



数据分析与人工智能

主讲人：周景阳（大周）

- 人工智能（Artificial Intelligence），英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。



人工智能的技术划分

人
工
智
能

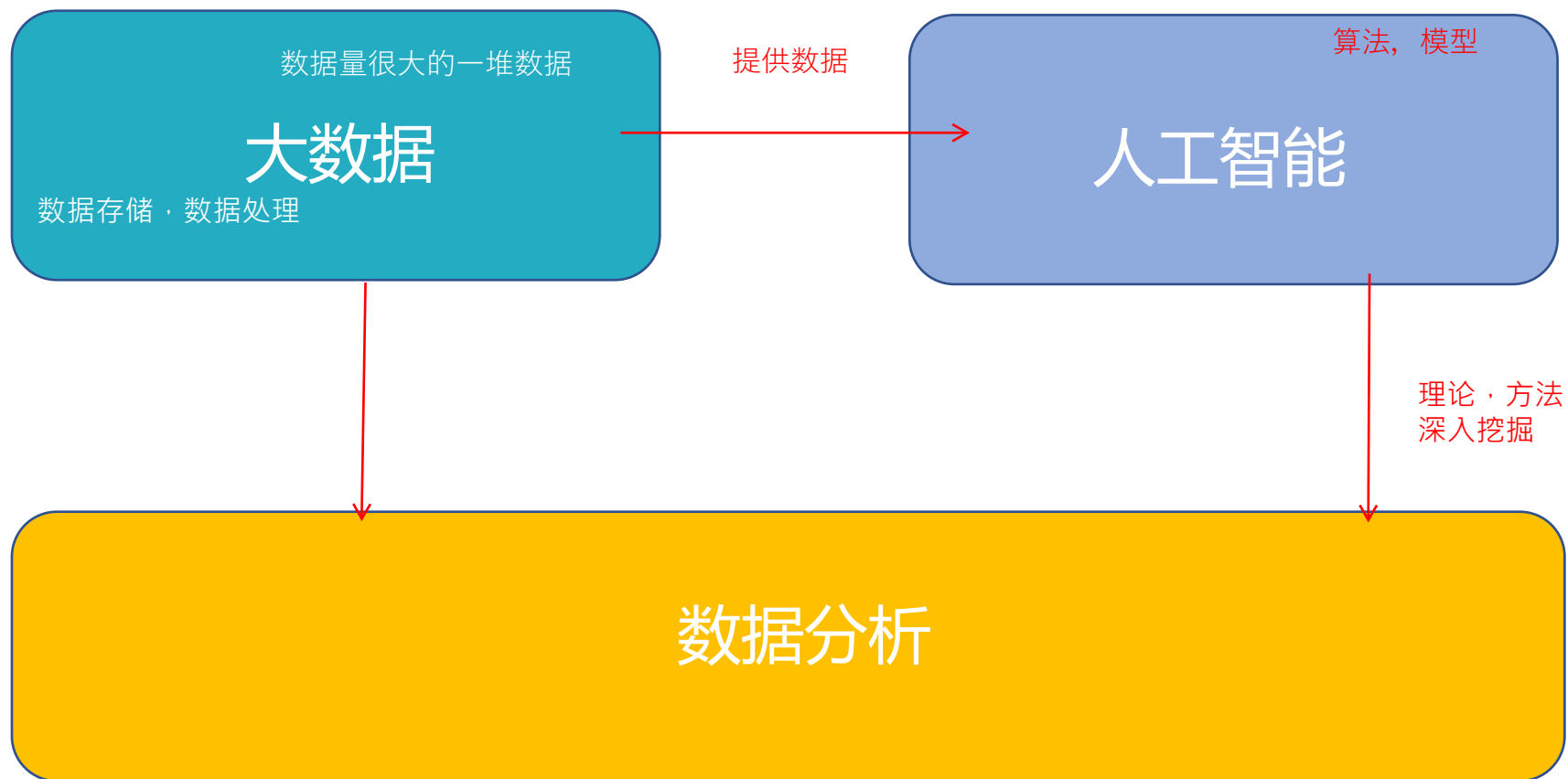
机
器
学
习

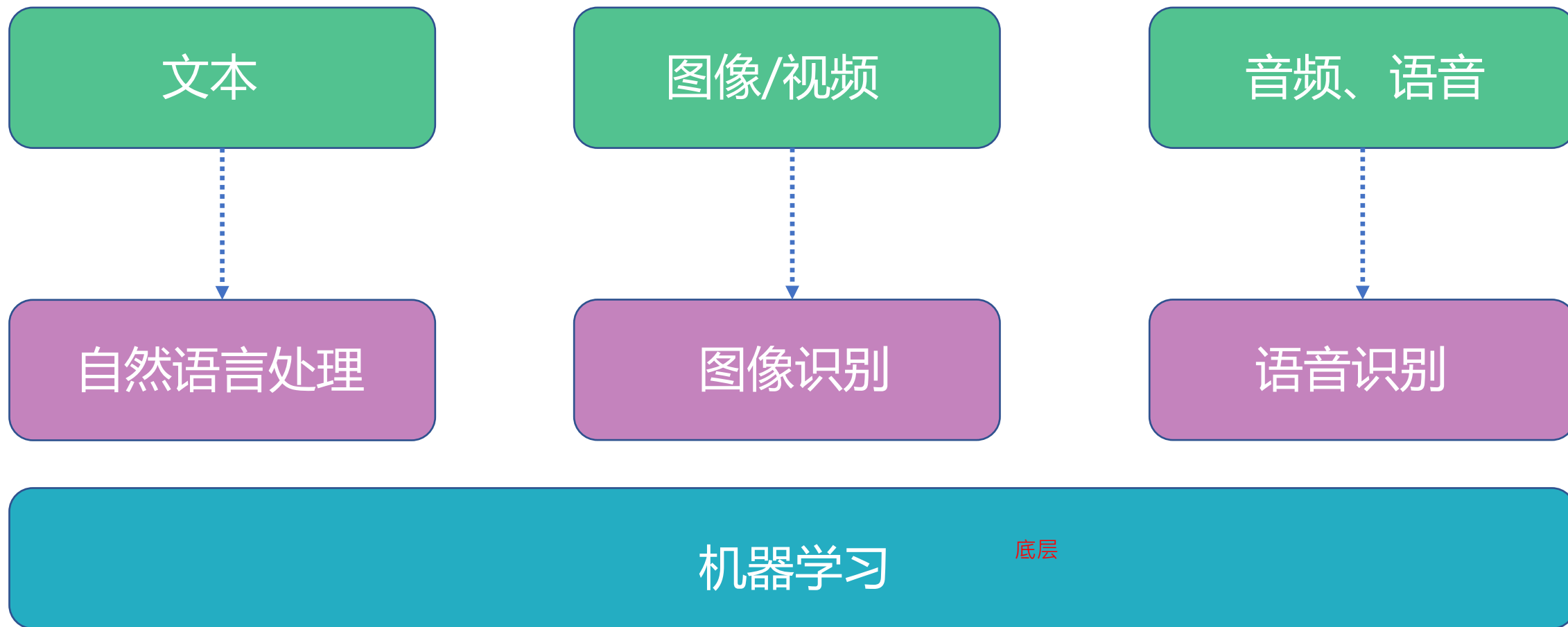
