于互联网大数据的深度学习

## 现状

- ✓ 应用层：互联网，安防，医疗，金融，运营商
- ✓ 接口层：Tensorflow，Scikit-Learn，Spark MLlib等框架接口
- ✓ 算法层：深度学习发展迅速，一些关键场景其他AI算法仍有用武之地
- ✓ 框架层：Tensorflow等框架
- ✓ 基础设施层：CPU与GPU集群大范围使用，云平台蚕食传统Data Center市场

## 趋势

- ✓ 应用层：深入各个领域
- ✓ 接口层：Workflow, Model Sharing
- ✓ 算法层：AutoML, 算法、系统发展
- ✓ 框架层：框架中间层编译优化, 开放协议ONNX
- ✓ 基础设施层：GPU集群加速及管理, ASIC, FPGA



# 什么时候用AI？

- 利用数据来解决简单规则无法或者难以解决的问题，它被广泛应用在了搜索引擎、无人驾驶、机器翻译、医疗诊断、人脸识别、数据匹配、信用评级等任务中。
- 我们无法直接编程解决这些问题，但我们能够使用配合数据编程来解决。

如果给我们的机器学习系统提供足够多猫和狗的图片，我们可以“编写”一个喵星人辨别器：



喵



喵



汪



汪

# AI编程 VS 基于规则编程？

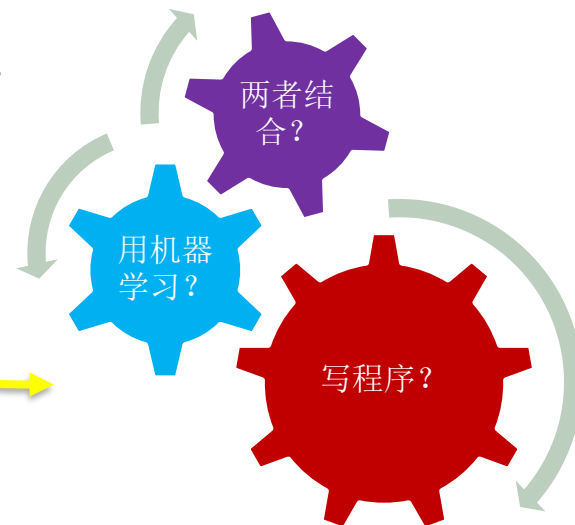
这个例子灵感来自 Joel Grus 的一次 应聘面试. 面试官让他写个程序来玩Fizz

Buzz. 这是一个小孩子游戏。玩家从1数到100，如果数字被3整除，那么喊

'fizz'，如果被5整除就喊'buzz'，如果两个都满足就喊'fizzbuzz'，不然就直接

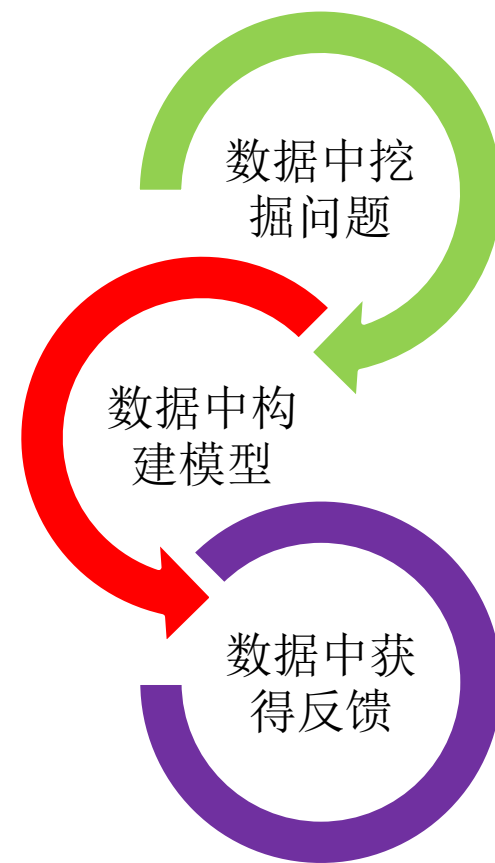
说数字。这个游戏玩起来就像是：

1 2 fizz 4 buzz fizz 7 8 fizz buzz 11 fizz 13 14 fizzbuzz 16 ...



