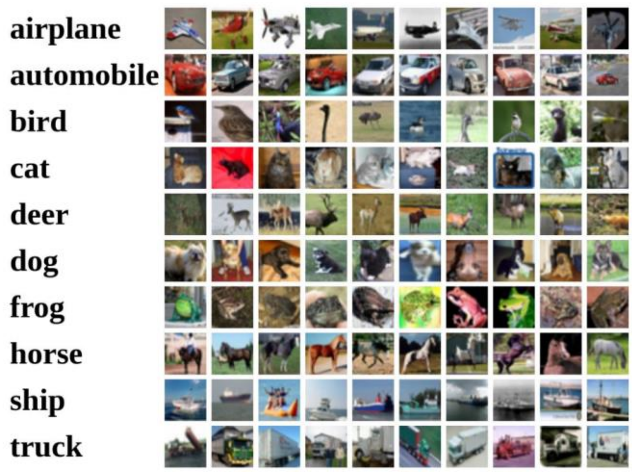
人工智能与自动化学院

# 模 式 识 别

## 第九讲 多类别分类(*Classification for Multiclass*)

## 9.1 “一对多”策略的多类别分类

**多类别分类问题**
**(*Multiclass Classification*)**

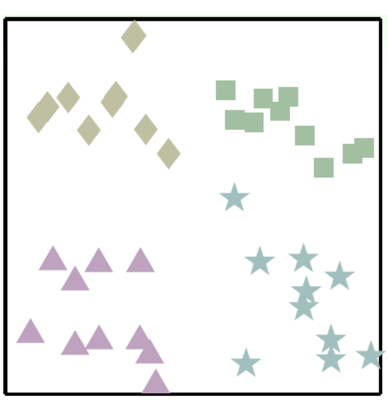



CIFAR 10 dataset



## 9.1 “一对多”策略的多类别分类

**多类别分类问题**
**(*Multiclass Classification*)**



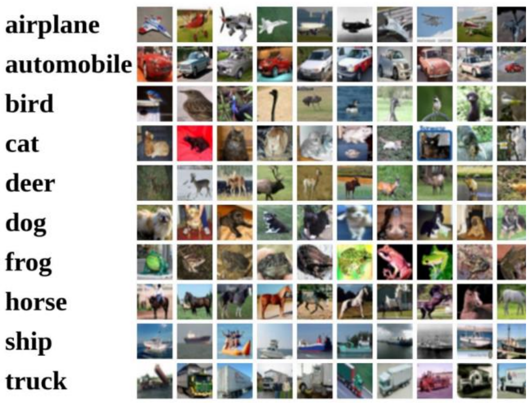
$\mathcal{Y} = \{$ ■, ◆, ▲, ★ $\}$ 类别数$\mathcal{K} = 4$

能否用二分类方法完成多分类问题?

