朴素贝叶斯算法
Naïve Bayes Algorithms

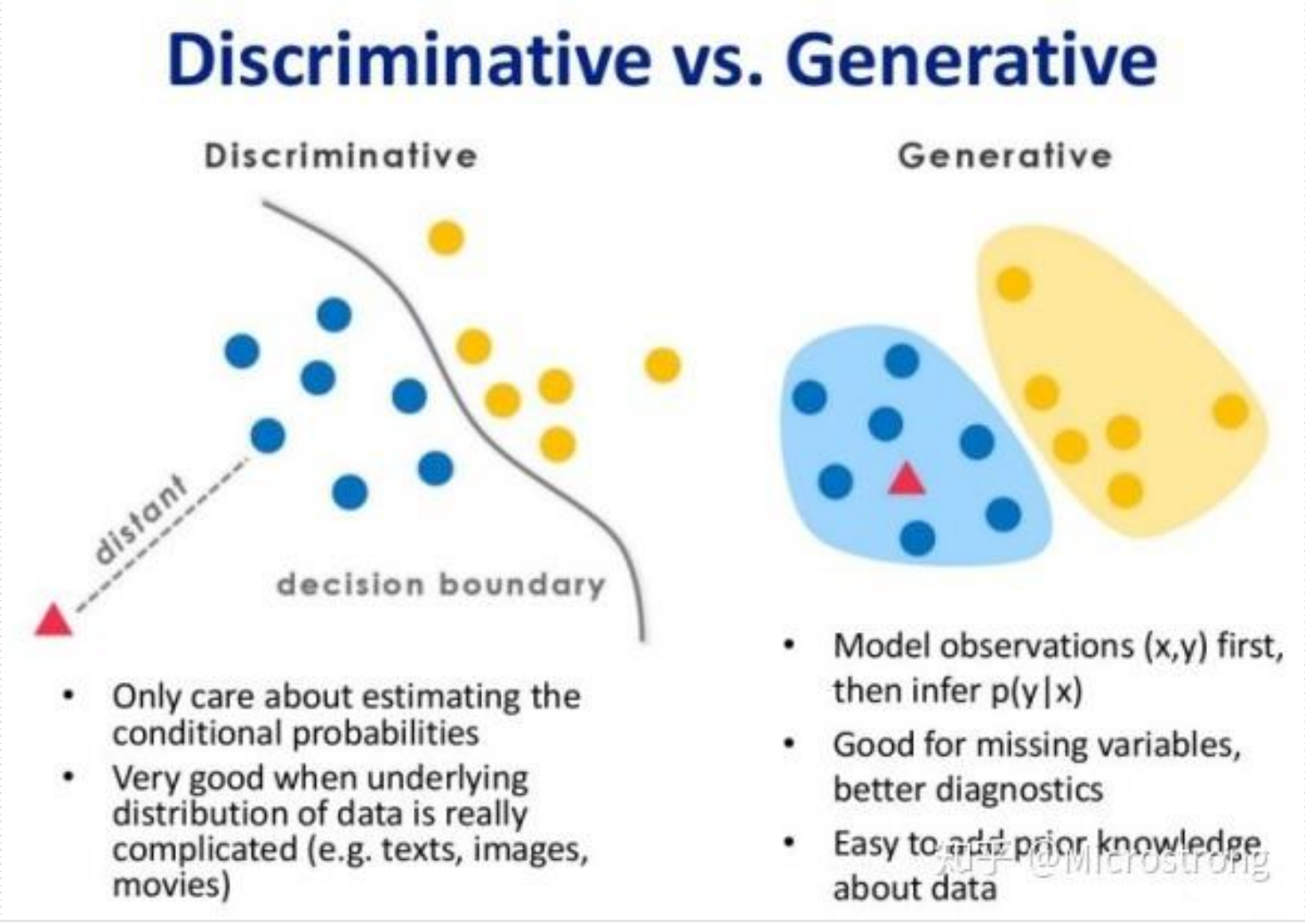
目录

1.1 朴素贝叶斯法的学习与分类

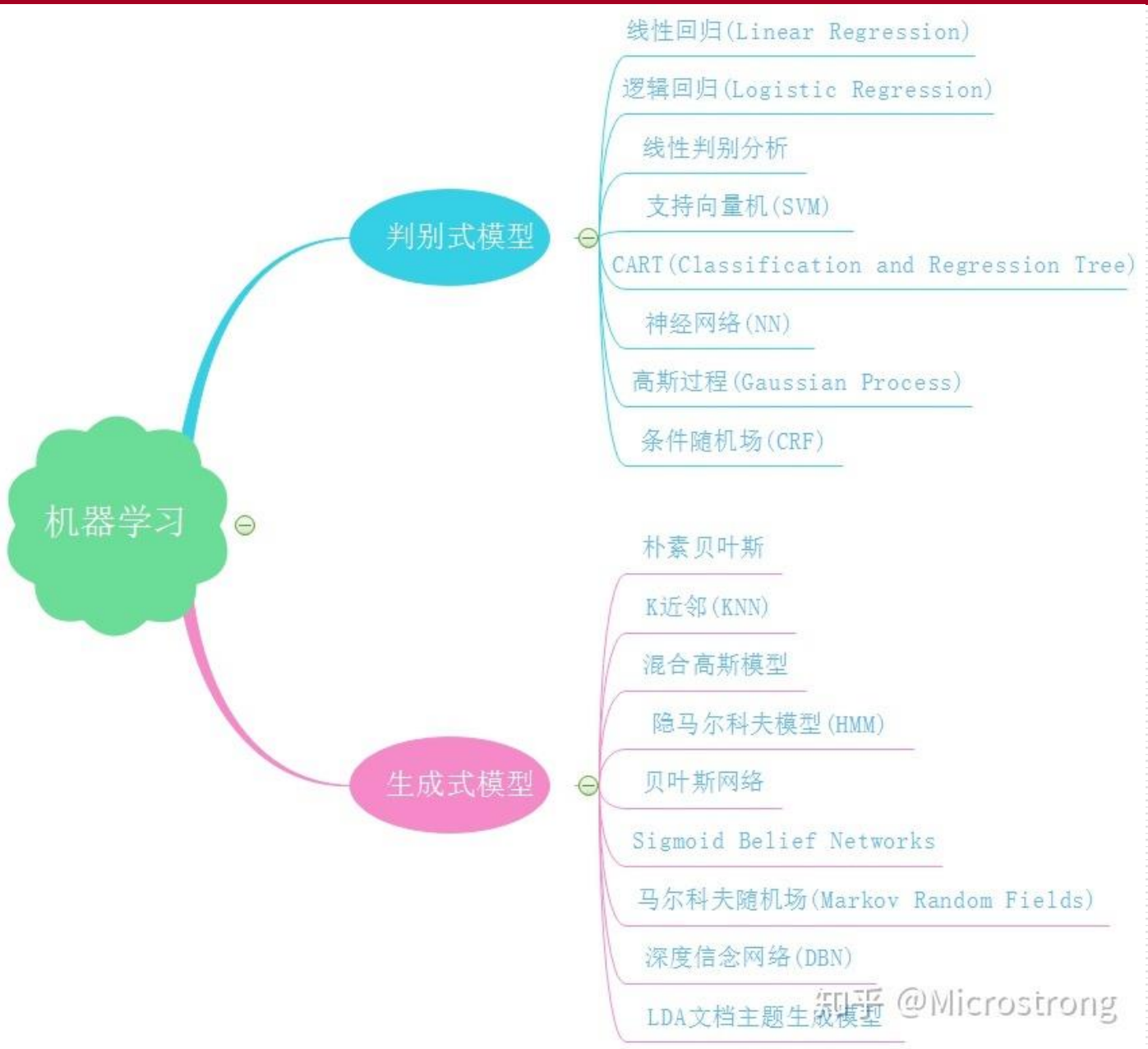
考虑一个分类问题，我们希望根据动物的某些特征（$X = (x_1, x_2, \ldots, x_n)$）来区分猫（y = 1）和狗（y = 0）。

- **判别模型**
  - 找到将猫和狗分开的决策边界或分类原则。
  - 为了分类一只新动物，判别模型会检查它落在决策边界的哪一边，并直接做出决定。
  - 直接估计后验概率 $p(y|x)$ 。
- **生成模型**
  - 分别学习猫和狗的特征模型。
  - 要对新动物进行分类，将其与猫/狗模型进行匹配，并查看它看起来更像哪个模型。
  - 估计先验概率 $p(y)$ 和条件概率 $p(x|y)$，根据贝叶斯定理计算后验概率 $p(y|x)$ 。

1.1 朴素贝叶斯法的学习与分类



1.1 朴素贝叶斯法的学习与分类



1.1 朴素贝叶斯法的学习与分类

朴素贝叶斯

思想：**朴素贝叶斯假设，又称条件独立性假设**

对于特征$X=(x_1,x_2,\dots,x_n)$，满足$x_i \perp x_j \mid y\ (i \neq j)$

$$p(X\mid y)=p(x_1,x_2,\dots,x_n\mid y)=\prod_{j=1}^{n} p(j\mid y)$$

Motivation：简化运算

条件独立假设，用于分类的特征在分类模型确定的条件下是条件独立的。

1.1 朴素贝叶斯法的学习与分类

朴素贝叶斯法

思想：**朴素贝叶斯假设，又称条件独立性假设**

做法：根据**贝叶斯定理**来估计每个类别的**后验概率**。

$$p(y|x)=\frac{p(x,y)}{p(x)}=\frac{p(x|y)p(y)}{p(x)}=\frac{p(x|y)p(y)}{\sum_i p(x|y_i)p(y_i)}\propto p(x|y)p(y)$$

