

# 智能显示场景下的机器学习入门

黄帅 / h00370663

2018 年 1 月 7 日

## 1 介绍

- 什么是学习

## 2 线性模型

- 线性回归
- 对数几率回归
- 性能度量

## 3 应用举例

- Kaggle 竞赛 - Titanic

## 4 学习材料

## 定义

A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ . (by *Mitchell*)

对于某类任务  $T$  和性能度量  $P$ , 如果一个计算机程序在  $T$  上以  $P$  衡量的性能随着经验  $E$  而自我完善, 那么我们称这个计算机程序在从经验  $E$  中学习。

- 任务  $T$ : 机器学习所关注的任务是人类很难通过总结具体规则进行解决的任务。
- 性能度量  $P$ : 为了可以量化地衡量机器学习算法的表现, 通常根据不同的任务类型设计相应的性能度量方式。
- 经验  $E$ : 根据学习过程中获取到的经验种类, 机器学习可以分为有监督学习和无监督学习两大类。

## 模型定义

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

其中， $\mathbf{x}$  是样本中提取出的特征向量； $\mathbf{w}$  和  $b$  是待学习参数。

“线性回归”(linear regression) 试图学得一个线性模型以尽可能准确地预测实值输出标记。

## 模型定义

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}} \quad (2)$$

其中， $\mathbf{x}$  是样本中提取出的特征向量； $\mathbf{w}$  和  $b$  是待学习参数。

对数几率回归又称 *Logistic Regression*，虽然名字中带着回归两个字，但它是一个分类模型。