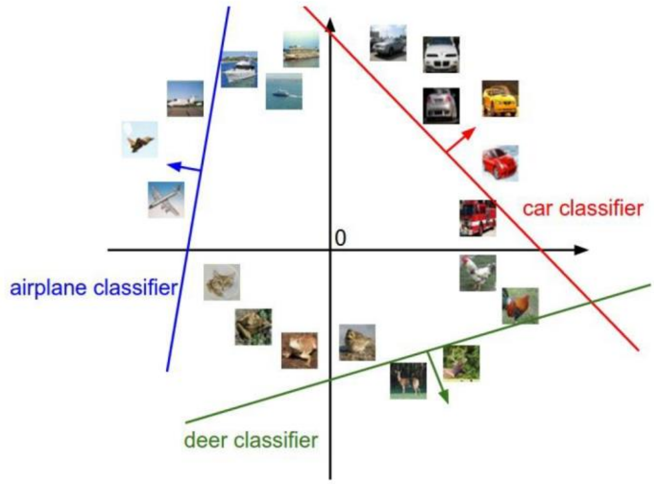
# 9.1 “一对多”策略的多类别分类

## 策略：一次只区分一个类别

**(*One Class at a Time*)**




CIFAR 10 dataset

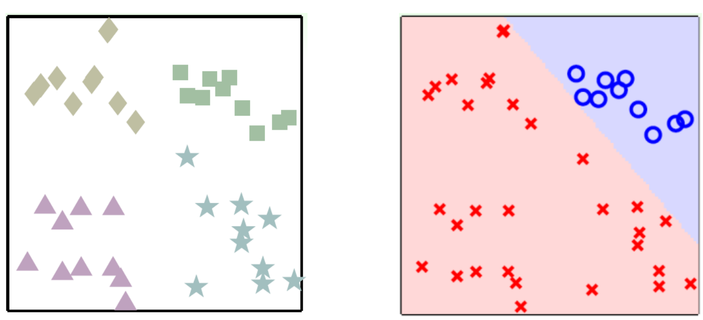





# 9.1 “一对多”策略的多类别分类

## 策略：一次只区分一个类别

**(*One Class at a Time*)**

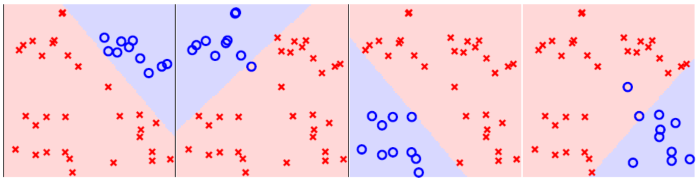


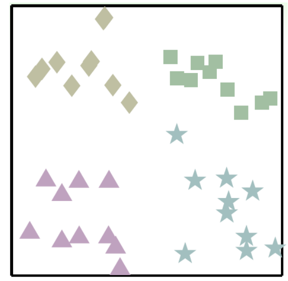
$\square$ or not? $\{\square = \circ, \diamond = \times, \triangle = \times, \star = \times\}$

# 9.1 “一对多”策略的多类别分类

**策略：一次只区分一个类别**

**(*One Class at a Time*)**



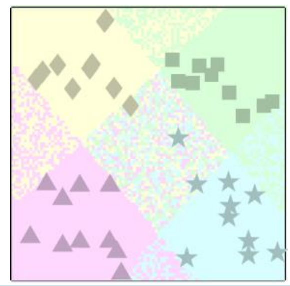


$\square$ or not? $\{\square=\circ, \Diamond=\times, \triangle=\times, \star=\times\}$

$\Diamond$ or not? $\{\square=\times, \Diamond=\circ, \triangle=\times, \star=\times\}$

$\triangle$ or not? $\{\square=\times, \Diamond=\times, \triangle=\circ, \star=\times\}$

$\star$ or not? $\{\square=\times, \Diamond=\times, \triangle=\times, \star=\circ\}$

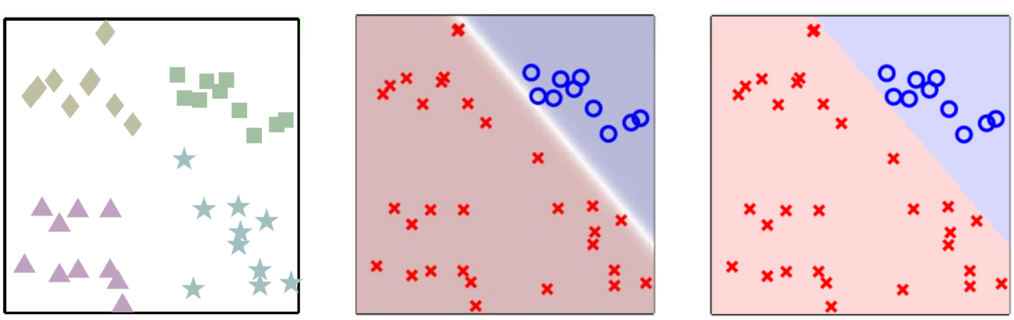




9.1 “一对多”策略的多类别分类

策略：一次只区分一个类别

(One Class at a Time)

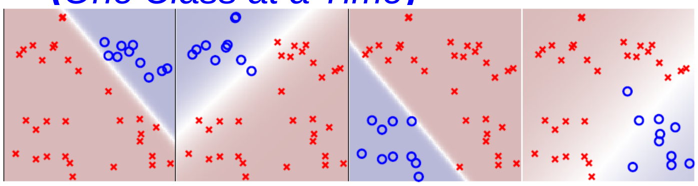


$P(\square|\mathbf{x})$? $\{\square=\circ, \Diamond=\times, \triangle=\times, \star=\times\}$

