

6. 逻辑回归

WU Xiaokun 吴晓堃

xkun.wu [at] gmail

2022/04/01

生成与判别

逻辑回归

逻辑回归 **logistic regression**: 线性模型中最基础的分类器

- 通常作为分类任务中监督学习的基准方法

逻辑回归

逻辑回归 **logistic regression**: 线性模型中最基础的分类器

- 通常作为分类任务中监督学习的基准方法
- 神经网络的构建基础
 - 神经网络可以看成一系列逻辑回归分类器的**串联**

逻辑回归

逻辑回归 **logistic regression**: 线性模型中最基础的分类器

- 通常作为分类任务中监督学习的基准方法
- 神经网络的构建基础
 - 神经网络可以看成一系列逻辑回归分类器的**串联**

注意命名混淆: 逻辑回归是分类器

生成、判别式分类器

朴素贝叶斯是典型的生成式 **generative** 分类器

- 解决的问题：输入数据如何被生成

逻辑回归是典型的判别式 **discriminative** 分类器

- 解决的问题：哪些特征对辨别类别最有用

生成、判别式分类器

朴素贝叶斯是典型的生成式 **generative** 分类器

- 解决的问题：输入数据如何被生成

逻辑回归是典型的判别式 **discriminative** 分类器

- 解决的问题：哪些特征对辨别类别最有用

考虑猫狗分类问题

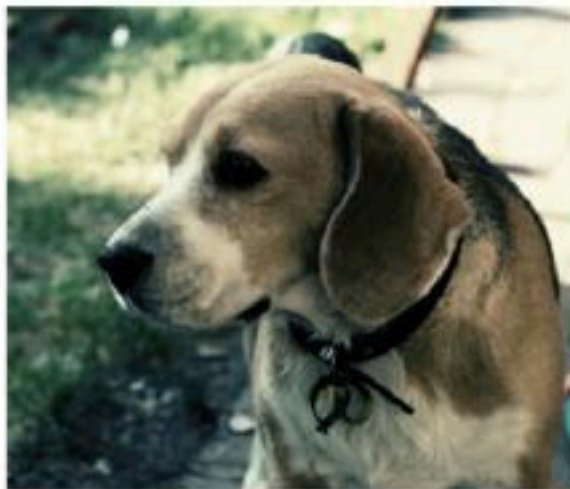


生成式分类器

首先构造一个猫分类器：判断图像是猫的概率

- 目标：对任意图像输出类别正确的概率
 - 训练出模型：知道从类别生成图像的方法

同样再构造一个狗分类器



生成式分类器

首先构造一个猫分类器：判断图像是猫的概率

- 目标：对任意图像输出类别正确的概率
 - 训练出模型：知道从类别生成图像的方法

同样再构造一个狗分类器



分类判定：两个模型都运行，比较哪个概率值更高

判别式分类器

将狗的图像从猫的图像区别开

- 比如狗通常带项圈，耳朵下垂，脸比较长等



- 目标：将不同类别的图像区分开
 - 训练出模型：知道类别中的实例有哪些特征

生成模型、判别模型

判别 **discriminative** 模型：解决哪些特征对辨别类别最有用问题，如逻辑回归

$$\hat{c} = \arg \max_{c \in C} \overbrace{P(c|d)}^{\text{posterior}}$$

- 优化目标：观测数据对应的类别，与任务目标一致
 - 参数：描述文档本身（的特征分解形式）

生成模型、判别模型

判别 **discriminative** 模型：解决哪些特征对辨别类别最有用问题，如逻辑回归

$$\hat{c} = \arg \max_{c \in C} \overbrace{P(c|d)}^{\text{posterior}}$$

- 优化目标：观测数据对应的**类别**，与任务目标一致
 - 参数：描述文档本身（的特征分解形式）
- 通常**效能更佳**，如现代深度神经网络（前半部分是特征提取器）

生成模型、判别模型

判别 **discriminative** 模型：解决哪些特征对辨别类别最有用问题，如逻辑回归

$$\hat{c} = \arg \max_{c \in C} \overbrace{P(c|d)}^{\text{posterior}}$$

- 优化目标：观测数据对应的类别，与任务目标一致
 - 参数：描述文档本身（的特征分解形式）
- 通常效能更佳，如现代深度神经网络（前半部分是特征提取器）

生成 **generative** 模型：解决输入数据如何被生成的问题，如朴素贝叶斯

$$\hat{c} = \arg \max_{c \in C} \overbrace{P(d|c)}^{\text{likelihood}} \overbrace{P(c)}^{\text{prior}}$$

只有两类模型：思考角度不同

训练数据对 (x, y) ：随机变量不分先后；统计学不讨论因果关系

- 判别模型： $y = f(x; \theta_x)$ ，参数 θ_x 描述 x ，因此可以直接预测 \hat{y}

$$\hat{y} = \arg \max_{y \in Y} \overbrace{P(y|x; \theta_x)}^{\text{posterior}}$$

只有两类模型：思考角度不同

训练数据对 (x, y) ：随机变量不分先后；统计学不讨论因果关系

- 判别模型： $y = f(x; \theta_x)$ ，参数 θ_x 描述 x ，因此可以直接预测 \hat{y}

$$\hat{y} = \arg \max_{y \in Y} \overbrace{P(y|x; \theta_x)}^{\text{posterior}}$$

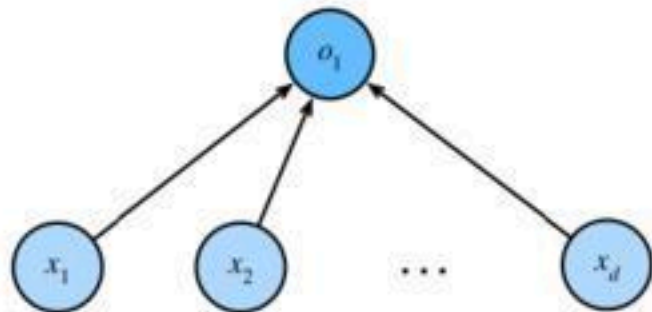
- 生成模型： $x = f^{-1}(y; \theta_x)$ ，不能直接预测 \hat{y} ，转而计算 x 的（条件）概率

$$\begin{aligned} \hat{y} &= \arg \max_{y \in Y} \overbrace{P(y|x)}^{\text{posterior}} \\ &= \arg \max_{y \in Y} \overbrace{P(x|y; \theta_x)}^{\text{likelihood}} \overbrace{P(y)}^{\text{prior}} \end{aligned}$$

逻辑回归分类器

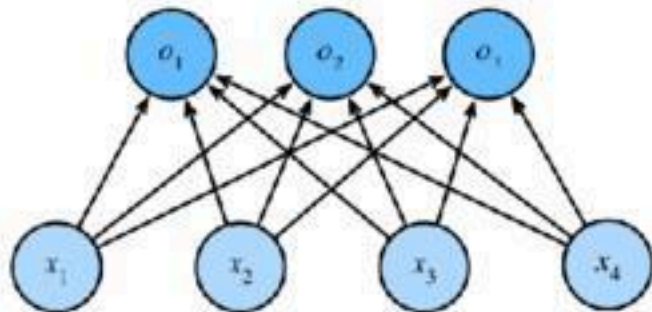
从单类到多类

单类



- 输出转换: sigmoid

多类



- 输出转换: softmax

Review

Feature selection: how & why

Feature engineering.

- correlation: weights.