算法推导

李沛泽[*]:1701111586

晁越[†]:1601110127

2017 年 12 月 5 日

[*]工学院,1701111586@pku.edu.cn

[†]物理学院,litterel@pku.edu.cn

1 条件随机场模型学习的梯度下降算法

输入：特征函数 $f_1, f_2, \cdots, f_n$；经验分布 $\widetilde{P}(X,Y)$；

输出：最优参数值 $\hat{\omega}$；最优模型 $P_{\hat{\omega}}(y|x)$.

1. 选定初始点 $\omega^{(0)}$，置 $k=0$

2. 计算 $f(\omega^{(k)})$

3. 计算梯度 $g_k = g(\omega^{(k)})$，若 $\|g_k\| < \varepsilon$，则停止计算，令 $\hat{\omega} = \omega^{(k)}$；否则，令 $p_k = -g(\omega^{(k)})$，求 λ_k，使得

$$f(\omega^{(k)} + \lambda_k p_k) = \min_{\lambda \geq 0} f(\omega^{(k)} + \lambda p_k) \tag{1}$$

4. 置 $\omega^{(k+1)} = \omega^{(k)} + \lambda_k p_k$，计算 $f(\omega^{(k+1)})$, 当 $\|f(\omega^{(k+1)}) - f(\omega^{(k)})\| < \varepsilon$ 或 $\|\omega^{(k+1)} - \omega^{(k)}\| < \varepsilon$ 时，停止迭代，令 $\hat{\omega} = \omega^{(k+1)}$

5. 否则，置 $k = k+1$，转 3

2 非监督朴素贝叶斯算法的 EM 导出

2.1 模型描述

朴素贝叶斯算法可以应用在文本分类，垃圾邮件过滤等方面，一般情况下样本形式为 $\{(x^{(1)}, z^{(1)}), \ldots, (x^{(m)}, z^{(m)})\}$，其中 x 表示样本特征，z 表示样本类别，根据样本可以生成先验概率 $p(z)$ 的分布以及条件概率分布 $p(x|z)$，然后由贝叶斯公式

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

来计算后验概率，比较后验概率的大小给出测试样本所属类别。

在非监督学习的情况下，由于训练样本没有给出类别，所以需要使用非监督的朴素贝叶斯学习方法。以垃圾邮件过滤为例，给定 m 个样本的训练集合 $\{x^{(1)}, \ldots, x^{(m)}\}$，每个样本 $x^{(i)}$ 属于 $(0,1)^n$，即根据词典将邮件文本转化为 n 维的 $(0,1)$ 向量，故 $x_j^{(i)}$ 表示词典中第 j 个词是否出现在样本 i 中。我们需要根据这些没有类别标记的训练样本得到先验概率和后验概率。下面我们通过 EM 算法来导出非监督朴素贝叶斯学习方法。

2.2 算法推导

明确隐变量 观察到的数据是每一个样本对应的 n 维 $(0,1)$ 向量，隐变量是类别 z 的先验概率，以及在 z 的条件下字典中第 j 个元素是否出现的条件概率。令 $\mu = p(z=1), \Sigma = p(x|z)$，其中 Σ 是一个 $n \times 2$ 维矩阵，即 $p(x_j^{(i)} = 1|z^{(i)} = 1) = \Sigma_{j,1}$, $p(x_j^{(i)} = 1|z^{(i)} = 0) = \Sigma_{j,0}$。上标 i 表示第 i 个样本。μ 和 Σ 等价于这个模型的全部参数 θ。

EM 算法 E 步：确定 Q 函数 首先我们可以写出 EM 算法的 Q 函数

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E_z[\log p(x, z|\theta)|x, \theta^{(k)}] \\ &= \sum_{i=1}^{m} \sum_{z^{(i)}} \log p(x^{(i)}, z^{(i)}|\theta) p(z^{(i)}|x^{(i)}, \theta^{(k)}) \end{aligned}$$

其中

$$\begin{aligned} \log p(x^{(i)}, z^{(i)}|\theta) &= \log(p(x^{(i)}|z^{(i)}, \theta) p(z^{(i)}|\theta)) \\ &= \log p(x^{(i)}|z^{(i)}, \theta) + \log p(z^{(i)}|\theta) \end{aligned}$$

带入 $Q(\theta, \theta^{(k)})$ 的表达式中, 可以将 $Q(\theta, \theta^{(k)})$ 写成两部分之和，第一部分为

$$
\begin{aligned}
part(1) =& \sum_{i=1}^{m} \sum_{z^{(i)}} \log p(z^{(i)}|\theta) p(z^{(i)}|x^{(i)}, \theta^{(k)}) && (2)\\
=& \sum_{i=1}^{m} \log p(z^{(i)} = 1|\theta) p(z^{(i)} = 1|x^{(i)}, \theta^{(k)}) && (3)\\
&+ \sum_{i=1}^{m} \log p(z^{(i)} = 0|\theta) p(z^{(i)} = 0|x^{(i)}, \theta^{(k)}) && (4)\\
=& \log \mu \sum_{i=1}^{m} p(z^{(i)} = 1|x^{(i)}, \theta^{(k)}) && (5)\\
&+ \log(1-\mu) \sum_{i=1}^{m} p(z^{(i)} = 0|x^{(i)}, \theta^{(k)}) && (6)
\end{aligned}
$$