## 9. 1 “一对多”策略的多类别分类

**策略：一次只区分一个类别**

**(*One Class at a Time*)**

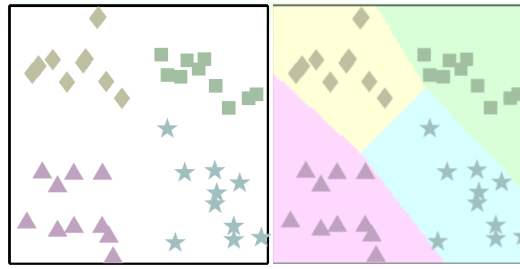


$P(\square|\mathbf{x})?\ \{\square = \circ, \lozenge = \times, \triangle = \times, \star = \times\}$

$P(\lozenge|\mathbf{x})?\ \{\square = \times, \lozenge = \circ, \triangle = \times, \star = \times\}$

$P(\triangle|\mathbf{x})?\ \{\square = \times, \lozenge = \times, \triangle = \circ, \star = \times\}$

$P(\star|\mathbf{x})?\ \{\square = \times, \lozenge = \times, \triangle = \times, \star = \circ\}$



测试样本为$\mathbf{x}$时，所属类别为：

$$g(\mathbf{x}) = argmax_{k\in\mathcal{Y}}(\mathbf{w}_{[k]}^T\mathbf{x})$$



## 9. 1 “一对多”策略的多类别分类

***“One-Versus-All (OVA) ”*** **策略对训练样本集的重分组**

① $for\ k \in \mathcal{Y}$, 对训练样本集重新分组得到 $\mathcal{D}_{[k]}$:

$$\mathcal{D}_{[k]} = \{(\mathbf{x}_n, y_n' = 2[\![y_n = k]\!] - 1)\}_{n=1}^N$$

② 在数据集$\mathcal{D}_{[k]}$上运行任一二分类算法，如*Logistic regression*，得到$\mathbf{w}_{[k]}$

③ 当测试样本为$\mathbf{x}$时，分类结果为： $g(\mathbf{x}) = argmax_{k\in\mathcal{Y}}(\mathbf{w}_{[k]}^T\mathbf{x})$

- 优点：简单、便于推广二分类方法实现多分类问题
- 不足：当类别数$\mathcal{K}$很大时， $\mathcal{D}_{[k]}$存在样本数不平衡问题，影响性能

不平衡问题的来源在于“一对多”的策略

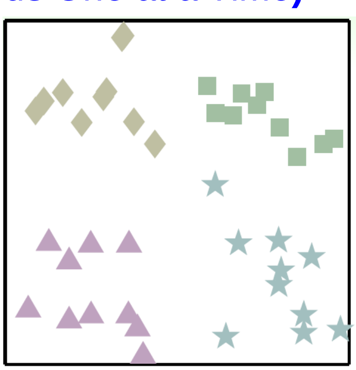
## 第九讲 多类别分类(*Classification for Multiclass*)

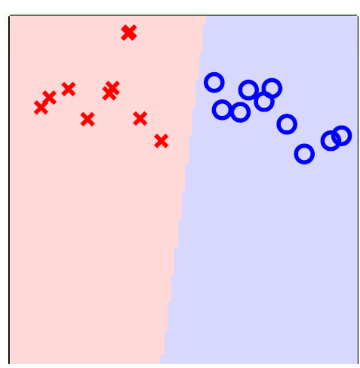




## 9.2 “一对一”策略的多类别分类

**策略：一次只区分两个类别**
**(*One versus One at a Time*)**

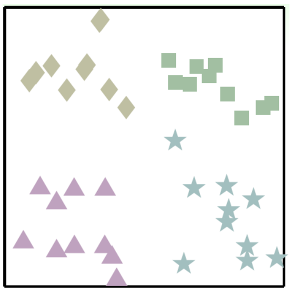




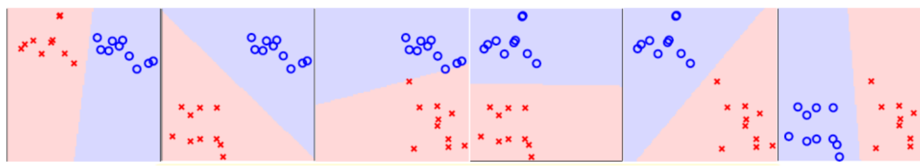
□ or ◊? {□ = ○, ◊ = ×, △ = nil, ⋆ = nil}

