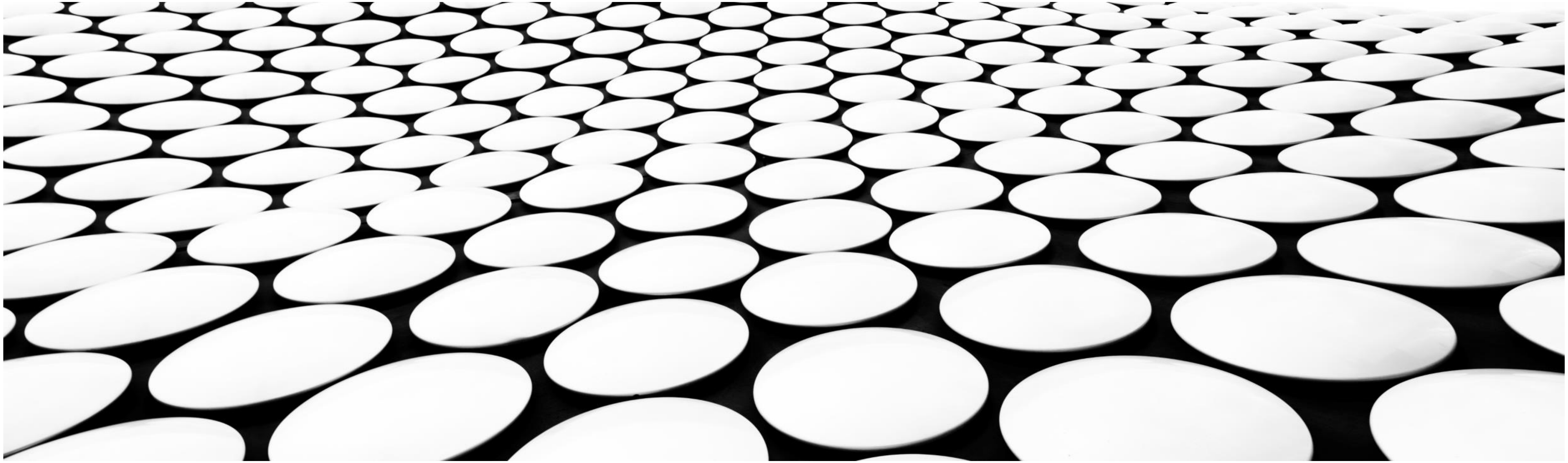
深度学习

邱怡轩



今天的主题

- 序章（续）：深度学习简史
- 机器学习问题
- 线性模型

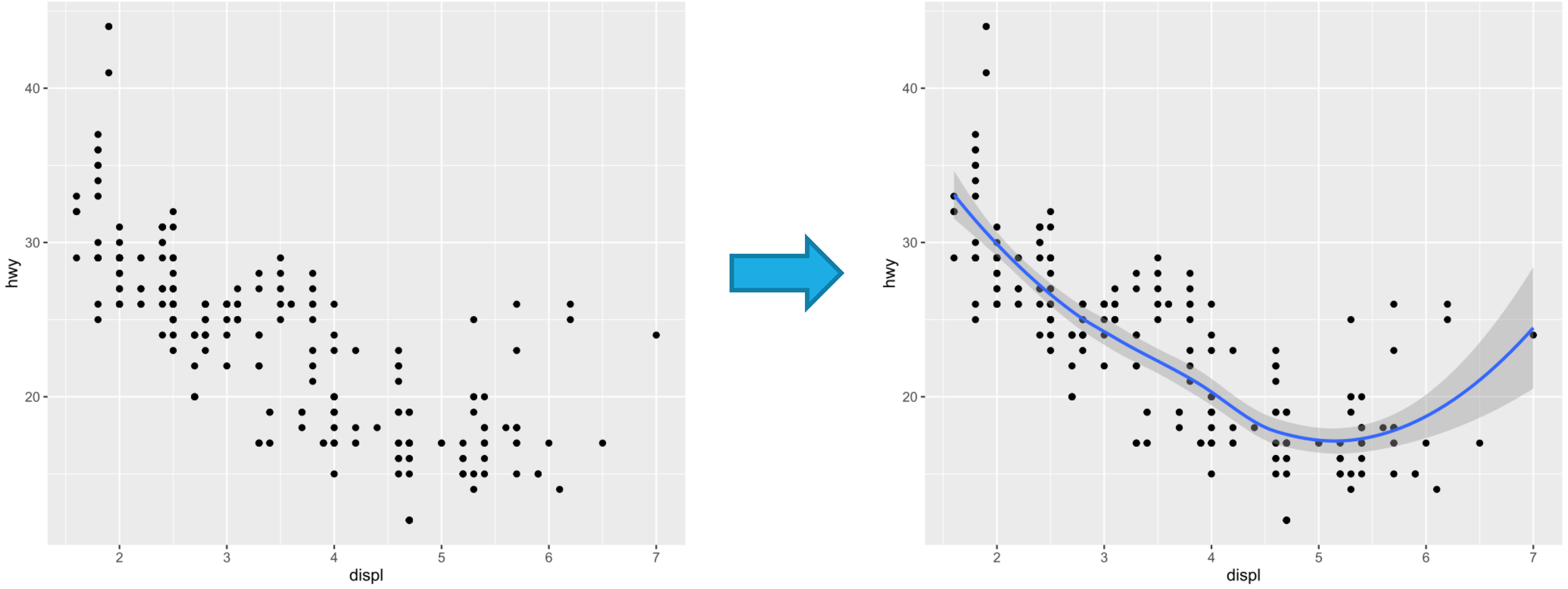
机器学习?