## 传统思维 VS AI思维？

# AI时代的思维方式



# 人工智能应用场景

# 智能客服&助手



udesk

首页 呼叫中心 在线客服 工单系统 客户案例 帮助


注册 登录

## 智能机器人客服

成本降低80%

通过机器算法自动回复常见问题，大幅提高客服工作效率。

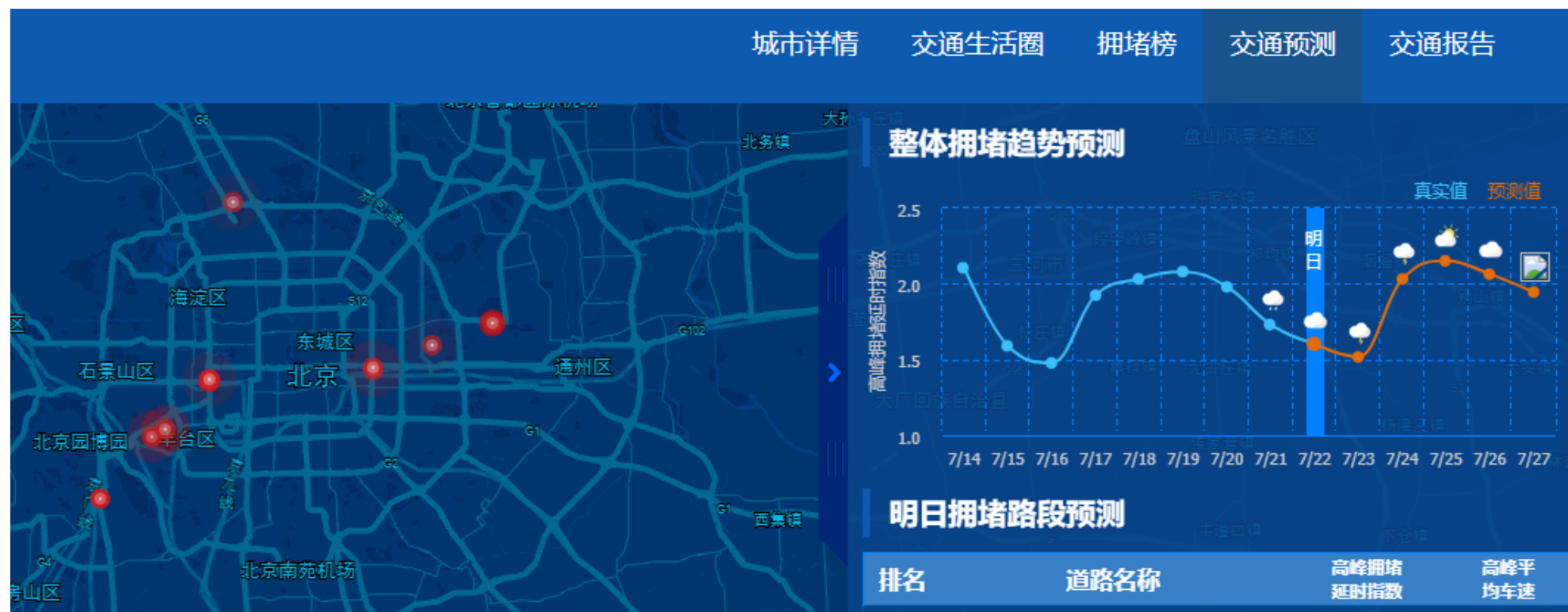
立即体验



### 联想智能机器人·小乐

24小时为您服务

# 实时视频分析：交通拥堵预测



# 个性化推荐：千人千面的电商

女装精品

内衣/男装

鞋品/箱包

饰品/配件

运动/户外

家具/家电

居家/家纺

智能/数码

影音/小家电

母婴/玩具

汽车/娱乐

美妆/个护

裙子

连衣裙碎花连衣裙雪纺连衣裙棉麻连衣裙

一字领连衣裙蕾丝连衣裙露背裙中长款连衣裙

吊带连衣裙半身裙

裤子

短裤夏打底裤休闲裤牛仔阔腿裤

牛仔短裤哈伦裤铅笔裤

T恤/衬衫

T恤纯棉T恤雪纺T恤T恤潮T恤长款宽松T恤

修身T恤大码T恤长袖衬衫小背心针织开衫

婚纱礼服

婚纱礼服旗袍长拖尾蓬蓬裙婚纱鱼尾婚纱

宴会礼服晚礼服派对礼服

韩都衣舍

雪梨网红定制

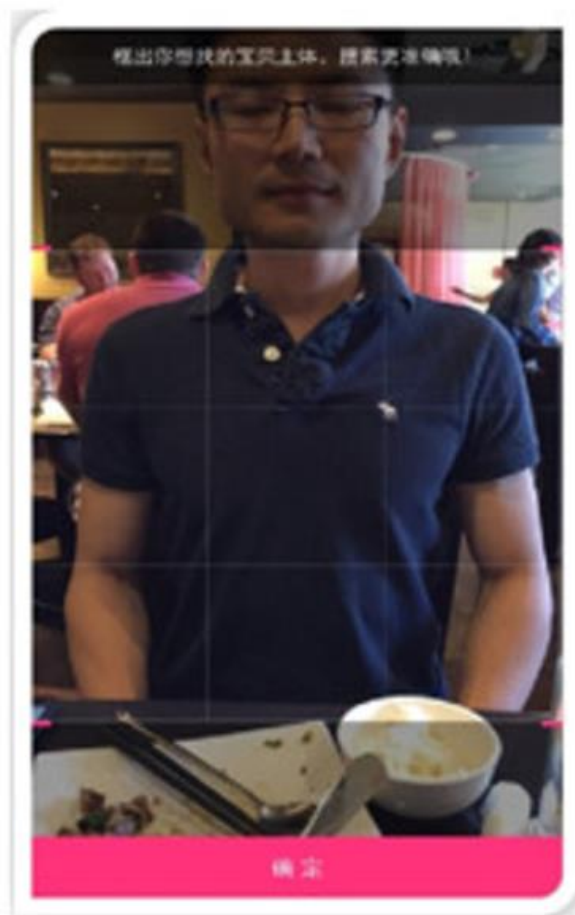
婚纱老店

淘抢购

更多爱逛街尖端时尚风格潮流更多



# 图像识别-同款商品推荐



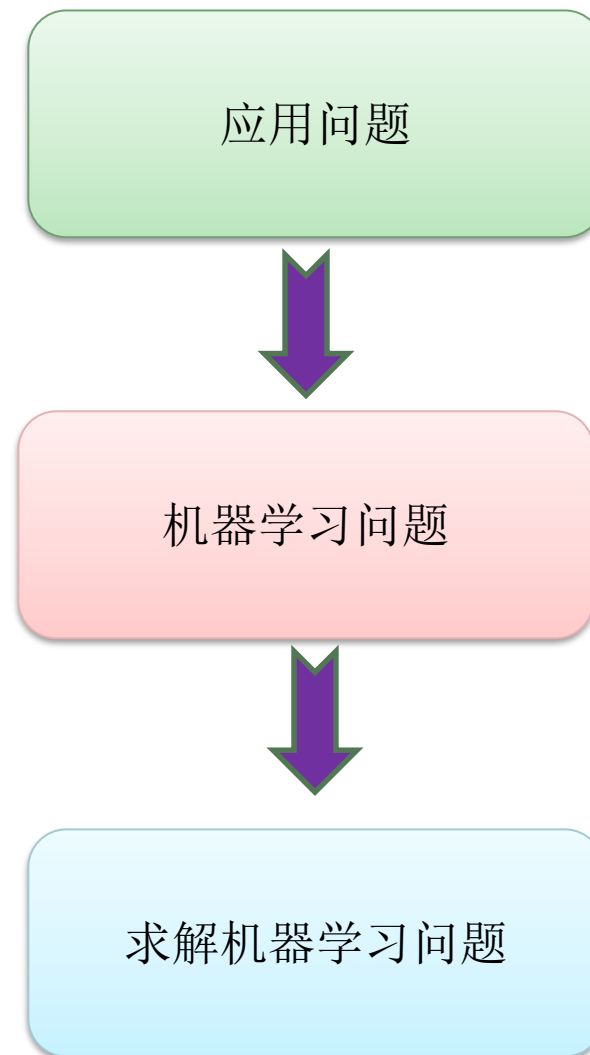


# 征信

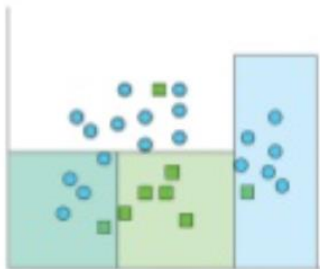


# 应用到机器学习问题的映射

- ✓ 征信:
  - ✓ 用户特征 -> 征信评分
  - ✓ 回归问题
- ✓ 交通拥堵预测:
  - ✓ 历史和实时信息 -> 拥堵指数
  - ✓ 回归问题
- ✓ 对话机器人:
  - ✓ 句子 -> 句子
  - ✓ Seq2Seq + 分类问题
- ✓ 电商推荐 / 商品推荐:
  - ✓ 用户特征 -> 感兴趣的物品
  - ✓ 推荐 + 回归问题