## 9.2 “一对一”策略的多类别分类

**策略：一次只区分两个类别**
**(*One versus One at a Time*)**



类别数两两组合，用二分类算法得到$C_{\mathcal{K}}^{2}$个分类面：$\boldsymbol{w}_{[k,l]}$



测试样本为$\mathbf{x}$时，所属类别为：$g(\mathbf{x}) = tournament\ champion_{k,l\in\mathcal{Y}*\mathcal{Y}}(\boldsymbol{w}_{[k,l]}^{T}\mathbf{x})$

样本属于得票最多的类别



## 9.1 “一对多”策略的多类别分类

**“One-Versus-One (OVO) ”策略对训练样本集的重分组**

① $for\ (k,l)\in\mathcal{Y}*\mathcal{Y}$, 对训练样本集重新分组得到 $\mathcal{D}_{[k,l]}$:

$$\mathcal{D}_{[k,l]} = \{(\mathbf{x}_n, y_n' = 2[\![y_n = k]\!] - 1) : y_n = k\ or y_n = l\}_{n=1}^{N}$$

② 在数据集$\mathcal{D}_{[k]}$上运行任一二分类算法，如*Logistic regression*，得到$\boldsymbol{w}_{[k,l]}$

③ 当测试样本为$\mathbf{x}$时，分类结果为：$g(\mathbf{x}) = tournament\ champion_{k,l\in\mathcal{Y}*\mathcal{Y}}(\boldsymbol{w}_{[k,l]}^{T}\mathbf{x})$

- 优点：简单、有效、稳定、便于推广二分类方法实现多分类问题
- 不足：类别组合后求解$\boldsymbol{w}_{[k,l]}$ 需要更多空间、更耗时

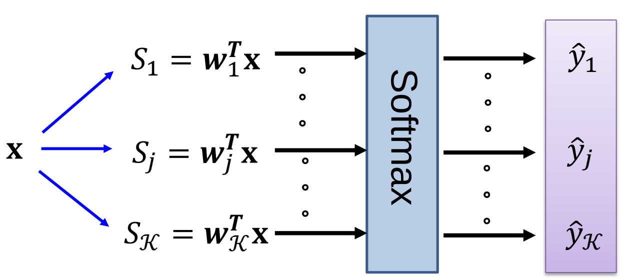
# 第九讲 多类别分类(*Classification for Multiclass*)





