



# 机器学习之概率统计基础

AI100学院  
2017年7月

<http://www.ai100.ai/>

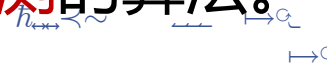
# ► 机器学习

- [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning) :



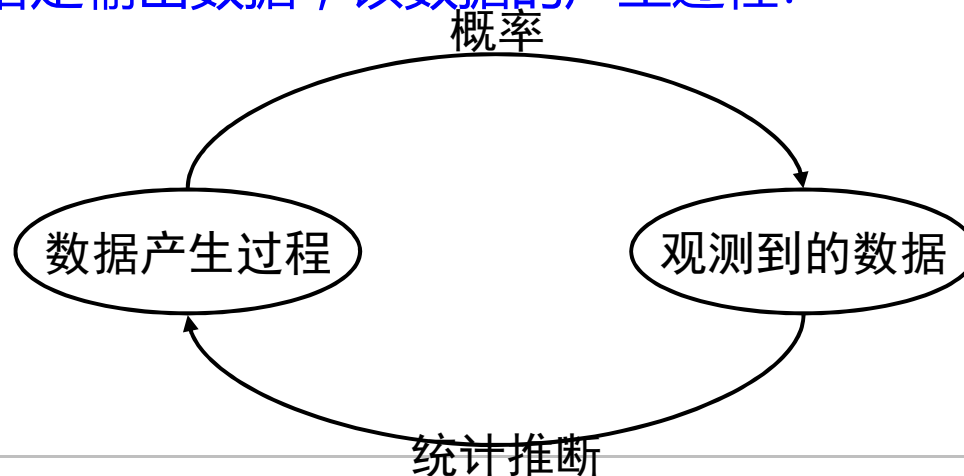
a branch of artificial intelligence, concerns the construction and study of systems that can **learn** from **data**."

- **机器学习**：机器学习理论主要是设计和分析一些让计算机可以自动“**学习**”的算法。机器学习算法是一类从**数据**中自动分析获得规律，并利用规律对未知数据进行**预测**的算法。



## ► 概率 vs. 统计

- 概率：研究随机事件出现的可能性的数学分支
  - 给定以一个数据产生过程，输出的性质？
- 统计推断：处理数据分析和概率理论的数学分支，与数据挖掘和机器学习是近亲
  - 给定输出数据，该数据的产生过程？



$$\vec{h} \leftarrow \sim \quad \vec{h} \rightarrow \vec{G} \quad \vec{h} \rightarrow \vec{G}$$

## ► 概率统计基础的重要性

- 研究数据分析必须打好概率和统计基础
  - Using fancy tools like neural nets, boosting and support vector machines without understanding basic statistics like **doing brain surgery before knowing how to use a band-aid.**



# ► 概率

- 概率基础
  - 概率公理及推论
- 随机变量及其分布： pmf/pdf、CDF、均值、方差...
- 常用随机变量分布
  - 离散型随机变量、连续性随机变量
- 多元随机向量
  - 联合概率、条件概率
  - 常用多元分布
  - 条件独立及Markov相关模型
- 概率密度估计
  - 参数估计/非参数估计/准参数估计



# ► 统计推断

- 大数定律
- 中心极限定理
- 蒙特卡洛近似
- 极大似然估计
  - 似然函数
  - 偏差-方差分解、Bootstrap
- 贝叶斯估计
  - 共轭先验、似然、后验、后验预测



## ► 参考书推荐

- Kevin P. Murphy, [Machine Learning: A Probabilistic Perspective](#), MIT Press, 2012
- Larry Wasserman, [All of Statistics: A Concise Course in Statistical Inference](#)
  - 中译本：《[统计学完全教程](#)》
  - 内容很全，但有些部分篇幅略少，更偏向于从统计的角度讲述
- 盛骤、谢式干、潘承毅，[概率论与数理统计](#)，高等教育出版社.



$$\vec{h} \leftrightarrow \sim \quad \vec{h} \rightarrow \vec{G} \quad \vec{h} \rightarrow \vec{G}$$

# THANK YOU

北京智百科技有限公司



$\hbar \leftrightarrow \sim$   $\rightarrow \mathcal{Q}$   
 $\rightarrow \mathcal{G}$