深度学习

函数的深度应用



人工智能、数据分析和大数据的关系







人工智能的颠覆性应用场景

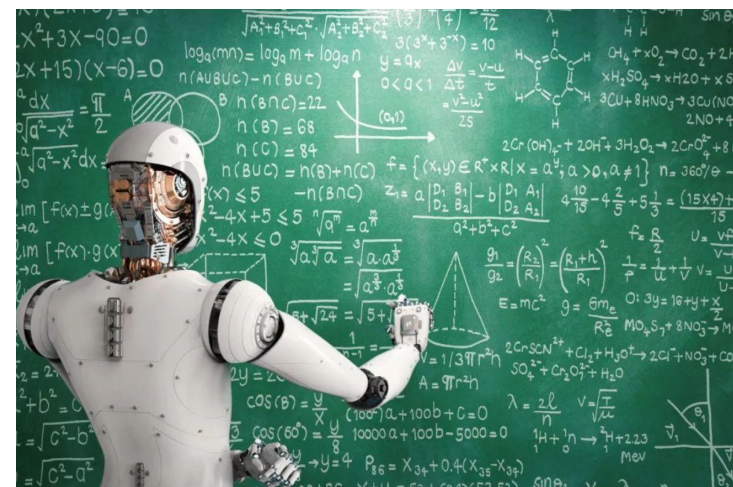
同声翻译



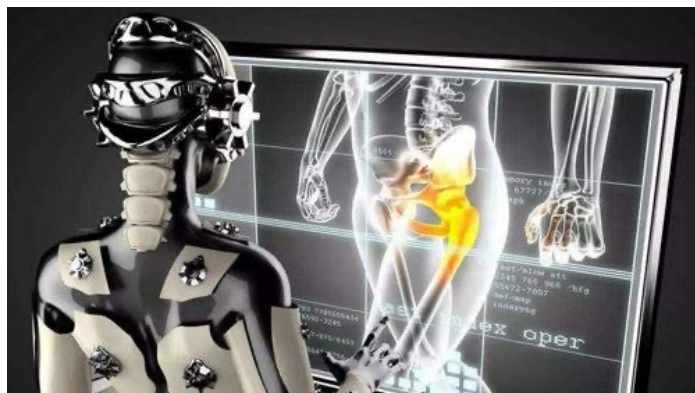
天网监控系统



AI老师



AI医生



AI厨师



AI设计师

鹿班^{BETA}

应用场景举例：