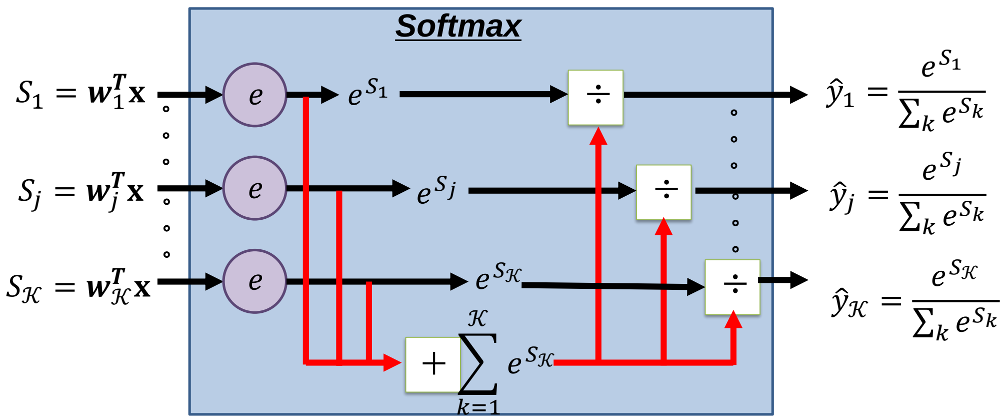
# 9.3 “Softmax”多类别分类

$$\mathcal{Y} = \{1,2,\dots,j,\dots\mathcal{K}\}$$



## 9.3 “Softmax”多类别分类

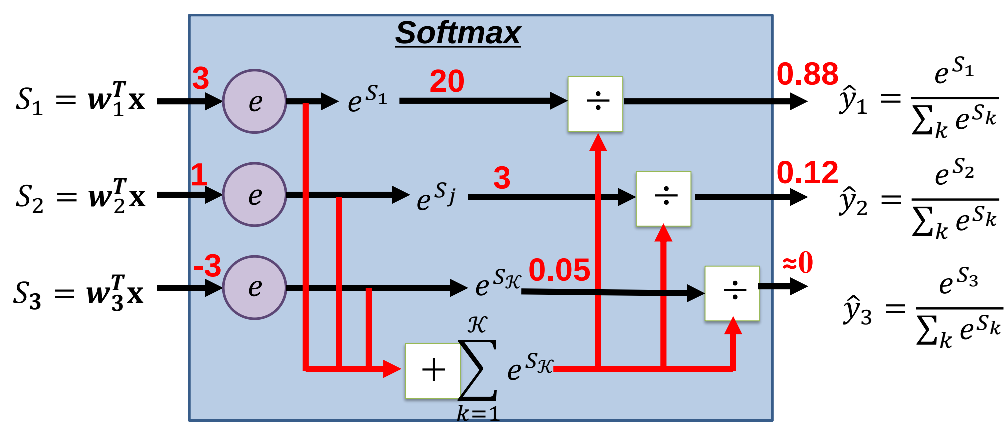
$\mathcal{Y} = \{1,2,\ldots,j,\ldots\mathcal{K}\}$






## 9.3 “Softmax”多类别分类

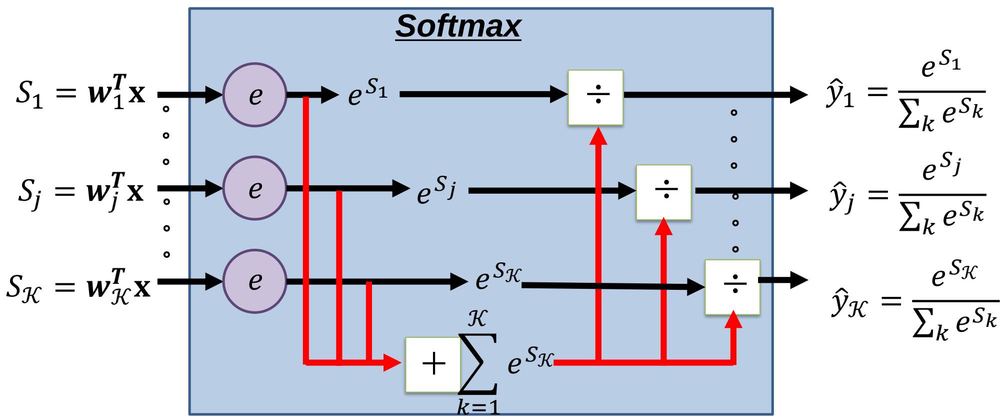
$\mathcal{Y} = \{1,2,3\}$




输出具有概率特性：$0 < \hat{y}_k < 1, \quad \sum_k \hat{y}_k = 1$

## 9.3 “Softmax”多类别分类

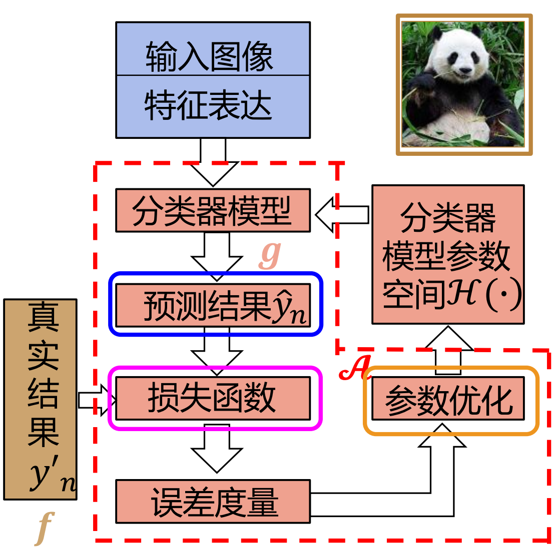
$\mathcal{Y} = \{1,2,\ldots,j,\ldots\mathcal{K}\}$




