模式识别与机器学习

4. 15、 Softmax判据的概念

引言

逻辑回归输出：属于正类的后验概率

$$\mathrm{sigmoid}(\boldsymbol{w}^T\boldsymbol{x} + w_0) = p(C_1|\boldsymbol{x})$$

引言

对于多类而言，

每类的后验概率如何表达？

后验概率：多类情况

一个类与剩余类的后验概率比率

- 逻辑回归是由Logit变换反推出来的。
- 由Logit变换可知：正负类后验概率比率的对数是一个线性函数。

$$\ln\frac{p(C_1|\boldsymbol{x})}{p(C_2|\boldsymbol{x})}=\boldsymbol{w}^T\boldsymbol{x}+w_0 \quad\Rightarrow\quad \frac{p(C_1|\boldsymbol{x})}{p(C_2|\boldsymbol{x})}=\exp\left(\boldsymbol{w}^T\boldsymbol{x}+w_0\right)$$

- 分类K个类，可以构建K个线性判据。第i个线性判据表示C_i类与剩余类的分类边界，剩余类用一个参考负类（reference class）C_K来表达。

$$\frac{p(C_i|\boldsymbol{x})}{p(C_K|\boldsymbol{x})}=\exp\left(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i}\right) \qquad \text{where } K\neq i$$

后验概率：多类情况

参考类的后验概率$p(C_K|\boldsymbol{x})$

- 由于所有类别后验概率之和为1，参考类C_K的后验概率为：

$$\frac{p(C_i|\boldsymbol{x})}{p(C_K|\boldsymbol{x})} = \exp(\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i})$$

$$p(C_K|\boldsymbol{x}) = 1 - \sum_{i=1}^{K-1} p(C_i|\boldsymbol{x})$$

$$= 1 - p(C_K|\boldsymbol{x}) \sum_{i=1}^{K-1} \exp(\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i})$$

$$\Rightarrow \quad p(C_K|\boldsymbol{x}) = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i})}$$

后验概率：多类情况

任意正类的后验概率$p(C_i|\boldsymbol{x})$

- 任意正类$C_i(i \neq K)$的后验概率为：

$$\frac{p(C_i|\boldsymbol{x})}{p(C_K|\boldsymbol{x})} = \exp\left(\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i}\right)$$

$$\Rightarrow p(C_i|\boldsymbol{x}) = \frac{\exp(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i})}{1+\sum_{j=1}^{K-1}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})}$$

后验概率：多类情况

重新审视参考负类的后验概率$p(C_K|\boldsymbol{x})$

- 把$p(C_K|\boldsymbol{x})$参照$p(C_i|\boldsymbol{x})$的形式，可以类比得到：

$$\left.\begin{aligned} p(C_K|\boldsymbol{x}) &= \frac{1}{1+\sum_{j=1}^{K-1}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})} \\ p(C_i|\boldsymbol{x}) &= \frac{\exp(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i})}{1+\sum_{j=1}^{K-1}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})} \end{aligned}\right\} \Rightarrow \exp(\boldsymbol{w}_K^T\boldsymbol{x}+w_{0K}) = 1$$

$$\boldsymbol{w}_K^T\boldsymbol{x}+w_{0K} = 0$$

后验概率：多类情况

任意类的后验概率$p(C_i|\boldsymbol{x})$

- 因此，得到任意一个类的后验概率$p(C_i|\boldsymbol{x})$：

$$p(C_i|\boldsymbol{x}) = \frac{\exp(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i})}{\sum_{j=1}^{K}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})}, \quad i = 1, 2, \ldots, K$$

- 可见，对于多类分类（$K > 2$），K个线性模型也跟每个类对应的后验概率建立起了联系。

Softmax函数

Softmax函数

$$\text{softmax:} \quad z_i = \frac{\exp(y_i)}{\sum_{j=1}^{K} \exp(y_j)}, \quad i = 1, 2, \ldots, K$$

- 该后验概率计算方法被称作softmax函数：
  - ✓ 如果一个类C_i对应的y_i（即线性模型的输出）远远大于其他类的，经过exp函数和归一化操作，该类的后验概率$p(C_i|\boldsymbol{x})$接近于1，其他类的后验概率接近于0，则softmax决策就像是一个max函数。
  - ✓ 但是，该函数又是可微分的，所以称作“软最大值函数”。

Softmax判据

Softmax判据

- Softmax判据：K个线性判据 + softmax函数。

Softmax判据的决策过程

- 给定测试样本$\boldsymbol{x}$，经由线性模型和softmax函数计算K个类对应的后验概率， $\boldsymbol{x}$属于最大后验概率对应的类别。