机器学习

- 深度学习 ⊂ 机器学习
- 机器学习的定义?

机器学习

- 构建变量之间的函数关系？



机器学习

- 构建变量之间的函数关系？

0

X

Y

机器学习

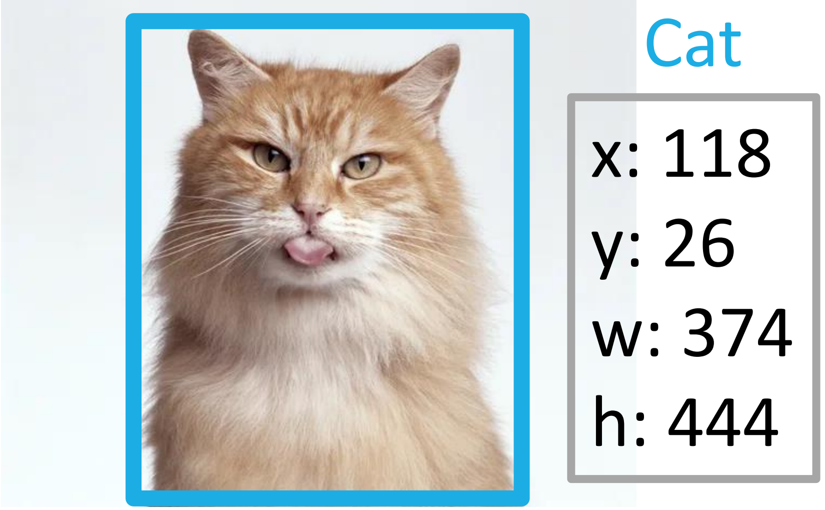
- 构建变量之间的函数关系？

$\Rightarrow$ Cat

X Y

机器学习

- 构建变量之间的函数关系？

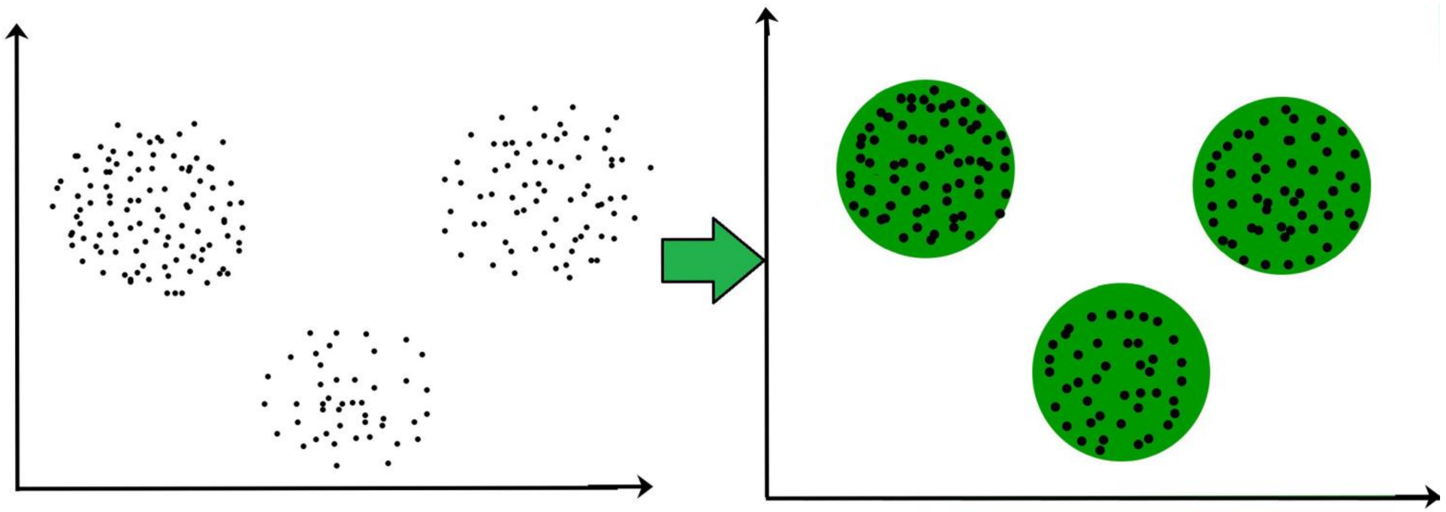


X Y

似乎很合理?

