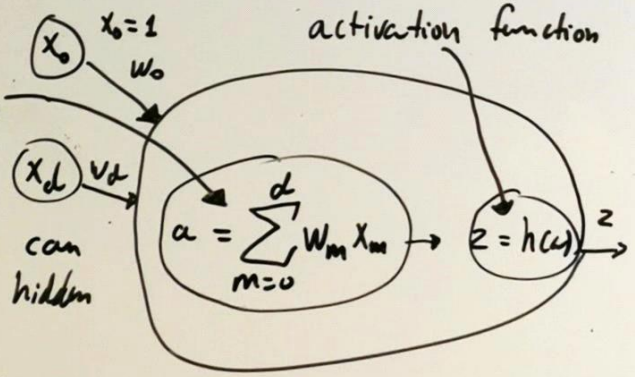
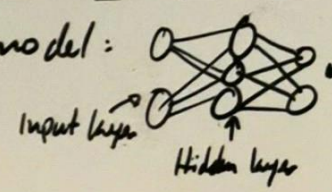
- Q8 Explain the general model of MLP and their expressive power. Mention an algorithm to train them!

Anatomy Neuron/Perceptron:



-General model:



- Universal Approximation Theory: Any C function can be approximated arbitrary well by atleast one hidden layer MLP with finite number of weights.
- Some function perform more efficiently with deeper layers.
- Activation functions: Sigmoid, RELu, sign, tanh

- Backpropagation: Optimize weights so we minimize the total loss

$L(w) = \frac{1}{N} \sum_{n=1}^{N} L_n(w)$ → No closed form for minimizer (L_n = Regression/Classification)

Gradient descent to update weights: $w \leftarrow w - \eta \nabla_w L(w)$, "backpropagation" is a technique for calc gradient effectively.

- Chain rule essential !! 4 step process: Forward pass → Output sensitivity → Hidden layer sensitivities → Weight gradients: $\frac{\partial L_n}{\partial w_{ji}^{(l)}} = \delta_{n,j}^{(l)} z_{n,i}^{(l-1)}$; $z_{n,j}^{(l)} = h\left(a_{n,j}^{(l)}\right)$

Previous layer

Sensitivity of neuron

Activation

C = continious.