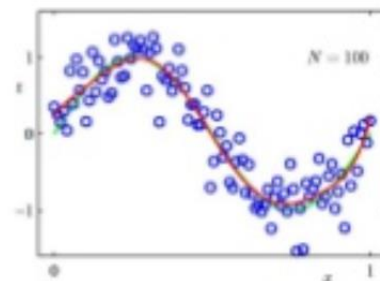


## Classification



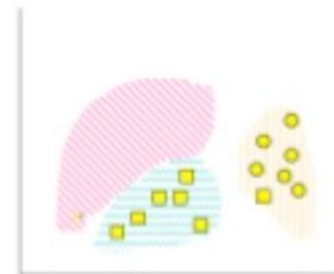
Learns a method for predicting the instance class from pre-labeled (classified) instances

## Regression



An attempt to predict a continuous attribute

## Clustering



Finds "natural" grouping of instances given un-labeled data

# 几何视角看AI问题

# 人工智能技术概览

# AI技术栈

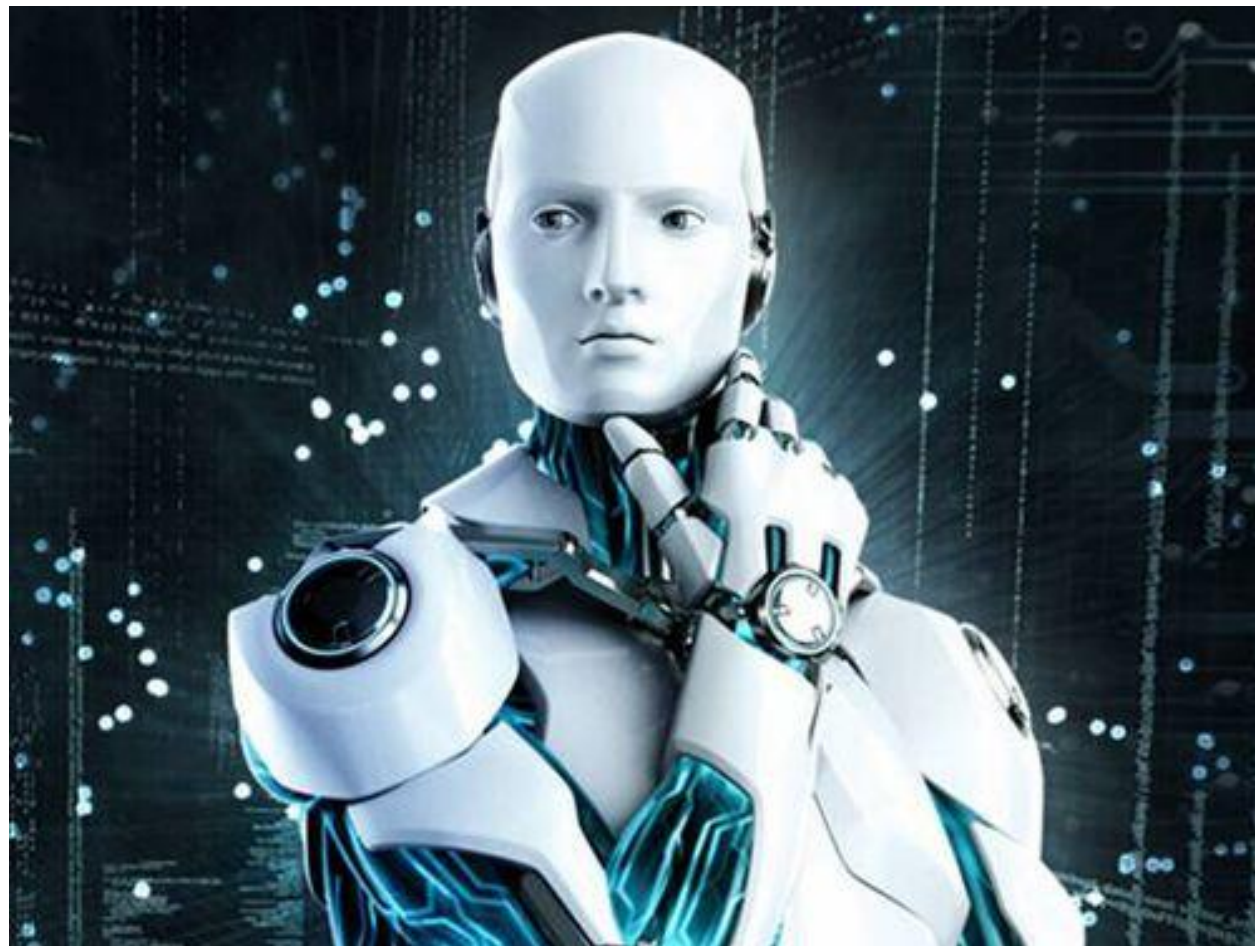
✓ 解决方案

✓ 接口

✓ 算法

✓ 框架

✓ 平台



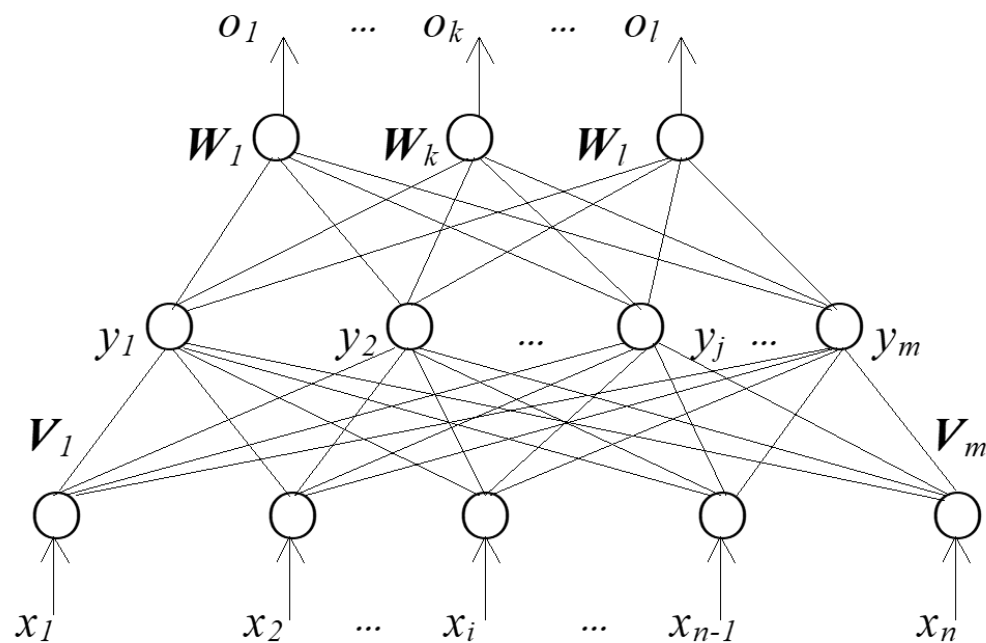
# AI算法

✓ 分类

✓ 回归

✓ 聚类

✓ 深度学习



# 开源系统工具蓬勃发展

- ✓ **特征工程与数据预处理:**

- ✓ **Numpy / Pandas / Spark**

- ✓ **OpenCV**

- ✓ **Jieba / NLTK / Scikit-Learn / gensim**

- ✓ **...**