输出具有概率特性：$0 < \hat{y}_k < 1$，　$\sum_k \hat{y}_k = 1$



## 5.1 逻辑斯蒂回归问题（引）




逻辑斯蒂回归：软分类

$f(\mathbf{x}_n) = y'_n = P(+1|\mathbf{x}_n) \in [0,1]$

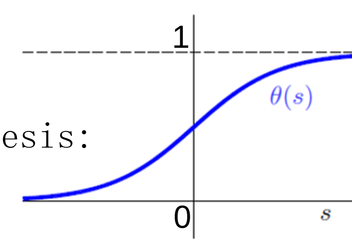
$\hat{y}_n = sign(P(+1|\mathbf{x}_n) - 0.5) \in \{1,-1\}$

$\mathbf{x} = (x_1, x_2, \ldots x_d)^T$

$s = \sum_{i=0}^{d} w_i x_i$

logistic hypothesis:

$h(\mathbf{x}) = \theta(\mathbf{w}^T\mathbf{x})$



## 5.2 逻辑斯蒂回归损失（引）

**逻辑斯蒂回归的最佳解：**

$$g = \underset{\boldsymbol{w}}{argmin} \frac{1}{N} \sum_{n=1}^{N} -ln\theta(y_n \boldsymbol{w}^T \mathbf{x}_n)$$

$$\theta(y_n \boldsymbol{w}^T \mathbf{x}_n) = \frac{1}{1 + \exp(-y_n \boldsymbol{w}^T \mathbf{x}_n)}$$

$$g = \underset{\boldsymbol{w}}{argmin} \frac{1}{N} \sum_{n=1}^{N} \ln(1 + exp(-y_n \boldsymbol{w}^T \mathbf{x}_n))$$

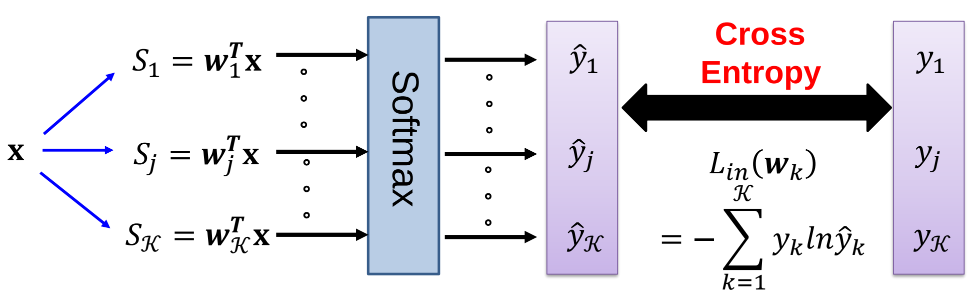
**交叉熵损失**
**(*Cross-Entropy Loss*)**

$$\boxed{L_{in} = \frac{1}{N} \sum_{n=1}^{N} \ln(1 + exp(-y_n \boldsymbol{w}^T \mathbf{x}_n))}$$



## 9.3 “Softmax”多类别分类

$\mathcal{Y} = \{1,2, \dots, j, \dots \mathcal{K}\}$




$if\ \ \mathbf{x} \in class\ 1$

$$\mathbf{y} = \begin{pmatrix} 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$L_{in}(\boldsymbol{w}_1) = -ln\hat{y}_1$$

$if\ \ \mathbf{x} \in class\ j$

$$\mathbf{y} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$L_{in}(\boldsymbol{w}_j) = -ln\hat{y}_j$$