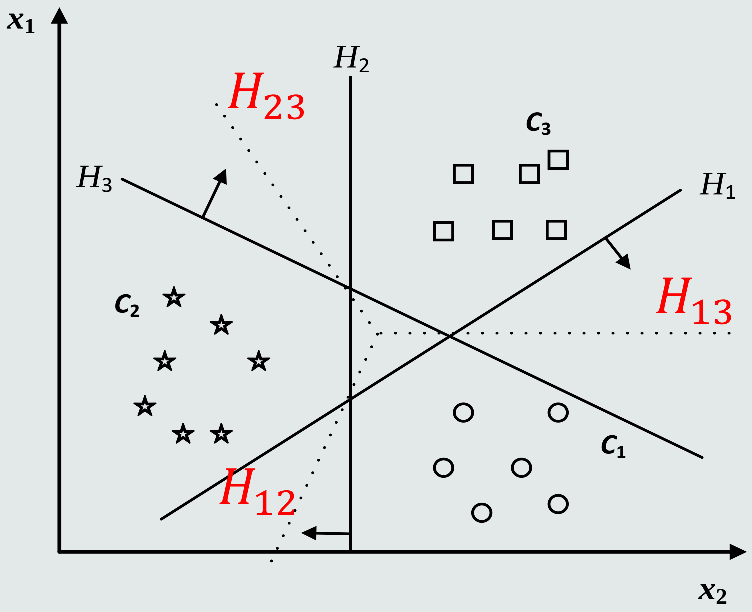
$$l(\boldsymbol{x}) = \arg\max_i p(C_i|\boldsymbol{x}) = \arg\max_i \frac{\exp(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i})}{\sum_{j=1}^{K}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})}$$

决策边界

Softmax判据的决策边界

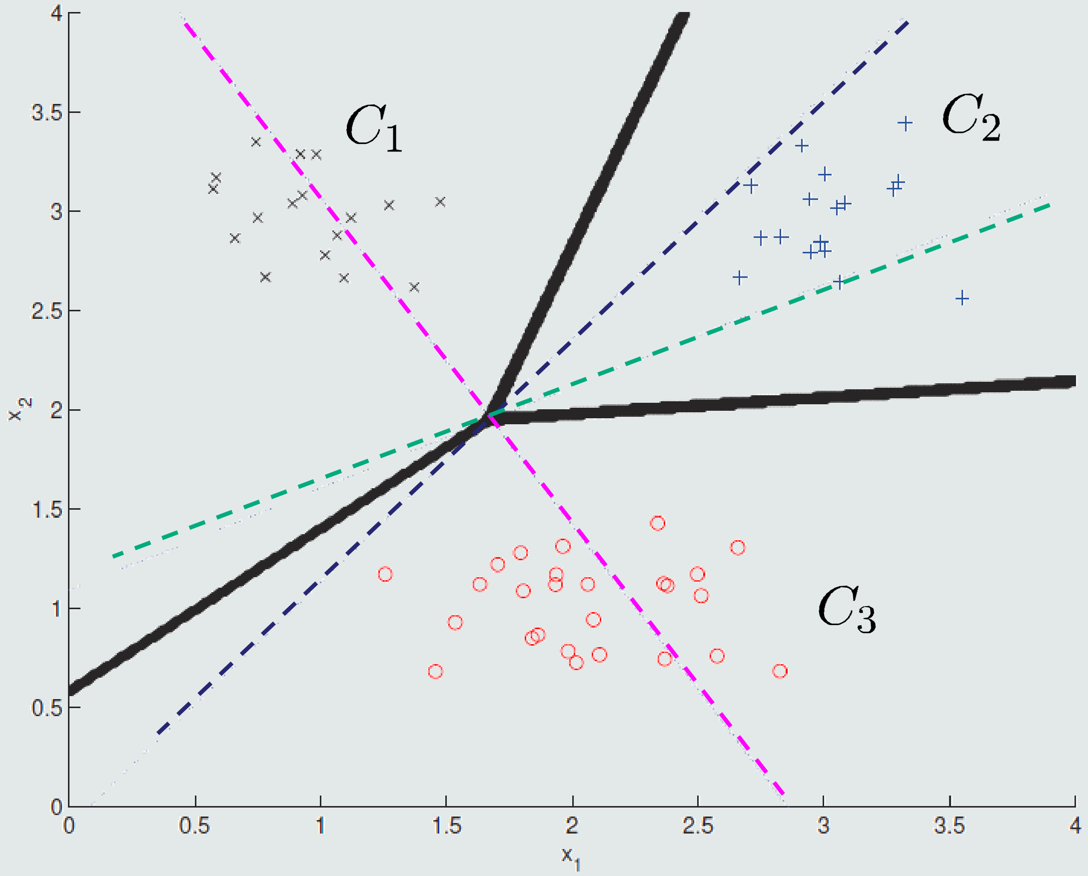
- Softmax判据用于分类，等同于基于one-to-all策略的线性机。
  - ✓ 决策区域相邻的两类C_i和C_j，它们的决策边界H_{ij}为线性：