结合朴素贝叶斯的基本假设

$$
\begin{aligned}
p(x^{(i)}|z^{(i)}) =& \prod_{j=1}^{n} p(x_j^{(i)}|z^{(i)})\\
=& \prod_{x_j^{(i)}=1} \Sigma_{j,z^{(i)}} \prod_{x_j^{(i)}=0} (1 - \Sigma_{j,z^{(i)}})
\end{aligned}
$$

第二部分可写为

$$part(2) = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{z^{(i)}} p(z^{(i)}|x^{(i)},\theta^{(k)}) \log p(x_j^{(i)}|z^{(i)},\theta) \tag{7}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{m} p(z^{(i)}=1|x^{(i)},\theta^{(k)}) \log p(x_j^{(i)}|z^{(i)}=1,\theta) \tag{8}$$

$$+ \sum_{j=1}^{n}\sum_{i=1}^{m} p(z^{(i)}=0|x^{(i)},\theta^{(k)}) \log p(x_j^{(i)}|z^{(i)}=0,\theta) \tag{9}$$

$$= \sum_{i=1}^{m} p(z^{(i)}=1|x^{(i)},\theta^{(k)}) \left(\sum_{x_j^{(i)}=1} \log \Sigma_{j,1} + \sum_{x_j^{(i)}=0}^{n} \log(1-\Sigma_{j,1}) \right) \tag{10}$$

$$+ \sum_{i=1}^{m} p(z^{(i)}=0|x^{(i)},\theta^{(k)}) \left(\sum_{x_j^{(i)}=0} \log \Sigma_{j,0} + \sum_{x_j^{(i)}=0}^{n} \log(1-\Sigma_{j,0}) \right) \tag{11}$$

于是 $Q(\theta,\theta^{(k)}) = part(1) + part(2)$

确定 EM 算法的 M 步 首先确定 μ，令 $\nabla_\mu Q(\theta,\theta^{(k)}) = 0$，由于 $part(2)$ 与 μ 无关，于是 $\nabla_\mu part(1) = 0$, 根据2整理得到

$$\mu = \frac{\sum_{i=1}^{m} p(z^{(i)}=1|x^{(i)},\theta^{(k)})}{\sum_{i=1}^{m} \left(p(z^{(i)}=1|x^{(i)},\theta^{(k)}) + p(z^{(i)}=0|x^{(i)},\theta^{(k)}) \right)} \tag{12}$$

然后确定 $\Sigma_{j,1}$，令 $\nabla_{\Sigma_{j,1}} Q(\theta,\theta^{(k)}) = 0$，由于 $part(1)$ 与 Σ 无关，因此等价于 $\nabla_{\Sigma_{j,1}} part(2) = 0$, 根据7式，可以得到

$$\frac{\sum_{i:x_j^{(i)}=1} p(z^{(i)}=1|x^{(i)},\theta^{(k)})}{\Sigma_{j,1}} + \frac{\sum_{i:x_j^{(i)}=0} p(z^{(i)}=1|x^{(i)},\theta^{(k)})}{\Sigma_{j,1}-1} = 0$$

整理得到

$$\Sigma_{j,1} = \frac{\sum\limits_{i:x_j^{(i)}=1} p(z^{(i)}=1|x^{(i)},\theta^{(k)})}{\sum\limits_{i:x_j^{(i)}=1} p(z^{(i)}=1|x^{(i)},\theta^{(k)}) + \sum\limits_{i:x_j^{(i)}=0} p(z^{(i)}=1|x^{(i)},\theta^{(k)})} \tag{13}$$

$$= \frac{\sum\limits_{i:x_j^{(i)}=1} p(z^{(i)}=1|x^{(i)},\theta^{(k)})}{\sum\limits_{i=1}^{m} p(z^{(i)}=1|x^{(i)},\theta^{(k)})} \tag{14}$$

同理可以得到

$$\Sigma_{j,0} = \frac{\sum\limits_{i:x_j^{(i)}=1} p(z^{(i)}=0|x^{(i)},\theta^{(k)})}{\sum\limits_{i=1}^{m} p(z^{(i)}=0|x^{(i)},\theta^{(k)})} \tag{15}$$

$$= \frac{\sum\limits_{i:x_j^{(i)}=1} \left(1 - p(z^{(i)}=1|x^{(i)},\theta^{(k)})\right)}{1 - \sum\limits_{i=1}^{m} p(z^{(i)}=1|x^{(i)},\theta^{(k)})} \tag{16}$$

2.3 算法步骤

1. 在 $[0,1]$ 之间随机生成 $\mu^{(0)}, \Sigma_{j,1}^{(0)}, \Sigma_{j,0}^{(0)}$，并且使其满足概率归一化条件
2. 开始迭代，根据式12, 式13, 式15分别计算 $\mu^{(k+1)}, \Sigma^{(k+1)}$
3. 如果 $\|\theta^{(k+1)} - \theta^{(k)}\| < \epsilon$，退出迭代，输出模型，否则返回第 2 步

分工说明

李沛泽同学完成了条件随机场的梯度下降算法推导，晁越同学完成了非监督朴素贝叶斯的 EM 算法推导。