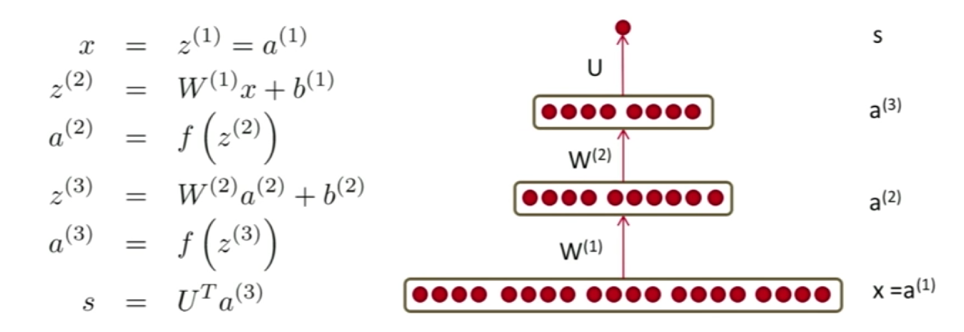
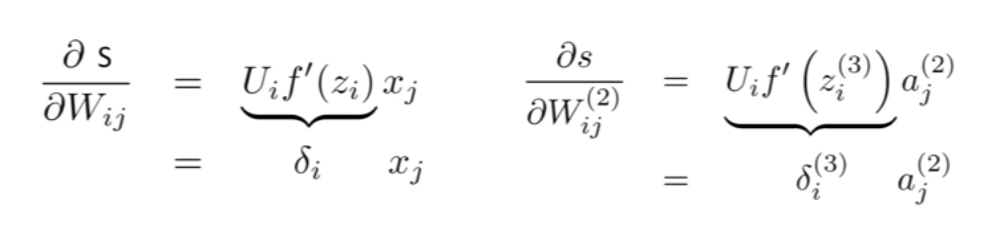
Lecture 5 | Backpropagation

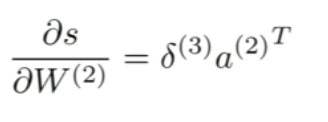
* 4 different explanation of backpropagation
  + 2 layers neural nets and full backprop
    - Same window definition for x and same scoring function but now with 2 hidden layers



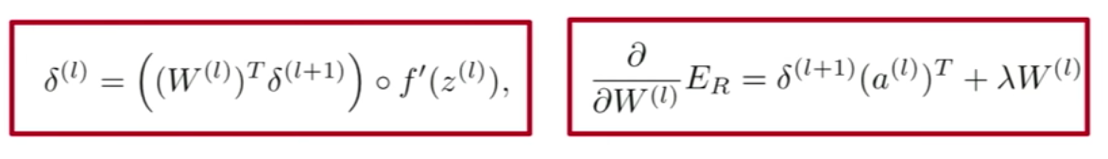
* + - The figure above shows the forward propagation of 2 layers NN
    - Now for back propagation, we essentially have the same derivation as before for W(2):



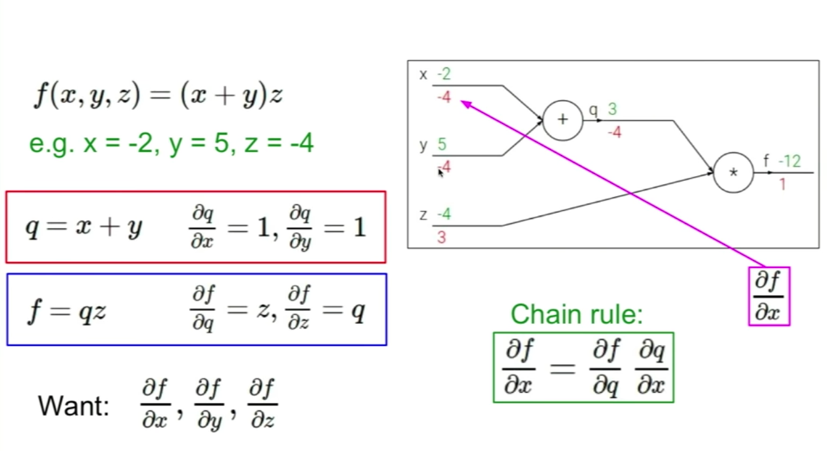
* + - For matrix notation (vectorisation form), back propagation becomes:

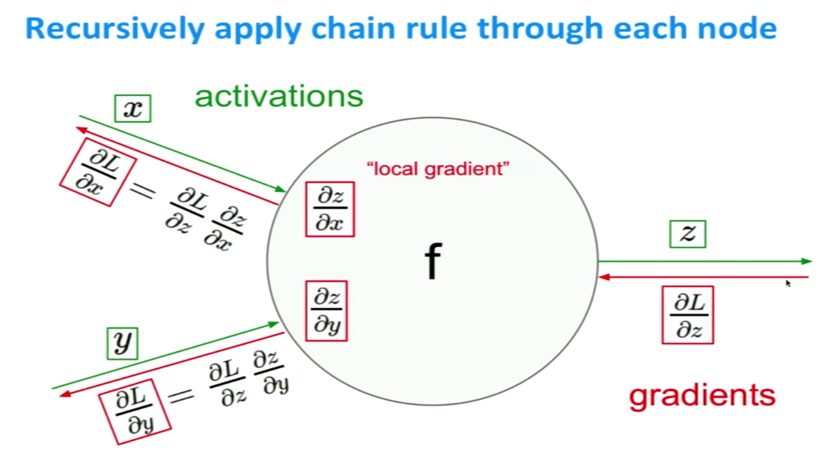


* + - For general backprop:
      * In general, for any matrix W(l) at internal layer l and any error with regularisation E(r), all backprop in standard multilayer neural networks boils down to 2 equations:

