forward propergation.

- unactivated input for hidden layer.

$$Z^{(1)}_{7\times 419} = W^{(1)}_{7\times 4} A^{(0)}_{4\times 419} + b^{(1)}_{7\times 1} \rightarrow ①$$

- $A^{(0)} = X^{T}_{4\times 419}$ Transpose of train set
- $A^{(1)} = Relu(Z^{(1)}) = g(Z^{(1)}) \rightarrow ②$ feed to activation function.
- $Z^{(2)}_{1\times 419} = W^{(2)}_{1\times 7} A^{(1)}_{7\times 419} + b^{(2)}_{1\times 1} \rightarrow ③$. unactivated input for output layer.
- $A^{(2)} = Z^{(2)}$ → use Linear Activation, to output layer. because desired output is price unit.

→ loss function (L) $= \sum_{i=1}^{m} (y_i - \hat{y})^2 / m$. = MSE. m = 419.

Backward propergation. (by using chain Rule)

- $dA2 = \partial L / \partial A_2 \rightarrow 2(y_i - \hat{y})/m$.
- $dZ2 = dA_2$.
- $dW_2 = \partial L/\partial Z2 \cdot \partial Z2/\partial W2$

$$dW_2 = dA2 \left(\frac{\partial (W^{(2)} A^{(1)} + b)}{\partial W_2} \right) = dA_2 \cdot A^T$$

- $db2 = \partial L/\partial Z^2 \cdot \partial Z^2/\partial b^2$. $\partial Z^2/\partial b^2 = 1$

$$db_2 = \partial L/\partial Z^2 = \sum dZ2.$$

b_2 shared across all examples ∴ Total effect. is the sum of effects.

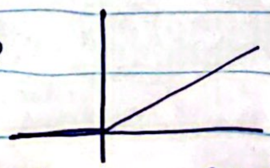
- $dA_1 \rightarrow \partial L/\partial A_1 = \partial L/\partial Z^2 \cdot \partial Z^2/\partial A_1$ *

$$= dZ2 \cdot W^T.$$

- $dZ_1 \rightarrow \partial L/\partial Z_1 = \partial L/\partial A_1 \cdot \partial A_1/\partial Z_1$

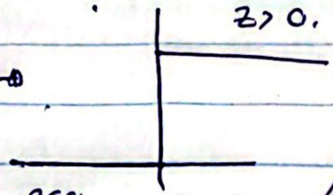
$$= dA_1 \cdot (Z > 0).$$

Relu →



Z<0 Z>0

derivative of relu →



Z>0. Z<0

$\cdot)\ dw_1 \Rightarrow \partial L/\partial z^1 \cdot dz^1/dw_1 = dz_1 \cdot x^T.$

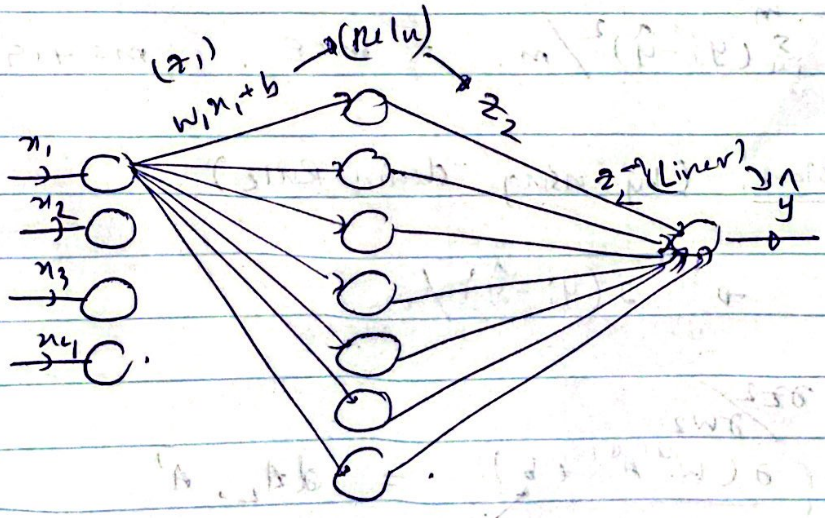
$\cdot)\ db_1 \Rightarrow \partial L/\partial b_1 = \partial L/\partial z_1 \cdot \partial z_1/\partial b_1 = \Sigma dz_1.$

weights update in a ~~m~~ way in back propagation.

When $\boxed{W_{new} = W_{old} - \alpha\, \partial L/\partial W_{old}}$

α = Learning Rate.

Architecture of the neural network



$loss = (y - \hat{y})^2.$