朴素贝叶斯法的目标是找到

$$y=\arg\max_y p(y|x)=\arg\max_y \frac{p(x,\ y)}{p(x)}=\arg\max_y p(x|y)p(y)$$

1.1 朴素贝叶斯法的学习与分类

给定一个包含 M 个文本的数据集，其中每个有 K 维特征向量 $X = (x_1, \dots, x_K)$ 和一个情感标签 e_i，为了预测测试文本，需要估计：

$$\begin{aligned}
\arg\max_{e_i} p(e_i|X) &= \arg\max_{e_i} \frac{P(X|e_i)p(e_i)}{p(X)} \\
&= \arg\max_{e_i} p(X|e_i)p(e_i) \\
&= \arg\max_{e_i} \prod_{k=1}^{K} p(x_k|e_i)p(e_i) \\
&= \arg\max_{e_i} \sum_{j=1}^{M} \prod_{k=1}^{K} p(x_k|e_i, d_j)p(d_j, e_i)
\end{aligned}$$

1.2 朴素贝叶斯法文本分类

数据处理

假设现在有一个文本："*Step by step, we succeed*"。

X	step	by	we	succeed	joy	sad
onehot	1	1	1	1	0.9	0.1
TF	0.4	0.2	0.2	0.2	0.9	0.1
TF-IDF	1.03	0.7	0.6	1.16	0.9	0.1

x_k 表示文本 d_j 中的第 k 个词，文本 d_j 的情感标签为 e_i。

为了保证 $\sum_{k=1}^{K} p(x_k|d_j, e_i) = 1$，需要对文本特征归一化至[0, 1]，这样 $p(x_k|d_j, e_i)$ 就处于 0 到 1 之间。具体归一化方法为：

$$p(x_k|d_j, e_i) = \frac{x_k}{\sum_{k=1}^{K} x_k}$$

1.2 朴素贝叶斯法文本分类

数据处理

假设现在有一个文本："*Step by step, we succeed*"。

X	step	by	we	succeed	joy	sad
onehot	1	1	1	1	0.9	0.1
TF	0.4	0.2	0.2	0.2	0.9	0.1
TF-IDF	1.03	0.7	0.6	1.16	0.9	0.1

$$p(x_k|d_j, e_i) = \frac{x_k}{\sum_{k=1}^{K} x_k}$$

X	step	by	we	succeed	joy	sad
onehot	0.25	0.25	0.25	0.25	0.9	0.1
TF	0.4	0.2	0.2	0.2	0.9	0.1
TF-IDF	0.30	0.20	0.17	0.33	0.9	0.1

1.2 朴素贝叶斯法文本分类

词频特征示例

Documnt	sentence	joy	sad
train1 (d1)	Step by step, we will succeed.	0.9	0.1
train2 (d2)	We step on shit.	0.3	0.7
test1 (d3)	We succeed.	?	?

Table: Example of documents

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	emotion	
Document	step	by	we	succeed	on	shit	will	joy	sad
train1 (d1)	0.33	0.17	0.17	0.17	0	0	0.17	0.9	0.1
train2 (d2)	0.25	0	0.25	0	0.25	0.25	0	0.3	0.7
test1 (d3)	0	0	0.5	0.5	0	0	0	?	?

Table : TF features of documents

1.2 朴素贝叶斯法文本分类

概率计算

为了预测文本 $X_3 = (x_3, x_4)$ 的情感 e_i，我们需要估计：

$$p(e_i|X_3) \propto \sum_{j=1}^{2} p(x_3|e_i, d_j)p(x_4|e_i, d_j)p(d_j, e_i)$$

$$\begin{aligned} p(joy|X_3) \propto\ & p(x_3|joy, d_1)p(x_4|joy, d_1)p(d_1, joy) \\ & + p(x_3|joy, d_2)p(x_4|joy, d_2)p(d_2, joy) \\ & = 0.17 \times 0.17 \times 0.9 + 0.25 \times 0 \times 0.3 = \mathbf{0.02601} \end{aligned}$$

$$\begin{aligned} p(sad|X_3) \propto\ & p(x_3|sad, d_1)p(x_4|sad, d_1)p(d_1, sad) \\ & + p(x_3|sad, d_2)p(x_4|sad, d_2)p(d_2, sad) \\ & = 0.17 \times 0.17 \times 0.1 + 0.25 \times 0 \times 0.7 = \mathbf{0.00289} \end{aligned}$$

1.3 贝叶斯估计

思考：在前面的文本分类算法中，如果测试文本中的单词没有在训练文本中出现会造成什么结果？

会影响到后验概率的计算结果，使分类产生偏差。解决这一问题的方法是采用**贝叶斯估计**。具体地，方法为：

$$p(x_k|d_j, e_i) = \frac{x_k + \lambda}{\sum_{k=1}^{K} x_k + K\lambda}$$

式中 $\lambda \geq 0$。等价于在随机变量各个取值的频数上赋予一个正数 $\lambda \geq 0$。当 $\lambda = 0$ 时就是极大似然估计。尝取 $\lambda = 1$，这时称为**拉普拉斯平滑** (Laplacian smoothing)。

实验任务

- 在给定文本数据集完成文本情感分类训练，在测试集完成测试，计算准确率。

  (提示：可借助 sklearn 机器学习库完成文本特征(tf-idf)提取)