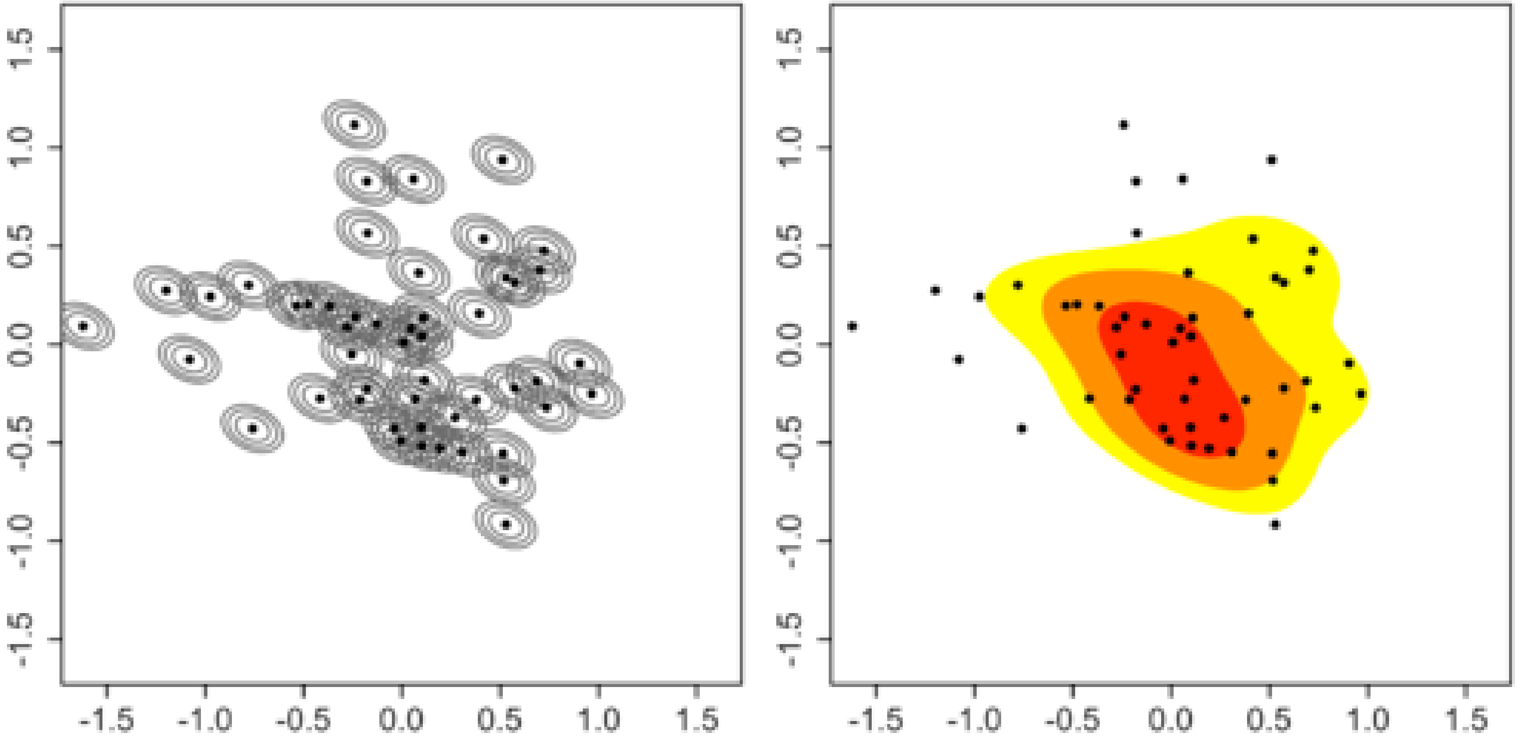
机器学习

- 按照该定义，聚类算不算机器学习？



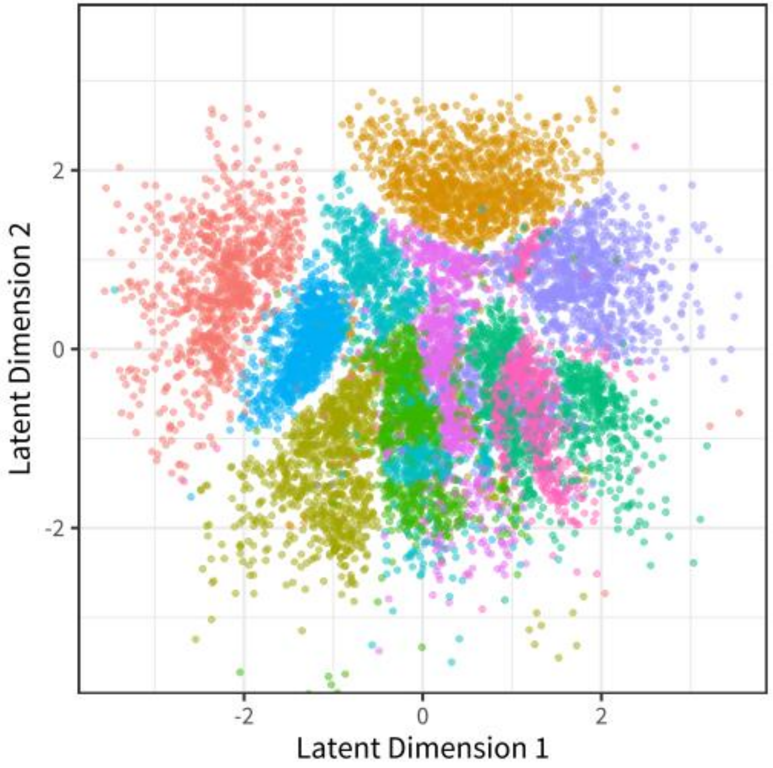
机器学习

- 密度估计算不算机器学习?



机器学习

- 降维算不算机器学习?