- ✓ **ML与DL框架:**

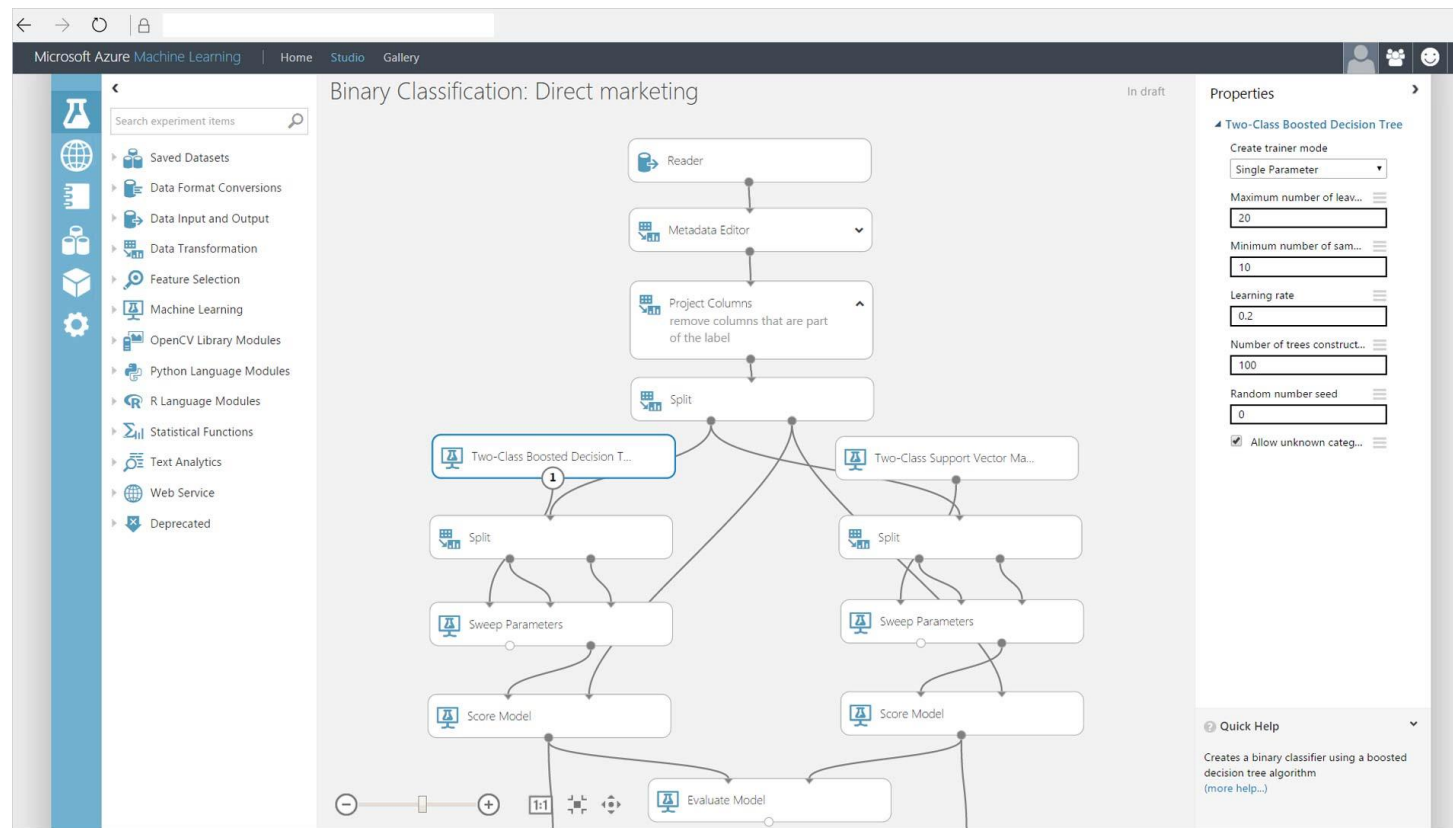
- ✓ **Tensorflow / Keras / Slim**

- ✓ **Scikit-Learn / XGBoost**

- ✓ **Spark Mllib**

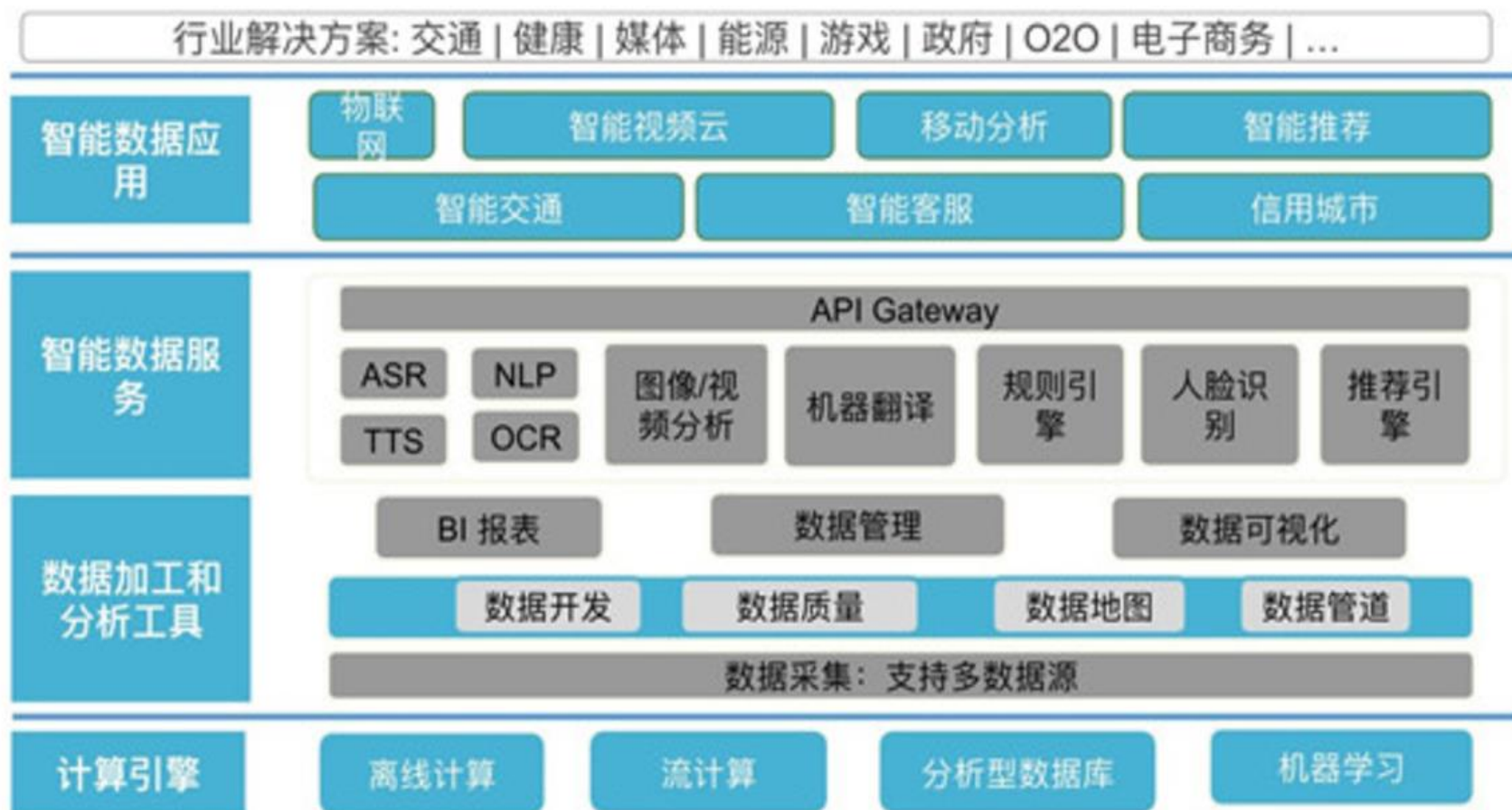
- ✓ **...**

# 人工智能分析平台





# 人工智能解决方案支撑技术栈



# 机器学习算法

# AI算法

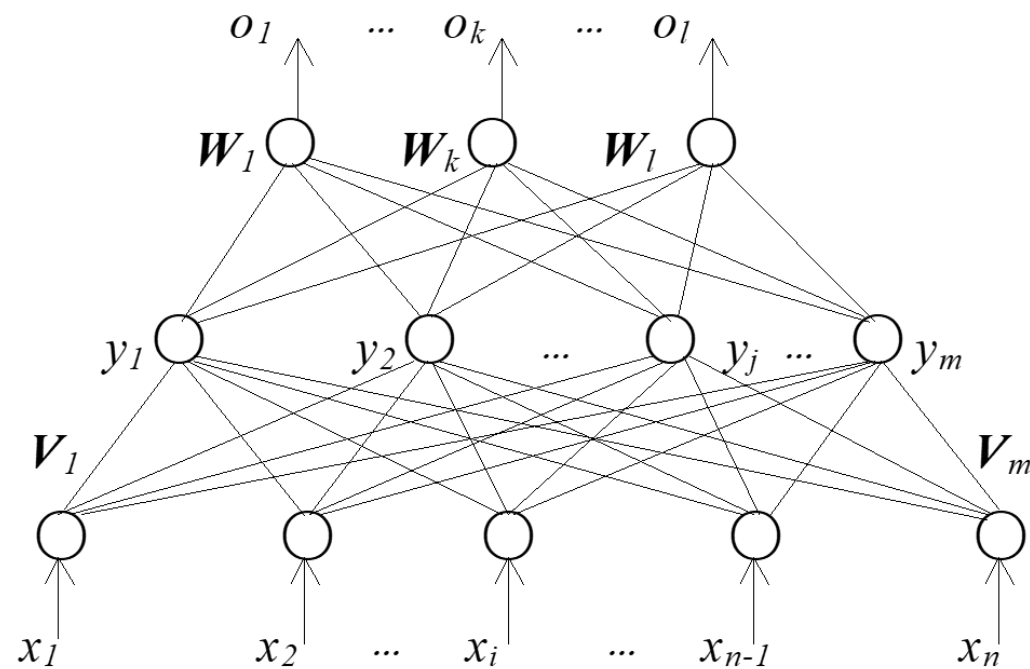
✓ 分类

✓ 回归

✓ 聚类

✓ 深度学习

✓ ...