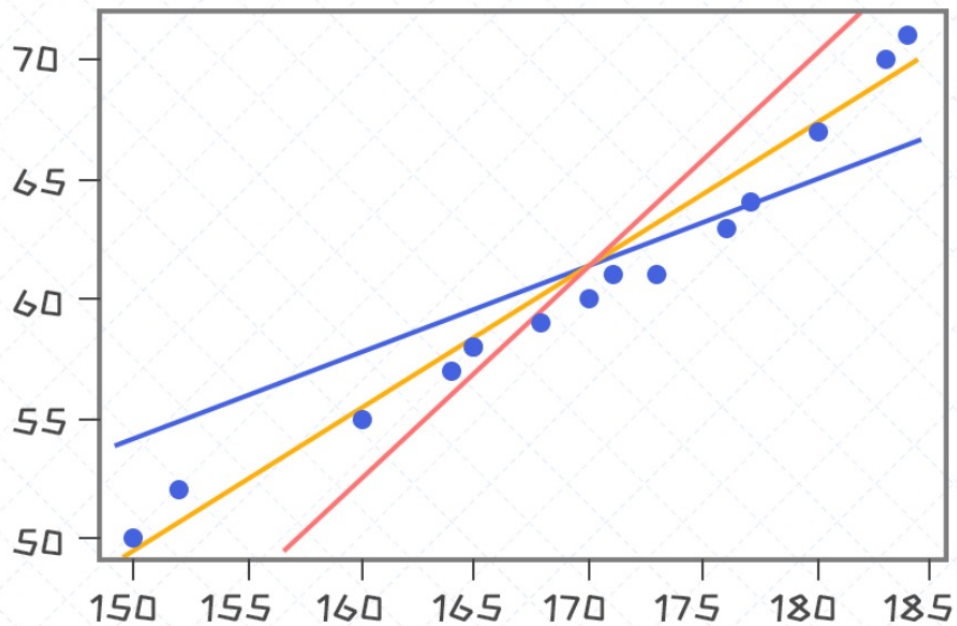
股价预测

营收预测

广告效果预测

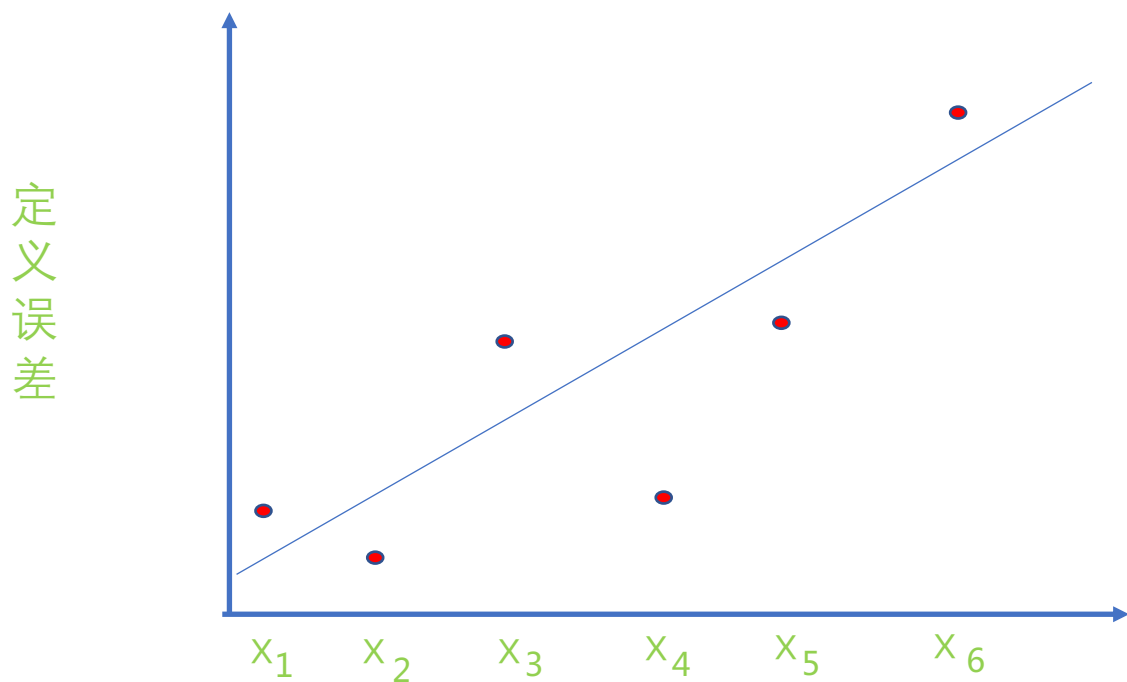
销售业绩预测

身高(cm)	体重(kg)
150	50
152	52
160	55
164	57
165	58
168	59
170	60
171	61
173	61
176	63
177	64
180	67
183	70
184	71



