

Binary Cross Entropy / Log loss (Classification)

$$J = \frac{1}{N} \sum_{i=0}^{N-1} (y_i \cdot \ln(a_i) + (1-y_i) \cdot \ln(1-a_i))$$

over N batches

$$\frac{\partial J}{\partial a_i} = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{y_i}{a_i} + (1-y_i) \cdot \frac{1}{1-a_i} \cdot (-1) \right)$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{y_i}{a_i} - \frac{(1-y_i)}{(1-a_i)} \right)$$

$$= \frac{y_i - \cancel{y_i a_i} - a_i + \cancel{y_i a_i}}{a_i(1-a_i)}$$

$$\Rightarrow \frac{\partial J}{\partial a_i} = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{y_i - a_i}{a_i(1-a_i)} \right)$$

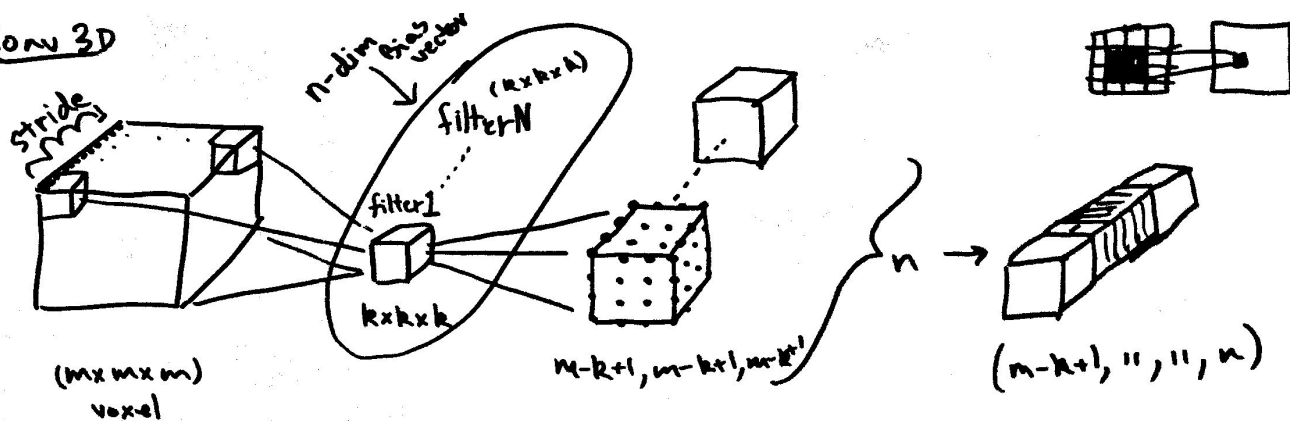
name: model(layerinfo, α , decay, optimization,
regularization,
 x, y)

$L[]$

size = # of layers.

$L[i]$ = "Activation", nL

Conv 3D



lets say $m = 12$ and $k = 3$ then $\text{output} = \frac{12}{3} = 4$
 $(12 \times 12 \times 12)$ $(3 \times 3 \times 3)$ $(4 \times 4 \times 4)$

Concatenate + Padding (P)

$(m-k+1+2P, m-k+1+2P, n(s))$

filterN → for any 1 filter,

input (m, m, m)

mask
 $(k \times k \times k)$

Conv

bias
 $(1, 1, 1)$

non-Linearity
 (function)

output
 $(m-k+1, \dots, n)$