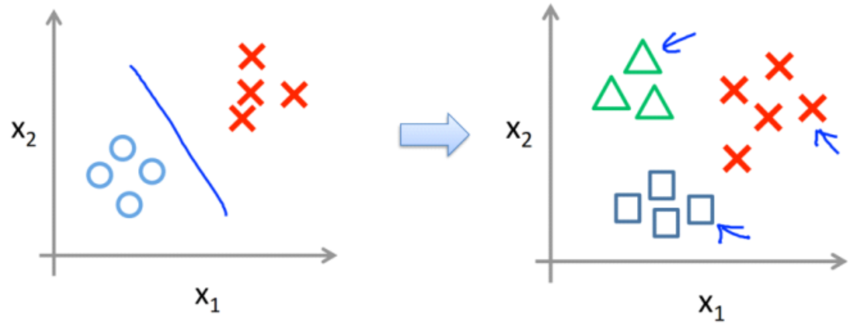
Class 3 Softmax回归

2018年10月17日 17:31

- ## 简介：

  - Softmax回归是一种多分类模型，也叫多分类逻辑回归；
  - 在NLP(Natural Language Processing) 中，也被叫做最大熵模型；
  - Softmax是一种被广泛使用的分类算法。



- ## 模型假设：

  - **假设**

$$p(y=j|x;\theta)=h_j(x)=\frac{e^{\theta_j{}^{\mathrm{T}}x}}{1+\sum_{j'=1}^{C-1}e^{\theta_{j'}{}^{\mathrm{T}}x}},j=1,\ldots,C-1$$

$$p(y=C|x;\theta)=h_C(x)=\frac{1}{1+\sum_{j'=1}^{C-1}\exp\{\theta_{j'}{}^{\mathrm{T}}x\}}$$

  - **上面两个式子可合并为简洁形式**

$$p(y=j|x;\theta)=h_j(x)=\frac{e^{\theta_j{}^{\mathrm{T}}x}}{\sum_{j'=1}^{C}e^{\theta_{j'}{}^{\mathrm{T}}x}},j=1,2,\ldots,C,\text{where }\theta_C=\vec{0}$$

  - **参数**

$$\theta_{C\times M}$$

其中，c 为类别数量，M为样本数量

- ## 最大似然估计：

  - **对数似然函数（多分类问题）**

$$l(\theta) = \sum_{i=1}^{N} \log p(y^{(i)}|x^{(i)};\theta)$$ **Softmax Regression**

$$= \sum_{i=1}^{N} \log \prod_{j=1}^{C} \left(\frac{e^{\theta_j{}^{\mathrm{T}} x}}{\sum_{j'=1}^{C} e^{\theta_{j'}{}^{\mathrm{T}} x}} \right)^{1\{y^{(i)}=j\}}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{C} 1\{y^{(i)} = j\} \log \left(\frac{e^{\theta_j{}^{\mathrm{T}} x}}{\sum_{j'=1}^{C} e^{\theta_{j'}{}^{\mathrm{T}} x}} \right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{C} 1\{y^{(i)} = j\} \log h_j(x^{(i)})$$

- 对于二分类问题，采用逻辑回归

Logistic Regression

$$l(\theta) = \sum_{i=1}^{N} y^{(i)} \log h_\theta\left(x^{(i)}\right) + (1 - y^{(i)}) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

- 当Softmax回归模型的类别数为2时，退化为逻辑回归模型。

• 梯度下降优化

- **梯度**

$$\frac{\partial l(\theta)}{\partial \theta_k} = \sum_{i=1}^{N} \boxed{\left(1\{y^{(i)} = k\} - h_k(x^{(i)})\right) x^{(i)}}$$

Error × Feature

其中：

$$\frac{\partial \log\ h_j(x)}{\partial \theta_k} = \begin{cases} \left(1 - h_k(x)\right)x, & j = k \\ -h_k(x)x, & j \neq k \end{cases}$$

$$\frac{\partial \sum_{j=1}^{C} 1\{y = j\} \log h_j(x)}{\partial \theta_k} = \begin{cases} \left(1 - h_k(x)\right)x, & y = k \\ -h_k(x)x, & y \neq k \end{cases}$$

$$= \left(1\{y = k\} - h_k(x)\right)x$$

- **梯度下降**

$$\theta_k := \theta_k + \alpha \sum_{i=1}^{N} \left(1\{y^{(i)} = k\} - h_k(x^{(i)})\right) x^{(i)}$$

$$\text{where } h_k(x) = \frac{e^{\theta_k{}^{\mathrm{T}} x}}{\sum_{k'=1}^{C} e^{\theta_{k'}{}^{\mathrm{T}} x}}, k = 1,2,\dots,C$$

- **随机梯度下降**

随机选择一个样本，更新参数：

$$\theta_k := \theta_k + \alpha\left(1\{y = k\} - h_k(x)\right)x$$