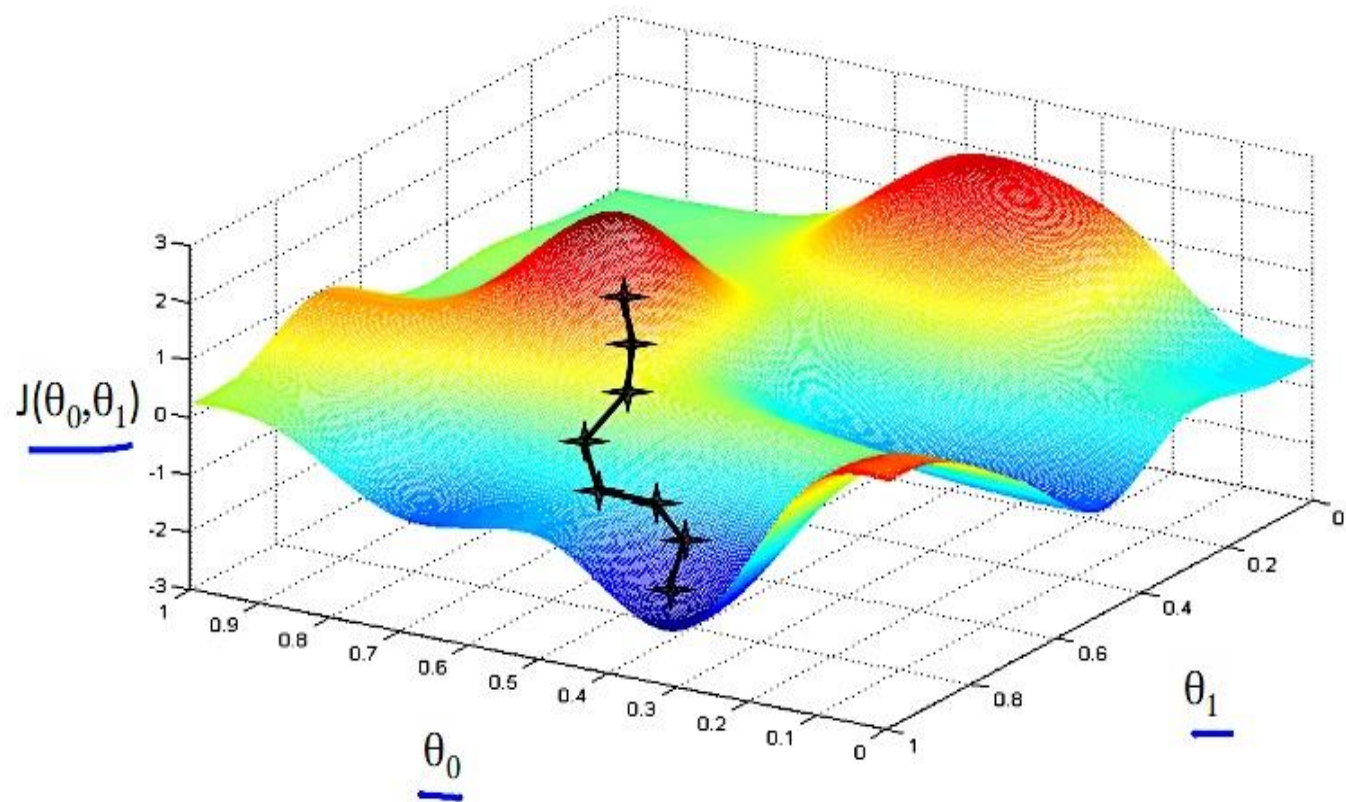
# 基于梯度下降训练的分类与回归算法

✓ 线性回归

✓ 逻辑回归

✓ Lasso

✓ Ridge回归

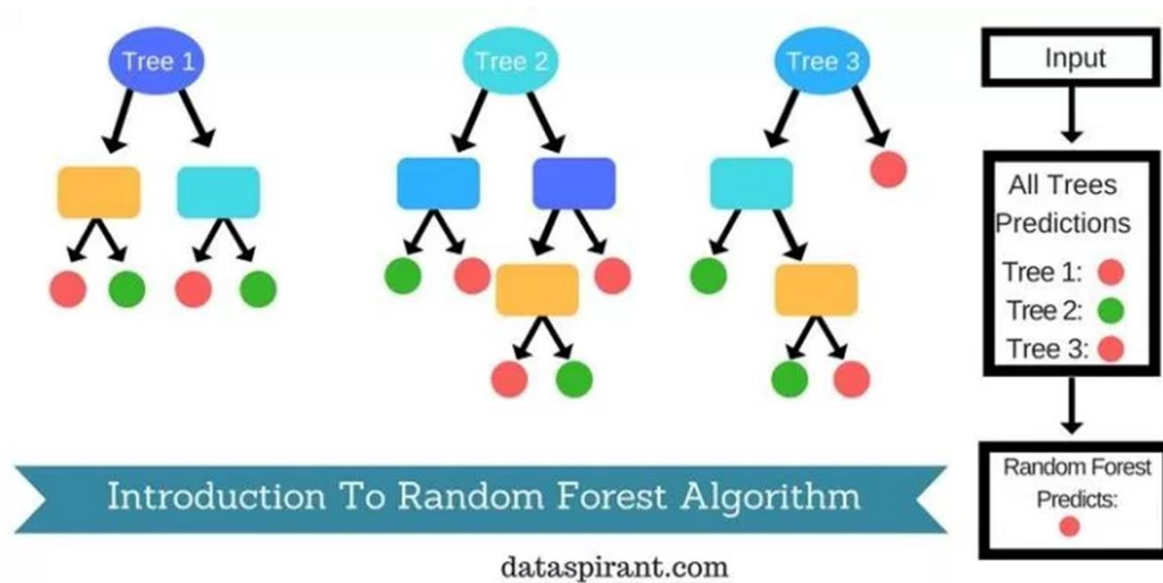


# Tree算法

✓ 决策树

✓ 随机森林

✓ GBDT

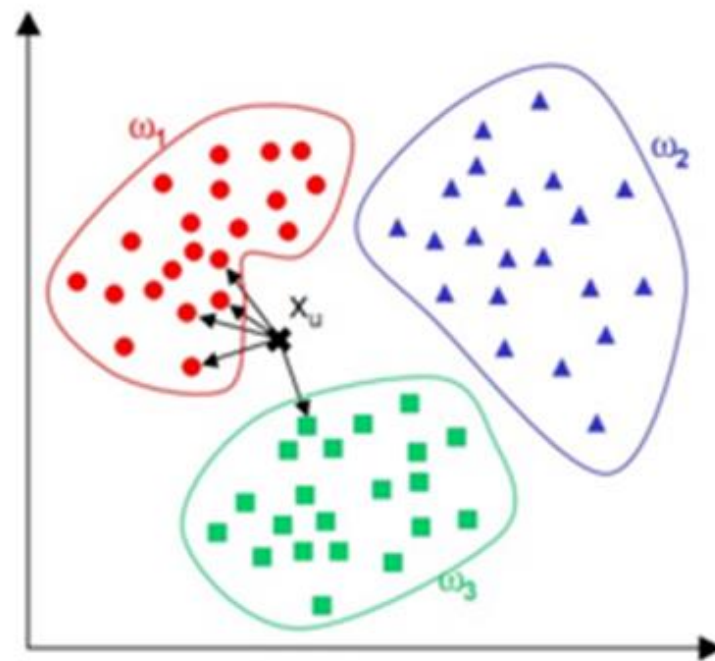


# 聚类算法

✓ K-Means

✓ K-Means ++

✓ 其他聚类算法

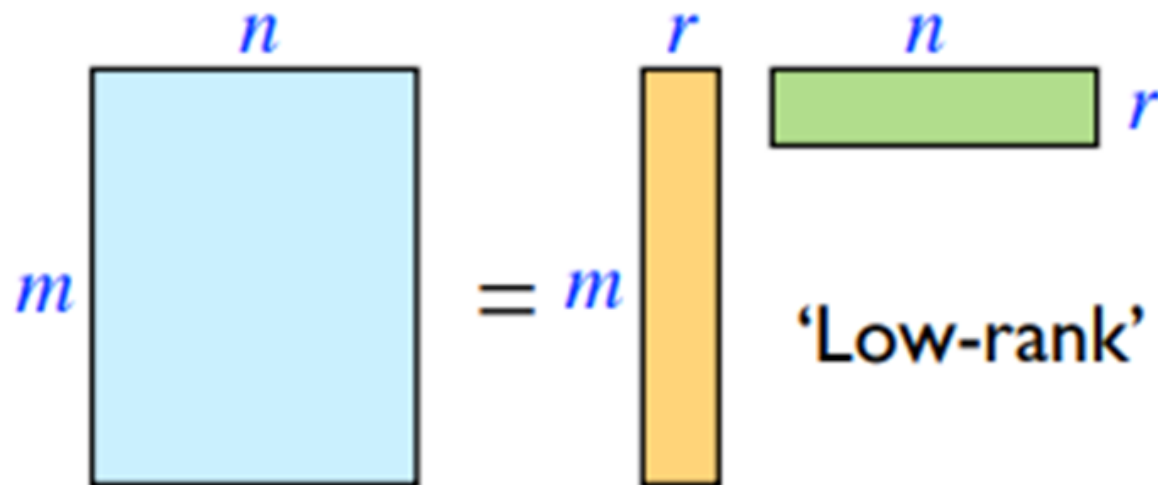


## 其他机器学习算法

✓ KNN

✓ 推荐ALS

✓ SVM等

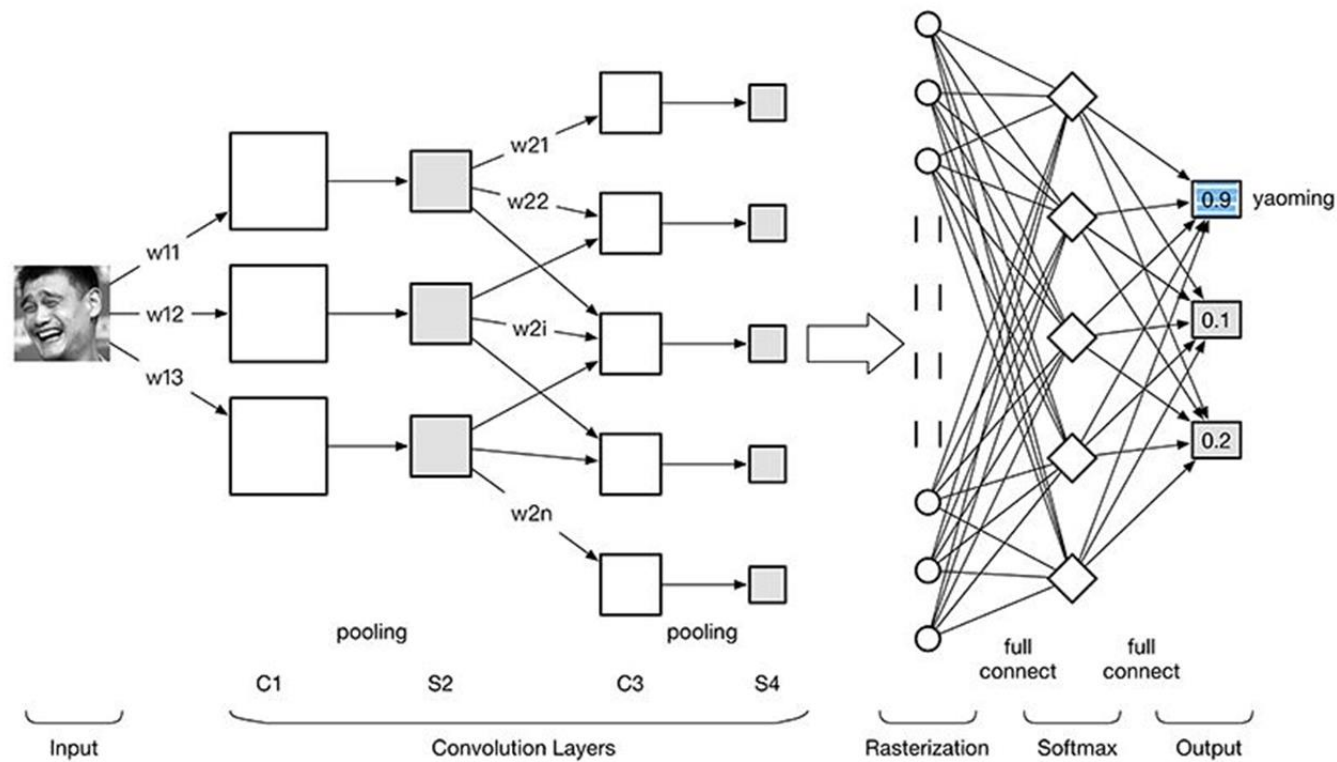


# 深度学习

✓ NN

✓ CNN

✓ RNN





# 主流机器学习库和计算框架介绍

# 开源系统工具蓬勃发展

- ✓ 特征工程与数据预处理：
  - ✓ Numpy / Pandas / Spark / ...
  - ✓ OpenCV / ...
  - ✓ Jieba / NLTK / Scikit-Learn / genism / ...
  - ✓ Matplotlib
- ✓ ML与DL框架：
  - ✓ Tensorflow / Keras / Slim / ...
  - ✓ Scikit-Learn / XGBoost / ...
  - ✓ Spark Mllib / ...
  - ✓ ...

# Scikit-Learn功能模块

## Classification

Identifying to which category an object belongs to.

**Applications:** Spam detection, Image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples

## Dimensionality reduction

Reducing the number of random variables to consider.

**Applications:** Visualization, Increased efficiency

**Algorithms:** PCA, feature selection, non-negative matrix factorization. — Examples

## Model selection

Comparing, validating and choosing parameters and models.

**Goal:** Improved accuracy via parameter tuning

**Modules:** grid search, cross validation, metrics. — Examples

## Preprocessing

Feature extraction and normalization.