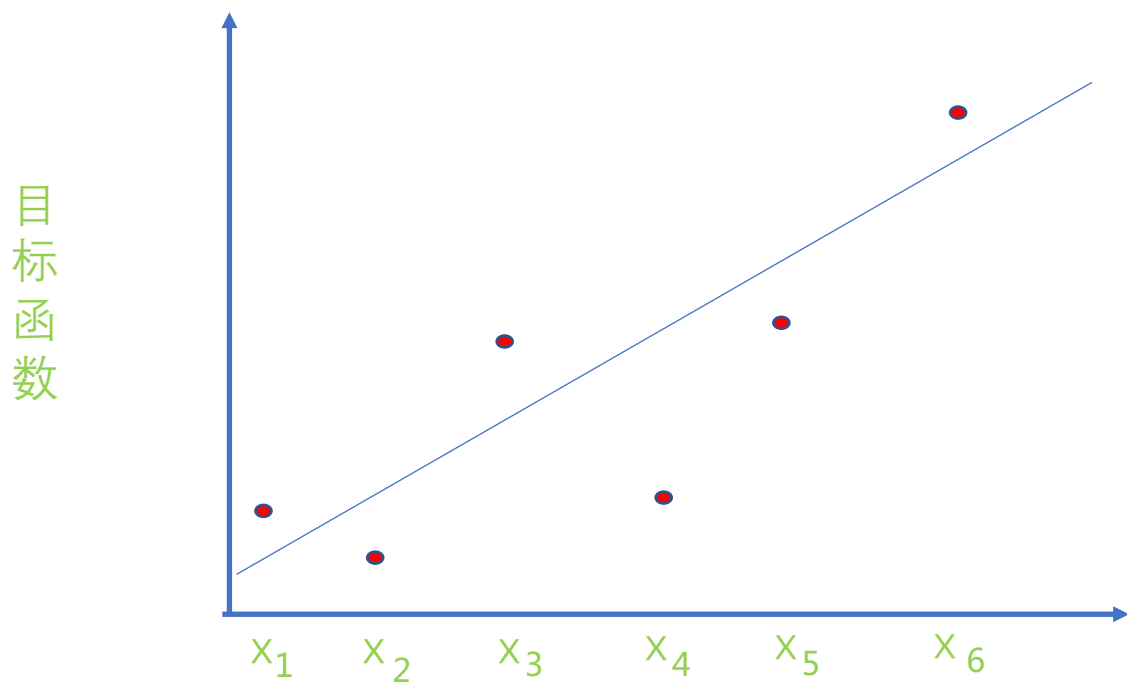
模型如何判定哪条线拟合程度最高？

一条线产生的误差较小,就可以认为这条线拟合度较高。



模型如何判定哪条线拟合程度最高？

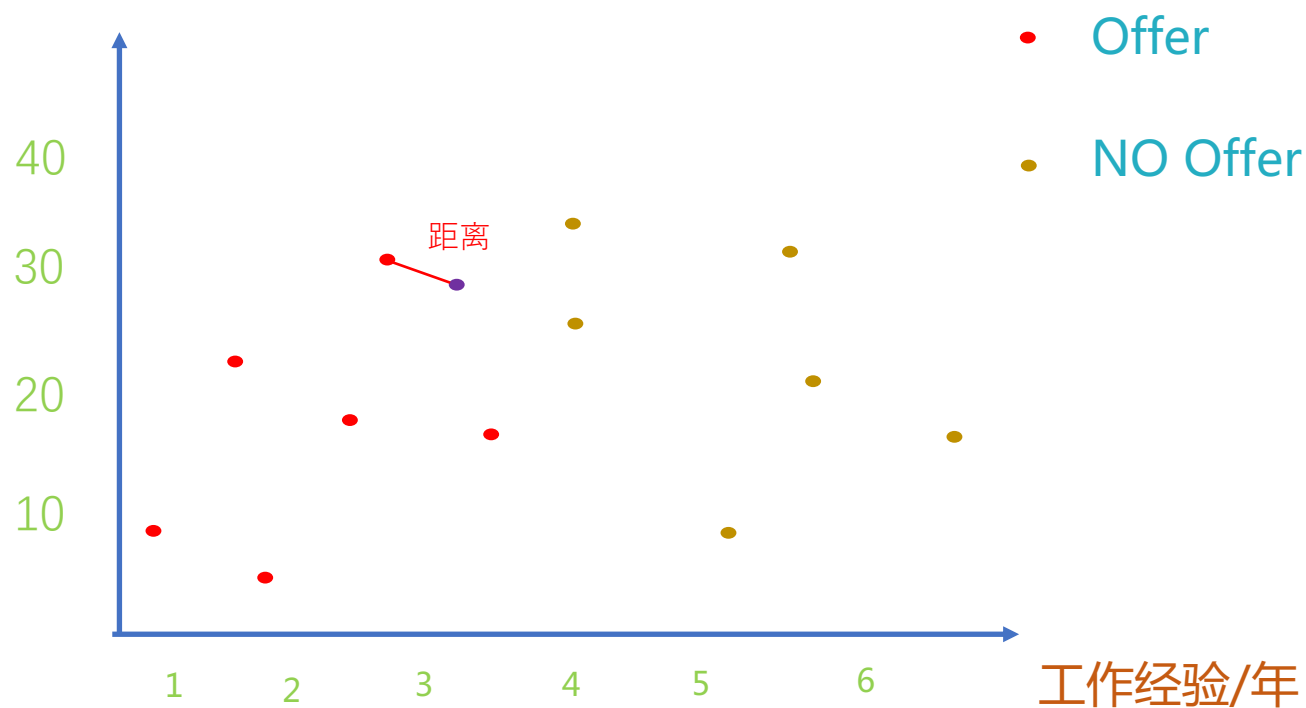
