



Generative model and Discriminative model



Generative model

生成方法学习联合概率密度分布 $P(X, Y)$, 通常可用于直接对数据分布建模或作为求解条件分布 $P(Y|X)$ 的中间步骤(利用贝叶斯公式)。当样本容量增加的时候, 学到的模型可以更快的收敛于真实模型, 当存在隐变量时, 仍可以用生成方法学习(通过如EM算法来估计)。

典型的生成模型有朴素贝叶斯, HMM

Discriminative model

判别方法直接学习的是决策函数 $Y=f(X)$ 或者条件概率分布 $P(Y|X)$ 。不能反映训练数据本身的特性。但它寻找不同类别之间的最优分类面, 反映的是异类数据之间的差异。由于分类或回归任务不需要数据的联合分布, 因此直接面对预测时, 判别模型往往学习的准确率更高。由于直接学习 $P(Y|X)$ 或 $P(X)$, 可以对数据进行各种程度上的抽象、定义特征并使用特征, 因此可以简化学习问题。

典型的判别模型有 Logistic回归, SVM等

Naïve Bayes

为了控制模型复杂度，假设所有属性相互独立

目的是要得到 $p(y|x)$ ，根据贝叶斯公式
$$P(y|x) = \frac{P(y)P(x|y)}{P(x)} = \frac{P(y)}{P(x)} \prod_{i=1}^d P(x_i|y)$$

由于所有类别的 $P(x)$ 相同，所以对优化没有影响

$$h(x) = \arg \max_{c \in y} P(y) \prod_{i=1}^d P(x_i|y)$$

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Naïve Bayes

不是直接求 $p(y|x)$ ，而是根据数据先获得联合分布 $p(x, y)$ ，新样本来时，再根据贝叶斯公式分别得到属于两类的概率

$$h(x) = \arg \max_{c \in y} P(y) \prod_{i=1}^d P(x_i|y)$$

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

–Given a new instance,

$\mathbf{x}' = (\text{Outlook}=\text{Sunny}, \text{Temperature}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=\text{Strong})$

Logistic Regression

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}}$$

通过sigmoid函数将连续函数转为分类问题

$$\ln \frac{y}{1 - y} = \mathbf{w}^T \mathbf{x} + b$$

直接对分类可能性建模，无需实现假设数据分布

$$\ln \frac{p(y = 1|x)}{p(y = 0|x)} = \mathbf{w}^T \mathbf{x} + b$$

将y写成后验概率估计

$$\ell(w, b) = \sum_{i=1}^m \ln p(y_i | x_i; w, b)$$

极大似然估计参数

Logistic Regression

$$\ell(w, b) = \sum_{i=1}^m \ln p(y_i | x_i; w, b)$$

极大似然估计参数

$$\ell(w, b) = \sum_{i=1}^m (-y_i(\mathbf{w}^T \mathbf{x} + b) + \ln(1 + e^{\mathbf{w}^T \mathbf{x} + b}))$$

关于参数向量(w;b)的高阶可导连续凸函数，可直接梯度下降法得到最优解

该方法直接对分类可能进行建模，没有事先假设数据分布，根据损失函数直接优化参数向量，属于判别模型。

Analysis

$P(X, Y)$ 的估计本身会很困难，需要的计算量和训练数据量都会非常巨大，像 Naive Bayes 那样为了使得模型复杂度被控制在可以处理的范围内就不得不做了非常强的假设，而 Logistic Regression 则不太会受到 X 上的相关性之类的限制。

Q: 得到的联合分布获得的信息更多，是不是分类效果更好？

A: 不是，在回归和分类任务中通常不需要联合分布这样的信息，多余的迂回步骤有可能会积累更多的误差

Q: 那为什么要生成模型？

A: 可以完成需要<联合分布>才能完成的任务，如可以根据学得的分自己生成数据，能够处理 latent variable , partially observed data 以及 partially labeled data

Reference

- [1] Ng A Y, Jordan M I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes[C]//Advances in neural information processing systems. 2002: 841-848.
- [2] 李航. 统计学习方法[J]. 清华大学出版社, 北京, 2012.
- [3] 周志华. 机器学习[M]. Qing hua da xue chu ban she, 2016.



Thank you