$$
\begin{aligned}
& p(C_i|\boldsymbol{x}) = p(C_j|\boldsymbol{x}) \\
\Rightarrow\ & \exp(\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i}) = \exp(\boldsymbol{w}_j^T\boldsymbol{x} + w_{0j}) \\
\Rightarrow\ & \boldsymbol{w}_i^T\boldsymbol{x} + w_{0i} = \boldsymbol{w}_j^T\boldsymbol{x} + w_{0j} \\
\Rightarrow\ & (\boldsymbol{w}_i - \boldsymbol{w}_j)^T\boldsymbol{x} + (w_{0i} - w_{0j}) = 0
\end{aligned}
$$



Softmax判据决策示例

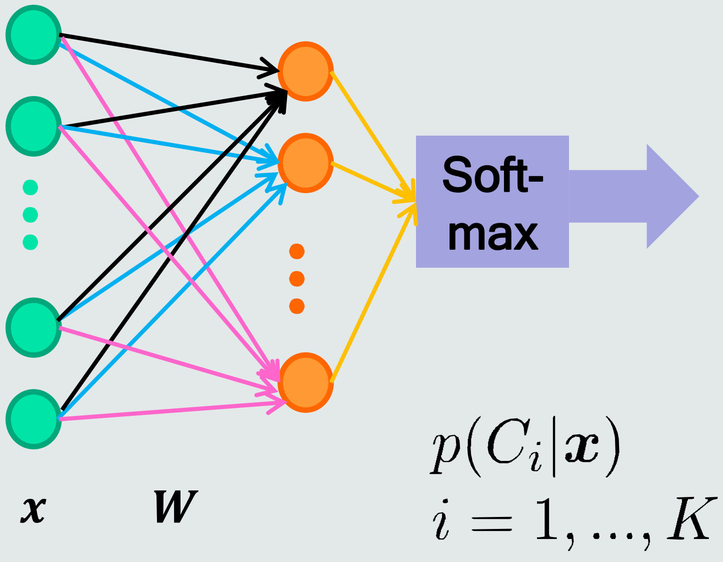
- $K = 3$
- 细线：3条线性方程$\boldsymbol{w}_i^T\boldsymbol{x} + w_{0i}$；粗线：softmax判据决策边界。



Softmax判据与神经网络

Softmax判据与神经网络

- 相当于由K个神经元组成的神经网络输出层，每个神经元输出值代表属于该类的后验概率$p(C_i|\boldsymbol{x}) \in [0,1]$。
- 注意：$\sum_{i=1}^{K} p(C_i|\boldsymbol{x}) = 1$



- Softmax判据的决策过程可以进一步表达为：

$$l(\boldsymbol{x}) = \arg\max_i \mathrm{softmax}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{w}_0)$$

$$\text{where } \boldsymbol{W} = [\boldsymbol{w}_1, \dots, \boldsymbol{w}_K]^T, \ \boldsymbol{w}_0 = [w_{01}, \dots, w_{0K}]^T$$

对比：线性机模型

适用范围

适用范围：分类

- 前提：每个类和剩余类之间是线性可分的。

适用范围：回归

- 范围：可以拟合指数函数（exp）形式的非线性曲线。

总结

总结

- Softmax判据本身是一个非线性模型。
- Softmax判据用于分类：只能处理多个类别、每个类别与剩余类线性可分的情况。但是， Softmax判据可以输出后验概率。因此，Softmax判据比基于one-to-all策略的线性机向前迈进了一步。
- Softmax判据用于拟合：可以输出有限的非线性曲线。

模型对比

模型	线性判据	逻辑回归	Softmax判据	MAP分类器
模型表达	$\boldsymbol{w}^T\boldsymbol{x}+w_0$	$\text{sigmoid}(\boldsymbol{w}^T\boldsymbol{x}+w_0)$	$p(C_i\|\boldsymbol{x}) = \frac{\exp(\boldsymbol{w}_i^T\boldsymbol{x}+w_{0i})}{\sum_{j=1}^{K}\exp(\boldsymbol{w}_j^T\boldsymbol{x}+w_{0j})}$	$p(C_i\|\boldsymbol{x}) = \frac{p(\boldsymbol{x}\|C_i)p(C_i)}{p(\boldsymbol{x})}$
模型属性	线性	非线性	非线性	线性/非线性
决策边界	线性	线性	线性	线性/ 非线性
输出信号	几何距离	后验概率	后验概率	后验概率
分类	二类分类 (线性可分)	二类分类 (线性可分)	多类分类 (每类与剩余类线性可分)	多类分类 (线性可分/不可分)
回归	拟合线性	拟合非线性曲线 (sigmoid形)	拟合非线性曲线 (exp形)	--

Softmax判据如何学习？