一条线产生的误差较小,就可以认为这条线拟合度较高。



算法核心思想：不标记样本的类别，由距离其最近的K个邻居投票来决定，所以K值的选择比较关键

算法原理：计算待标记的数据样本和数据集中每个样本的距离，取距离最近的k个样本。待标记的数据样本所属的类别，就由这k个距离最近的样本投票产生。

目前收入/千元



山鸢尾



虹膜锦葵

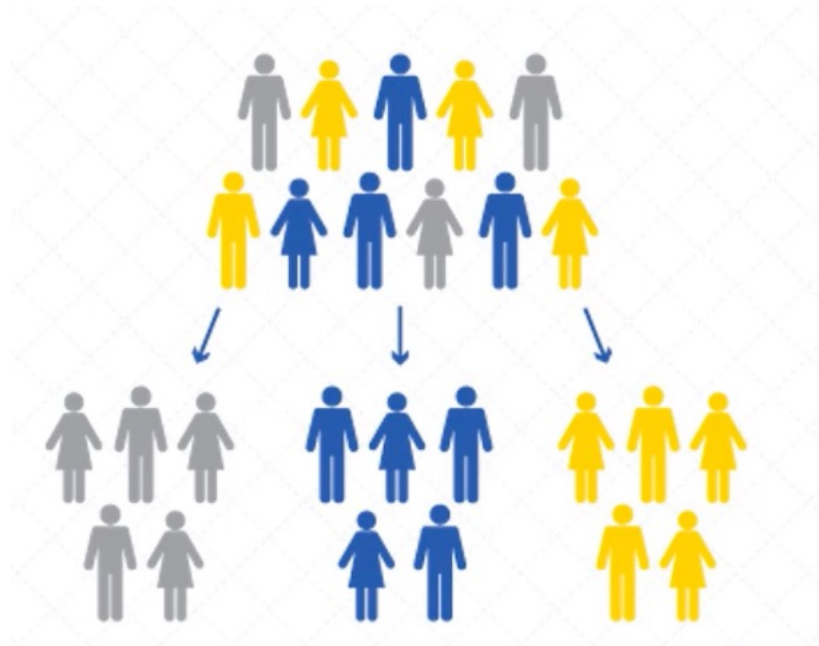


变色鸢尾



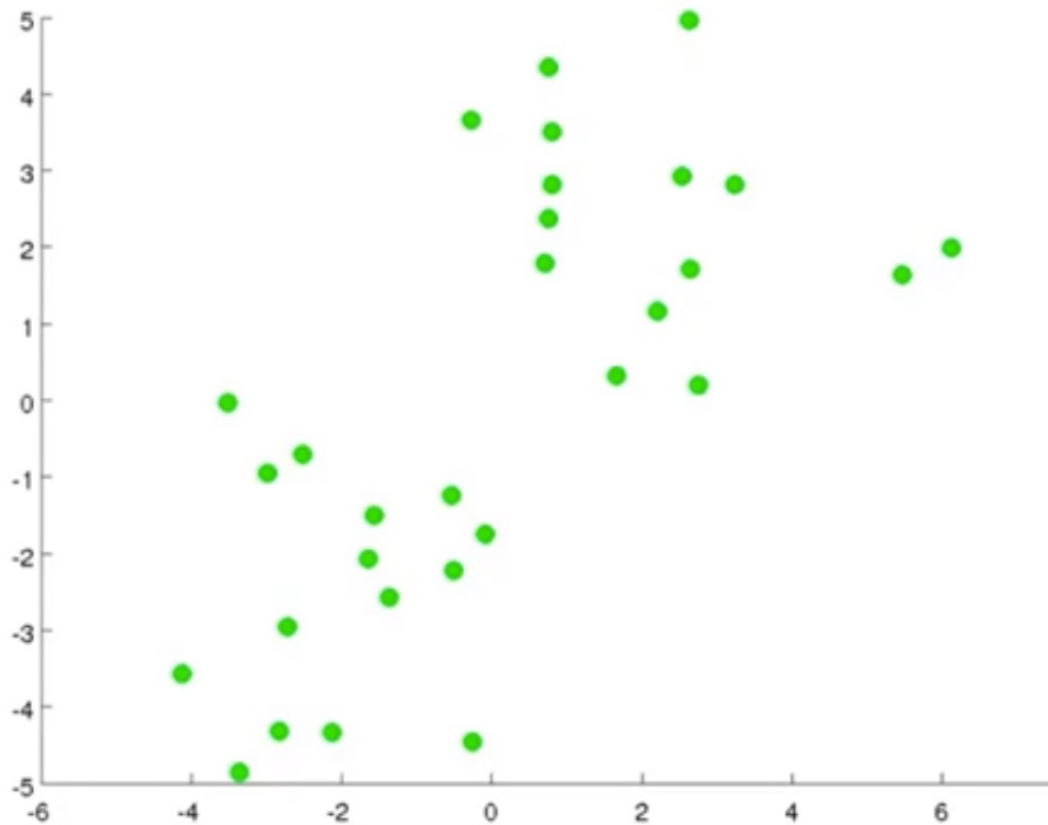
