## Softmax

**定义：**训练集 \{ (x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)}) \}， y^{(i)} \in \{1, 2, \ldots, k\}。

**假设函数:**


\begin{align}
h_\theta(x^{(i)}) =
\begin{bmatrix}
p(y^{(i)} = 1 | x^{(i)}; \theta) \\
p(y^{(i)} = 2 | x^{(i)}; \theta) \\
\vdots \\
p(y^{(i)} = k | x^{(i)}; \theta)
\end{bmatrix}
=
\frac{1}{ \sum_{j=1}^{k}{e^{ \theta_j^T x^{(i)} }} }
\begin{bmatrix}
e^{ \theta_1^T x^{(i)} } \\
e^{ \theta_2^T x^{(i)} } \\
\vdots \\
e^{ \theta_k^T x^{(i)} } \\
\end{bmatrix}
\end{align}


**代价函数：**


\begin{align}
J(\theta) = - \frac{1}{m} \left[ \sum_{i=1}^{m} \sum_{j=1}^{k}  1\left\{y^{(i)} = j\right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{ \theta_l^T x^{(i)} }}\right]
\end{align}


**“冗余”的参数集：**


\begin{align}
p(y^{(i)} = j | x^{(i)} ; \theta)
&= \frac{e^{(\theta_j-\psi)^T x^{(i)}}}{\sum_{l=1}^k e^{ (\theta_l-\psi)^T x^{(i)}}}  \\
&= \frac{e^{\theta_j^T x^{(i)}} e^{-\psi^Tx^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}} e^{-\psi^Tx^{(i)}}} \\
&= \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{ \theta_l^T x^{(i)}}}.
\end{align}


**权重衰减**


\begin{align}
J(\theta) = - \frac{1}{m} \left[ \sum_{i=1}^{m} \sum_{j=1}^{k} 1\left\{y^{(i)} = j\right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{ \theta_l^T x^{(i)} }}  \right]
              + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2
\end{align}


**和逻辑回归的关系：**


\begin{align}
h(x) &=

\frac{1}{ e^{\vec{0}^Tx}  + e^{ (\theta_2-\theta_1)^T x^{(i)} } }
\begin{bmatrix}
e^{ \vec{0}^T x } \\
e^{ (\theta_2-\theta_1)^T x }
\end{bmatrix} \\


&=
\begin{bmatrix}
\frac{1}{ 1 + e^{ (\theta_2-\theta_1)^T x^{(i)} } } \\
\frac{e^{ (\theta_2-\theta_1)^T x }}{ 1 + e^{ (\theta_2-\theta_1)^T x^{(i)} } }
\end{bmatrix} \\

&=
\begin{bmatrix}
\frac{1}{ 1  + e^{ (\theta_2-\theta_1)^T x^{(i)} } } \\
1 - \frac{1}{ 1  + e^{ (\theta_2-\theta_1)^T x^{(i)} } } \\
\end{bmatrix}
\end{align}


**注意点：**

最小化代价函数，同样可以采用简单而有效的梯度下降，需要提到的是，在程序实现中，我们一般采用批量随机梯度下降，即MSGD，minibatch Stochastic Gradient Descent，简单来说，就是每遍历完一个batch的样本才计算梯度和更新参数，一个batch一般有几十到几百的单个样本。PS：随机梯度下降则是一个样本更新一次。