机器学习

- 我们似乎只能用一个很模糊的描述：
- 从数据中学习规律
- 这是否就是统计学？

机器学习

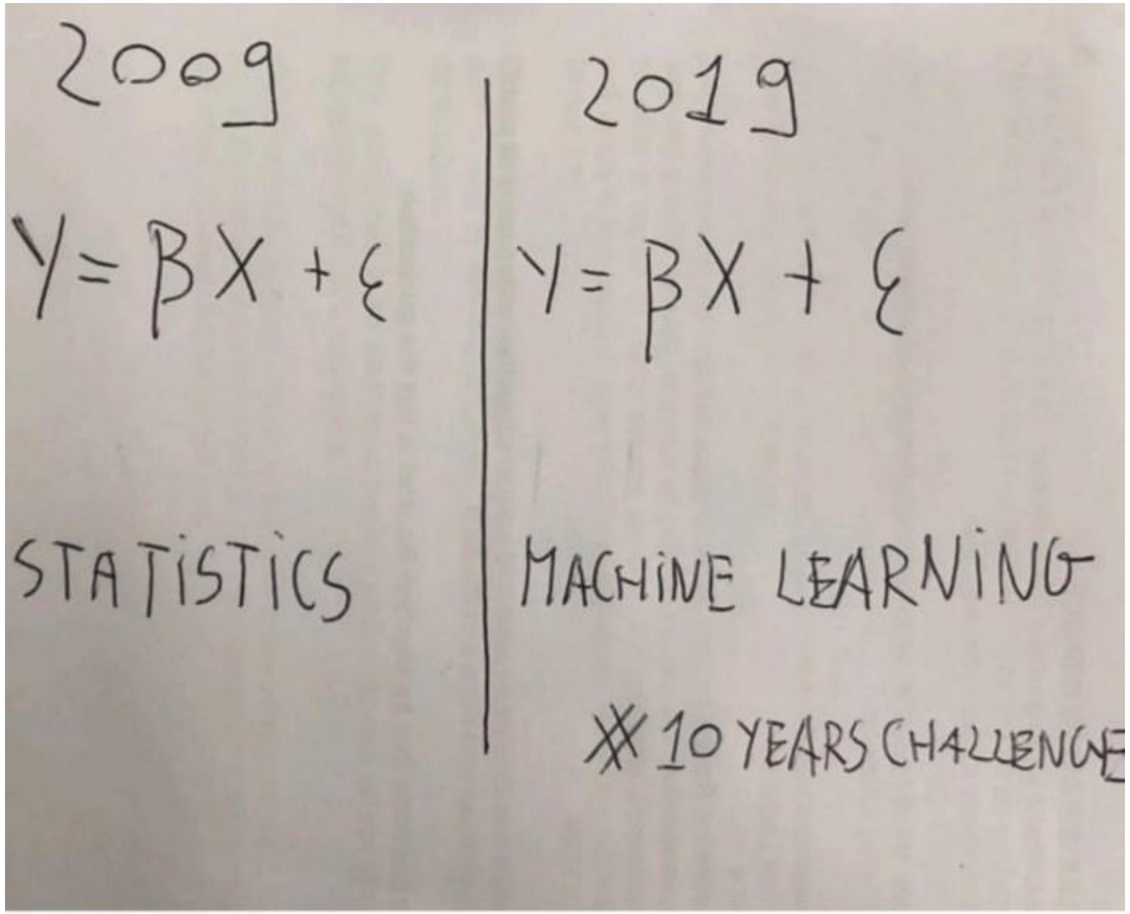


Image: https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3

机器学习

- 很大一类机器学习模型可以称为统计机器学习
- 核心任务是找到数据的生成规律
- 或者说是数据的统计分布

机器学习

- 回归

- 研究 $p(y|x)$，Y 是连续型随机变量
- 如假定 $Y|x \sim N(\beta' x, \sigma^2)$

机器学习

- 二分类

- 研究 $p(y|x)$，Y 服从 Bernoulli 分布
- 如假定 $Y|x \sim Bernoulli(\mathrm{sigmoid}(\beta' x))$

机器学习

- 多分类

- 研究 $p(y|x)$，Y 服从多项分布
- 如假定 $Y|x \sim Multinom(\mathrm{softmax}(Wx))$

机器学习

- 密度估计
- 研究 $p(x)$

机器学习

- 聚类

- 数据 X 服从混合分布
- 每个 X 对应一个离散潜变量 $C \in \{1, \ldots, K\}$
- 研究 $p(c|x)$

机器学习

- 降维

- 研究 $p(z|x)$，Z 是比 X 低维的随机变量

基础设施

构成要素