其中包含4个特征：萼片长度、萼片宽度、花瓣长度、花瓣宽度

个性化



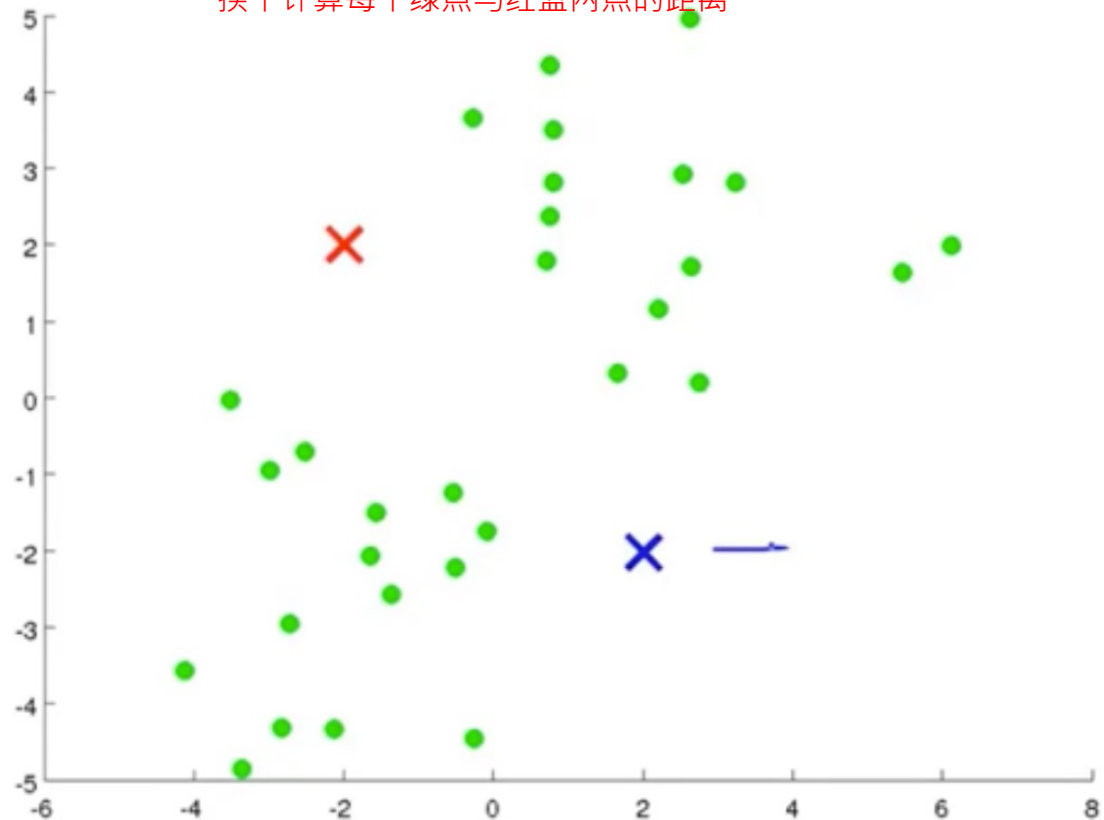


KMeans聚类算法



随机落两个红蓝两个点

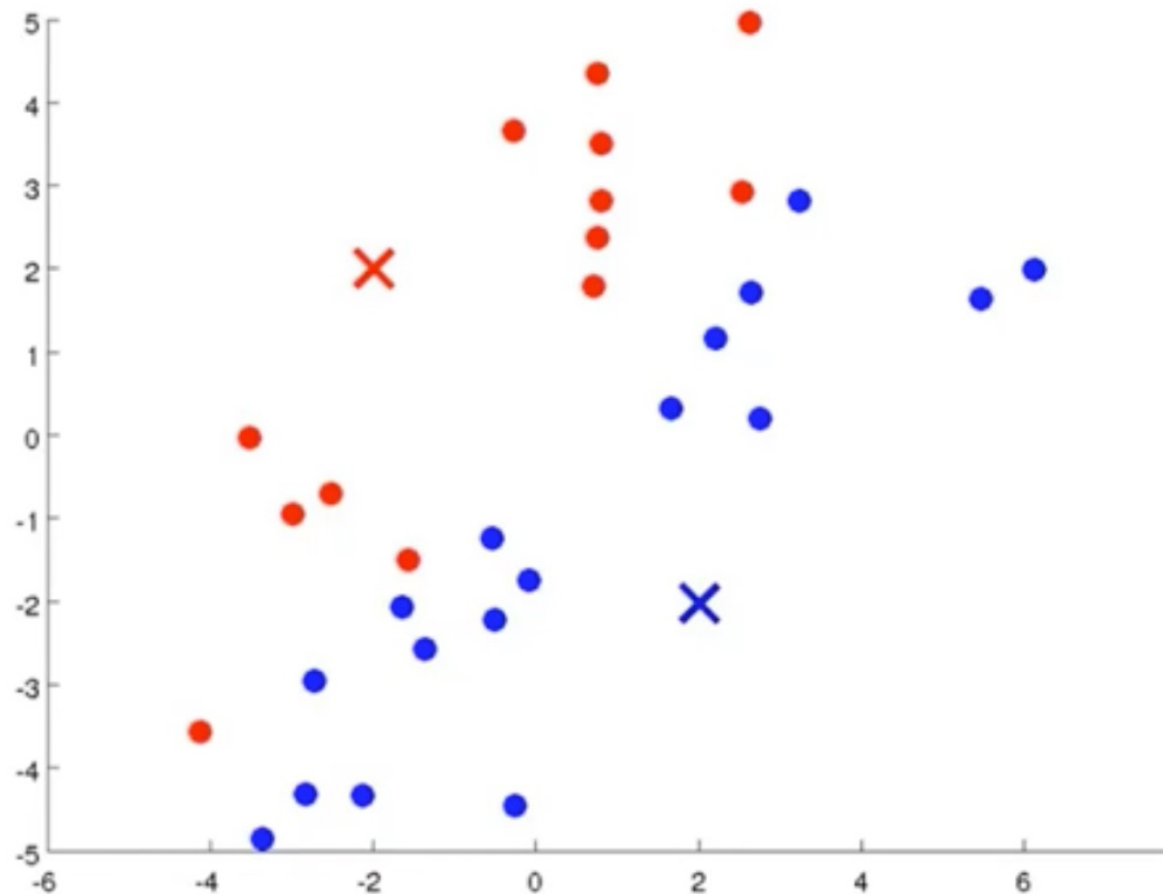