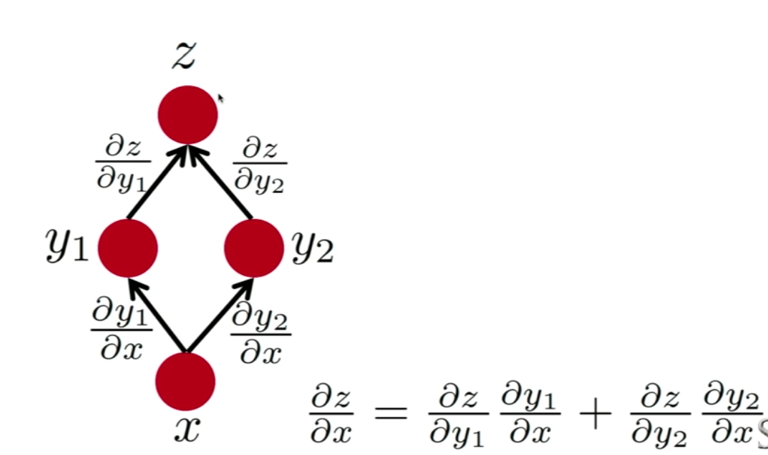


* + Circuits explanation of backprop (prefer this explanation)

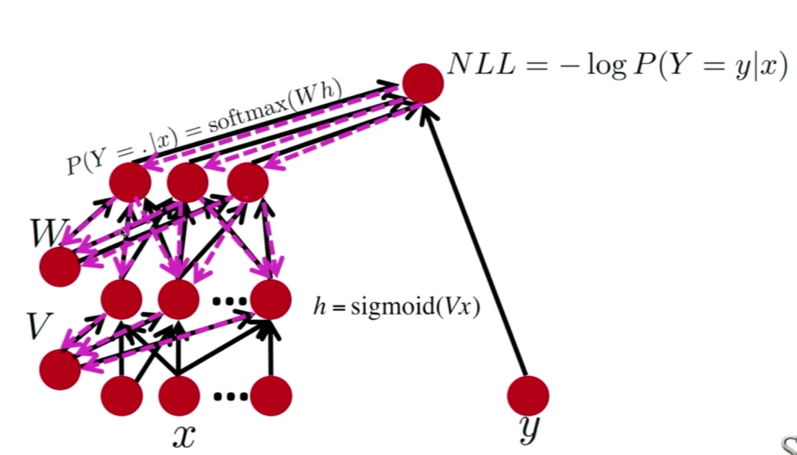




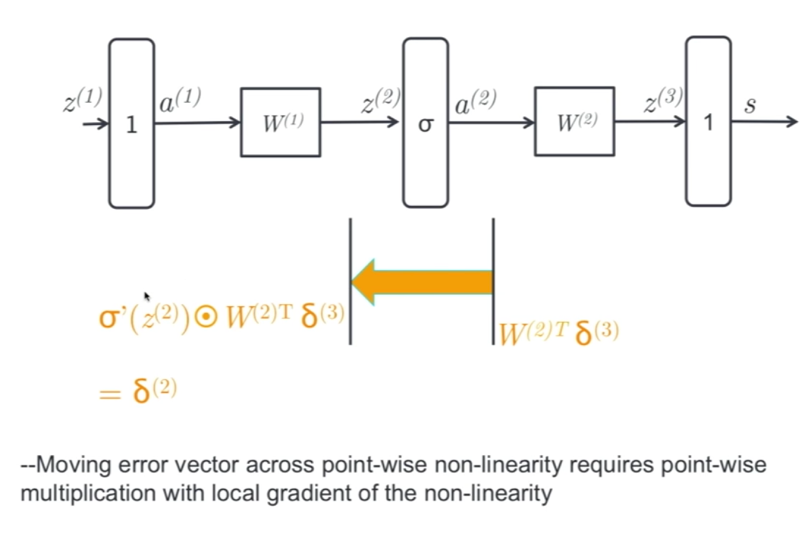
* + The high-level flowgraph explanation



* + - Back-prop in multi-layer net



* + The delta error signals in real neural nets



* + - Going from the outer layer to W(2), we will just be reusing the delta(3) for the downstream updates.
    - Note that moving error vector across affine transformation only requires multiplication with the transpose of forward matrix
    - To backprop from W(2) to W(1) – figure above! Using delta(2), we can compute the gradient w.r.t. W(1), where you multiply delta(2) by activation(1)\_transpose