L = total number of layers in the network

sl = number of units(not counting bias unit) in layer I

k = number of output units/classes

cost function

J(theta) = -1/m ∑i=1m∑k=1K[y(i)klog((hΘ(x(i)))k)+(1−y(i)k)log(1−(hΘ(x(i)))k)]

calculate error

 \delta^{(L)} = a^{(L)} - y^{(t)}*δ*(*L*)=*g(z(l))*−*y*(*t*)

*calculate error before*

*δ*(*l*)=((Θ(*l*))*Tδ*(*l*+1)) .∗ *a*(*l*) .∗ (1−*a*(*l*))

*g*′(*z*(*l*))=*a*(*l*) .∗ (1−*a*(*l*))

Δ(*l*):=Δ(*l*)+*δ*(*l*+1)(*a*(*l*))*T*

*Di*,*j*(*l*)​:=*m*1​(Δ*i*,*j*(*l*)​+*λ*Θ*i*,*j*(*l*)​), if j≠0.

∂Θ*ij*(*l*)​∂​*J*(Θ)= D\_{ij}^{(l)}*Dij*(*l*)​

A picture containing standing, baseball, person, player

Description automatically generated