挨个计算每个绿点与红蓝两点的距离



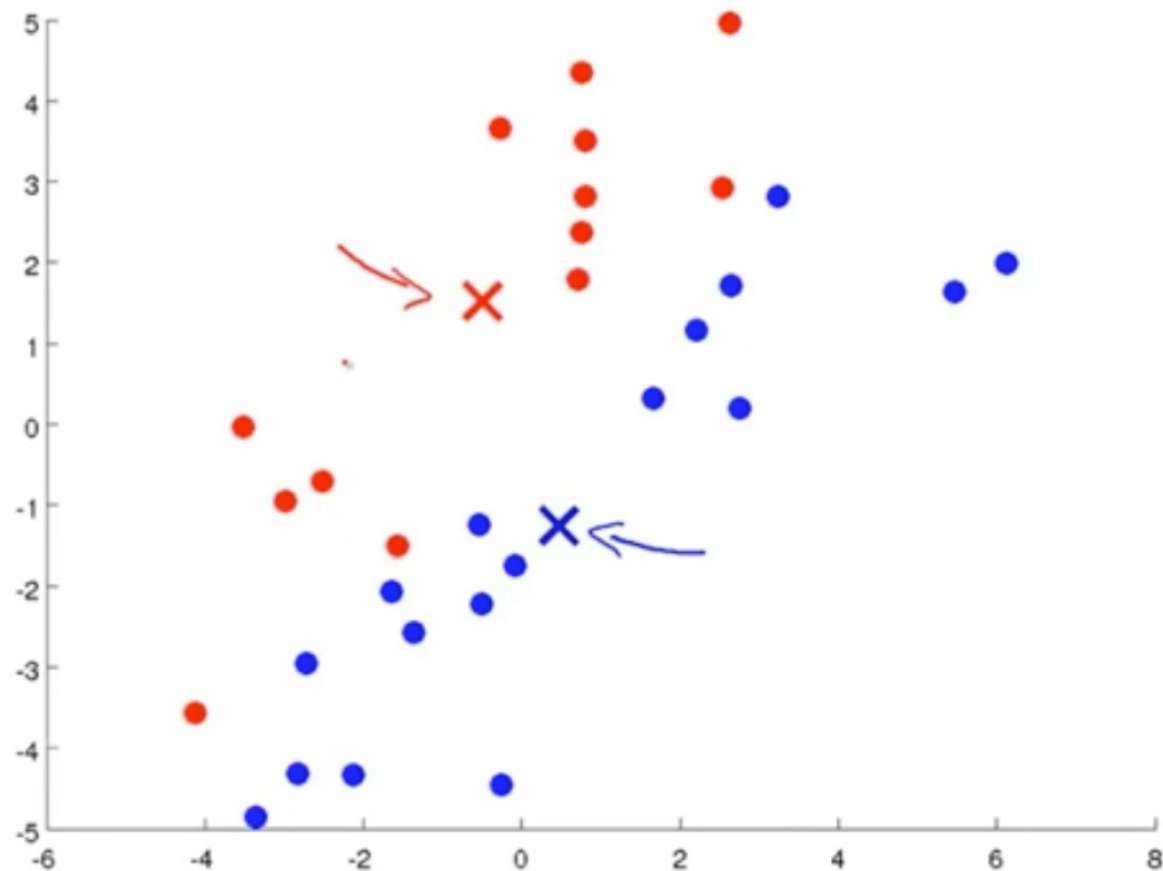
114



KMeans聚类算法



移动到红蓝各自的中心点，再次进行距离计算





KMeans聚类算法

循环次数，中心点移动距离（小于多少时停止）

