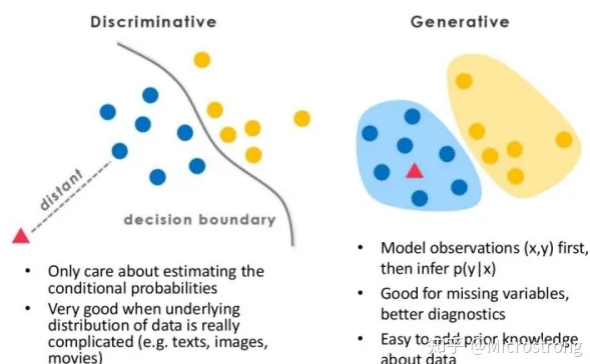


概率式模型和生成式模型

Discriminative vs. Generative



判别式模型可以由概率模型或非概率模型表示：将根据模型直接计算 $p(y|x)$ ，将最大化 $p(y|x)$ 的 y 作为新实例的预测分类（所有类别共享一个模型）

生成式模型**只能由**概率模型表示：对每一个类别都建立一个模型 $p(x, y)$ ，然后通过贝叶斯公式算出 $p(y|x)$ 后选最大概率的 y 作为新分类(每个类别都有一个独有的模型)

朴素贝叶斯分类

利用贝叶斯公式 $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$ 建模，朴素一词在于假设特征之间的独立性
 $p(x^1, x^2, \dots, x^k) = p(x^1)p(x^2)\dots p(x^k)$

上式等号右侧的所有概率项都可以在训练集中学习并记忆；模型、策略、算法三大要素集于一身

EM算法

隐变量的似然度：

$$L(\theta; \mathbf{X}) = p(\mathbf{X}|\theta) = \int p(\mathbf{X}, \mathbf{Z}|\theta) d\mathbf{Z}$$

E - Step :

$$Q(\theta|\theta^{(t)}) = E_{\mathbf{Z}|\mathbf{X}, \theta^{(t)}} [\log L(\theta; \mathbf{X}, \mathbf{Z})]$$

M - Step :

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$$

目的:

可以找出最优模型参数, 或最优隐变量分布

隐变量:

产生过程不可观测, 但是却可以对系统的状态和能观察到的输出存在影响的一种随机变量。

例如: 1.混合高斯模型中的比例系数, 不可以对他做抽样后预测, 但是他却决定了X的概率分布($p(x|\theta)$ 变为了 $p(x|z; \theta)$) $p(x|\theta)$ 变为了 $p(x|z; \theta)$

2.label noise中, 分类问题原本在DNN中可以被建模为 $p(z|x; \theta)$,但是如果假设asymmetric noisy label, 那么中间层输出的真实标签y便会以一种“hard”方式(是一个偏置矩阵, 而不是全连接层)影响原来的条件分布, 可以写为 $p(z|x, y; \theta, \theta_n)$ 从而通过EM算法找出y的分布。

步骤:

E步为通过上一时刻的 θ^t 求得因变量Z的分布 $Q(z^i)$

M步为通过引入对数极大似然, 变量和隐变量联合分布, Jensen不等式 等找到下一时刻的 θ , 算法经过迭代一定会收敛