**Application:** Transforming input data such as text for use with machine learning algorithms.

**Modules:** preprocessing, feature extraction. — Examples

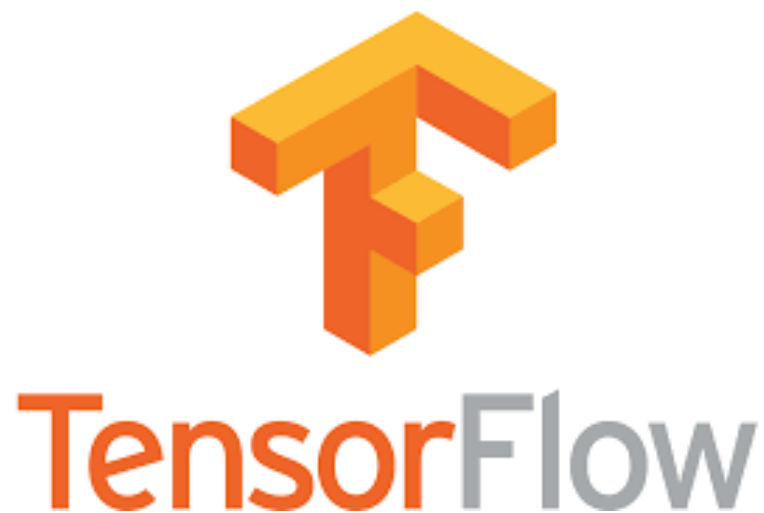
# Spark MLlib

- ✓ MLlib 是Spark的可以扩展的机器学习库
- ✓ 包括分类，回归，聚类，协同过滤，降维
- ✓ 包括调优的部分



# Tensorflow

- ✓ 数据流图 (data flow graphs)
- ✓ 用于数值计算的开源软件库
- ✓ 多种平台上展开计算
- ✓ 用于机器学习和深度神经网络

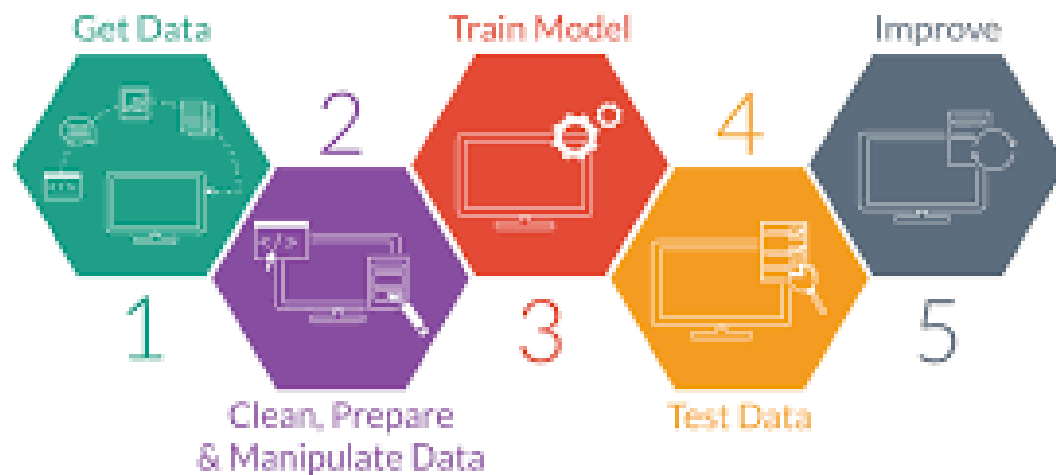


# 如何进行机器学习

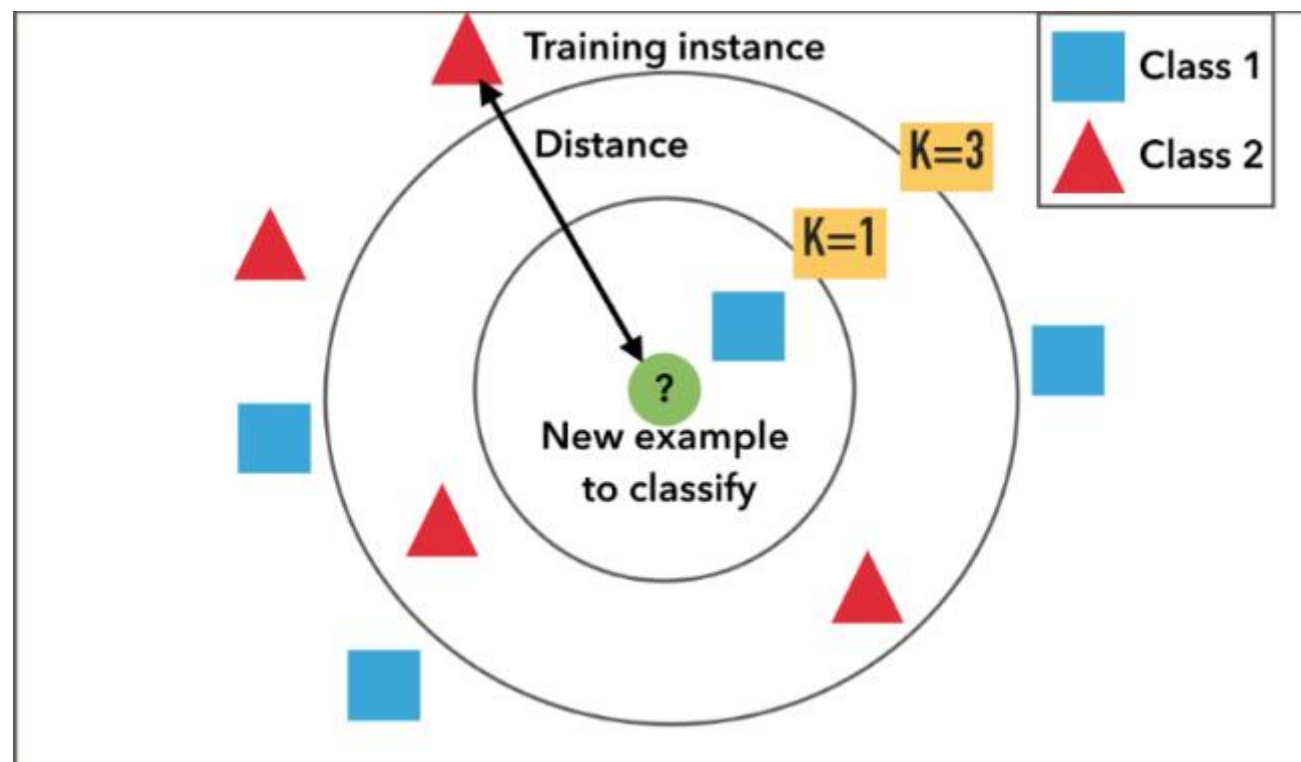
# 机器学习基本工作流程

- ✓ 问题建模
- ✓ 获取数据
- ✓ 特征工程
- ✓ 模型训练与验证
- ✓ 模型诊断与调优
- ✓ 线上运行

Steps to Predictive Modelling



# KNN算法





# 预习/复习重点

✓ 重点:

✓ 1. 课程概览

✓ 2. 人工智能技术概览

✓ 3. FizzBuzz实例

✓ 4. 机器学习流程

✓ 5. KNN算法

✓ 难点:

✓ AI与传统编程区别

✓ 机器学习流程

# 参考资料

- ✓ 统计学习方法
- ✓ 机器学习
- ✓ 深度学习
- ✓ 深度学习：核心技术、工具与案例解析
- ✓ 数据仓库与数据分析教程
- ✓ 数据挖掘：概念与技术
- ✓ Scikit-learn, TensorFlow, Keras, Spark MLlib相关文档