- 要完成一项机器学习任务，通常都要有以下几个部分

- 数据
- 模型
- 损失函数
- 优化算法

数据

- 数据的大小和质量往往决定了模型效果的上限

- 巧妇难为无米之炊

模型

- 描述数据是如何生成的
- 允许有一部分信息是未知的（参数）

- 例：密度估计，$X \sim N(\mu, \sigma^2)$
- 例：回归，$Y = \beta' x + \varepsilon$

损失函数

- 因为有未知参数，所以可能有无限多可能的模型
- 损失函数用来评价模型的好坏
- 例：最小二乘准则
- 例：极大似然准则
- 后面还将更深入讨论

优化算法

- 根据损失函数制定的目标，寻找“最优”的参数值
- 例：梯度上升/下降
- 路漫漫其修远兮，吾将上下而求索。

小结

- 数据是你的粮草供应
- 模型是你的知识储备
- 损失函数是你的目标蓝图
- 优化算法是你的行动指南

小结

- 机器学习的创新和进步，很大程度上也是由这四项基本元素推动的

- 更丰富的数据类型，更海量的数据
- 更具表达力的模型
- 更稳健、更具泛化能力的损失函数
- 更高效、收敛更快的优化算法

从简单入手

- 我们先从最简单的问题入手

- 假设数据充分
- 模型为线性
- 暂不考虑优化问题
- 重点讨论损失函数如何选取

线性模型

线性回归

- 给定模型 $Y = \beta' x + \varepsilon$
- 给定数据 $(x_i, Y_i),\ i = 1, \dots, n$

- 如何估计 β ?