$if\ \ \mathbf{x} \in class\ \mathcal{K}$

$$\mathbf{y} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

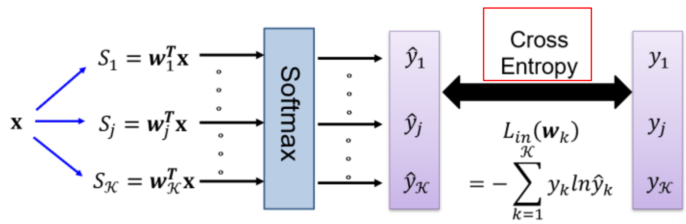
$$L_{in}(\boldsymbol{w}_{\mathcal{K}}) = -ln\hat{y}_{\mathcal{K}}$$

## 9.3 “Softmax”多类别分类

$$s_j = \boldsymbol{w}_j^T \vec{x} \qquad \hat{y}_j = \frac{e^{s_j}}{\sum_k e^{s_k}}$$

$$L_{in}(\boldsymbol{w}_k) = -\sum_{k=1}^{K} y_k ln\hat{y}_k = -ln\hat{y}_k$$

$$\frac{\partial L_{in}}{\partial \boldsymbol{w}_j} = \frac{\partial L_{in}}{\partial \hat{y}_k}\frac{\partial \hat{y}_k}{\partial S_j}\frac{\partial S_j}{\partial \boldsymbol{w}_j} = -\frac{1}{\hat{y}_k}\boxed{\frac{\partial \hat{y}_k}{\partial S_j}}\mathbf{x}$$




$$\frac{\partial \hat{y}_k}{\partial S_j} = \frac{\partial}{\partial S_j}\left(\frac{e^{s_k}}{\sum_k e^{s_k}}\right) = \frac{(e^{s_k})' \sum_k e^{s_k} - (\sum_k e^{s_k})' e^{s_k}}{(\sum_k e^{s_k})^2} =$$

$$\begin{cases} \dfrac{e^{s_j}\sum_k e^{s_k} - e^{s_j}e^{s_j}}{(\sum_k e^{s_k})^2} = \dfrac{e^{s_j}}{\sum_k e^{s_k}} - \dfrac{e^{s_j}}{\sum_k e^{s_k}}\dfrac{e^{s_j}}{\sum_k e^{s_k}} = \hat{y}_j(1-\hat{y}_j) & j = k \\ \dfrac{0\sum_k e^{s_k} - e^{s_j}e^{s_k}}{(\sum_k e^{s_k})^2} = 0 - \dfrac{e^{s_j}}{\sum_k e^{s_k}}\dfrac{e^{s_k}}{\sum_k e^{s_k}} = -\hat{y}_j\hat{y}_k & j \neq k \end{cases}$$

$$\frac{\partial L_{in}}{\partial \boldsymbol{w}_j} = \frac{\partial L_{in}}{\partial \hat{y}_k}\frac{\partial \hat{y}_k}{\partial S_j}\frac{\partial S_j}{\partial \boldsymbol{w}_j}$$

$$= -\frac{1}{\hat{y}_k}\frac{\partial \hat{y}_k}{\partial S_j}\mathbf{x}$$

$$= \begin{cases} (\hat{y}_j - 1)\mathbf{x} & j = k \\ \hat{y}_j\mathbf{x} & j \neq k \end{cases}$$



## 9.3 “Softmax”多类别分类

| | OVO,OVA | Softmax |
|---|---|---|
| Attributes | Overlap between classes | No overlap between classes |
| Examples | Indoor scene、<br>Gray images、<br>People photos | Indoor scene、<br>Outdoor urban scene、<br>Outdoor wilderness scene |
| | Vocal music、<br>Dance music、<br>Movie music、<br>Pop Song | Classical music、<br>Country music、<br>Rock music、<br>Jazz |

## 第九讲 多类别分类(*Classification for Multiclass*)

9.1 “一对多”策略的多类别分类(Multiclass via One-Versus-All)

*样本分类到概率最大的类别*

9.2 “一对一”策略的多类别分类(Multiclass via One-Versus-One)

*通过投票机制确定样本所属类别*

9.3 “Softmax”多类别分类(Multiclass via Softmax)

*用交叉熵(cross entropy)作为损失函数求解分类面*