A função custo $J(\theta)$ para uma rede neural com K classes, n features, m input rows, L layers, com s_ℓ neurônios em cada layer, e na qual aplicamos regressão logística, será dada por:

OBS: se $K=1$ temos apenas 1 classe e $y=0$ ou 1, ao que chamamos de problema de classificação binário. No entanto, se $K \geq 3$, temos um problema de classificação multiclasses (multi-class classification).

OBS2: se $K=2$ pode-se reduzir ao problema binário.

$$\boxed{J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \cdot \log\left(h_\theta(x^{(i)})\right)_k + \left(1 - y_k^{(i)}\right)\cdot \log\left(1-\left(h_\theta(x^{(i)})\right)_k\right)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(\theta_{ji}^{(l)}\right)^2}$$

// aplicando em Octave:

→ given $m, K, L, y(m,K), h=a^{(L)}(K,m)$, e $\overbrace{s(L)}^{\text{vetor c/ \# neurônios em cada camada}}, \lambda$, e $\text{Theta}(\underbrace{\max(s)+1}_{\text{maior \# de linhas em uma matriz } \theta^{(l)}}, \underbrace{m+1}_{\text{maior \# de colunas em uma matriz } \theta^{(l)}}, \underbrace{L-1}_{\text{\# matrizes } \theta})$

```
J=0;
for i=1:m
    for k=1:K
        J=(-1/m)*(y(i,k).*log(a(k,i))+(1-y(i,k)).*log(1-a(k,i))) + J;
    endfor
endfor
// próxima parte calculará o custo levando em conta as regularizações dos
// parâmetros.
```