OLS

- 最小二乘几乎成为了做回归的“铁律”
- $L(\beta) = n^{-1} \sum_{i=1}^{n} (Y_i - \beta' x_i)^2$

- 但多想一步：
- 什么时候 OLS 具有好的性质?
- 如何向更复杂的问题扩展?

极大似然

- 统计中一个经典的估计方法是极大似然
- 当我们有了数据 $X_1, \ldots, X_n$ 以及其分布的模型 $p_\theta(x)$
- 定义似然函数 $l(\theta; x) = \log p_\theta(x)$
- 极大似然准则 $\max_\theta \; n^{-1} \sum_{i=1}^{n} l(\theta; X_i)$

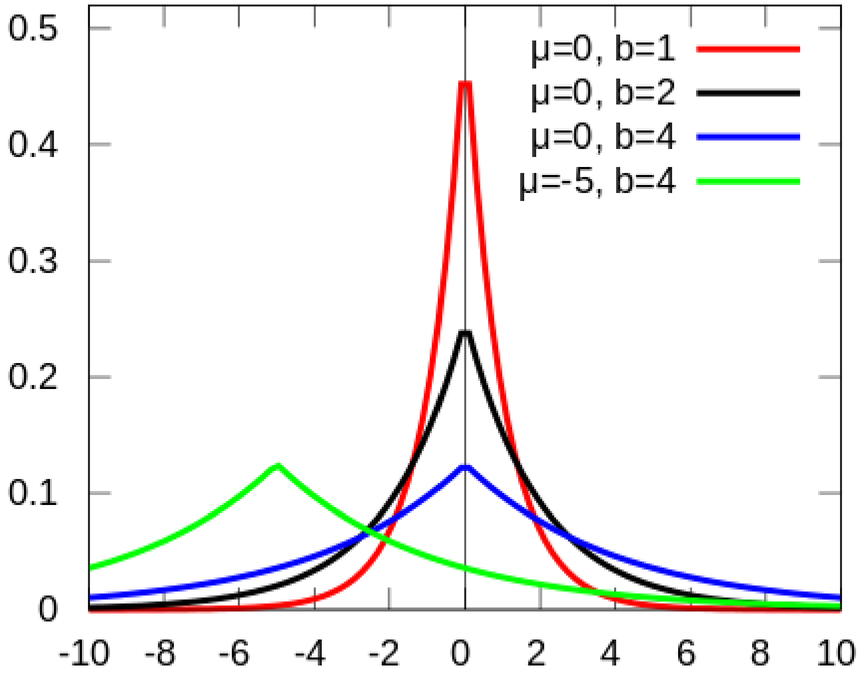
线性回归

- 在线性回归中，如果假定误差是正态
- $Y|x \sim N(\beta' x, \sigma^2)$
- $\log p(y|x) = -\frac{1}{2\sigma^2}(y - \beta' x)^2 + C$
- 极大似然准则可以导出 OLS

练习

- 如果误差服从 Laplace 分布

$$p(x;b)=\frac{1}{2b}e^{-|x|/b}$$



- 极大似然准则可以导出什么损失函数?

损失函数

- 似然函数可能比你想象中的要重要得多
- 一个被经常忽略的事实是，机器学习模型中很大一部分的损失函数都可以由似然函数导出
- 当你不知道选择什么损失函数的时候，先试试看极大似然准则

线性二分类

- 给定模型 $Y|x \sim Bernoulli(\rho(\beta' x))$
- $\rho(\cdot)$ 是把 $\beta' x$ 映射到 $(0,1)$ 上的一个给定的函数
- $\rho(\beta' x)$ 代表 Y 取1的概率
- 数据 $(x_i, Y_i),\ i = 1, \ldots, n$

Logistic回归

- 当 $\rho(x) = e^x/(e^x + 1)$ 时
- 模型即为 Logistic 回归

- 利用极大似然准则，Bernoulli 分布的似然函数为
  $$l(p; y) = y \log p + (1 - y) \log(1 - p)$$
- 其中 $p = \rho(\beta' x)$

练习

- 根据前述信息，写出 Logistic 回归关于 β 的损失函数
- 注意尽可能化简表达式

思考题

- 除了 Logistic 回归，你还能想出哪些 ρ 的构造方式?
- 它们对模型的最终结果影响大吗?