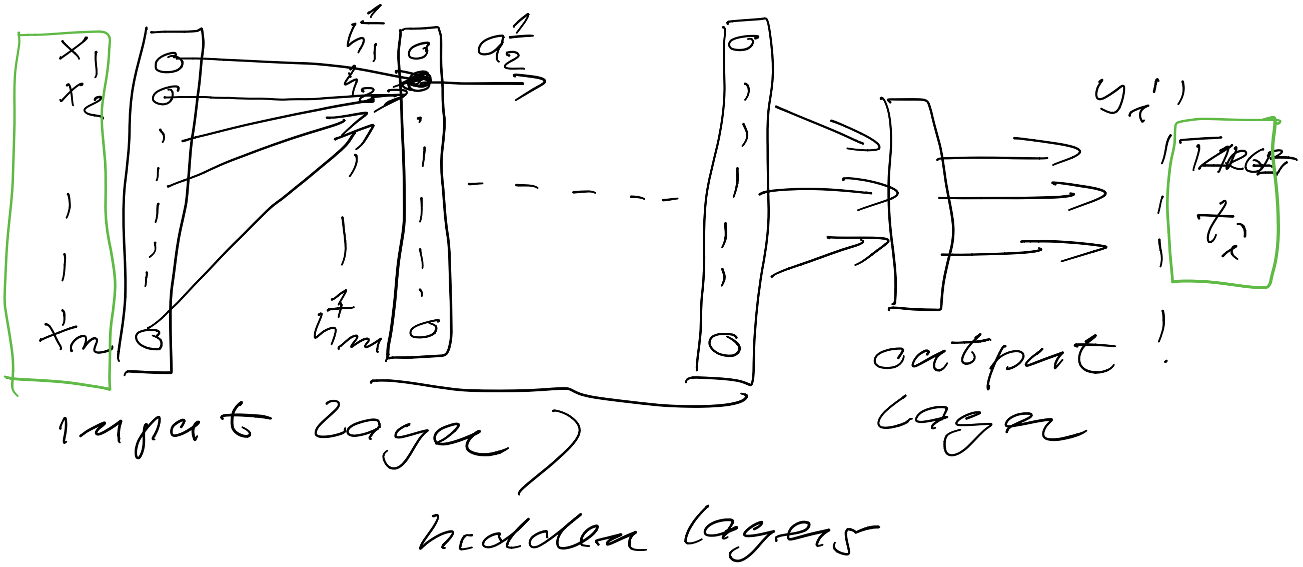
BIRD'S VIEW ON NN



- input layer (# nodes)
- output layer (# nodes)
- # hidden layers (# nodes)
- weights w_{ij}^{l} l = layer,
- biases b_i^l

Random initial values

- output from every node

$$a_i^l = f(z_i^l)$$

$$z_i^l = \sum_j w_{ij}^l a_j^{l-1} + b_i^l$$

- activation function,
- To find optimal weights & biases which the cost function (our model)

- Back propagation algo (matrix-vector multiplications)
- $\frac{\partial C}{\partial w} \wedge \frac{\partial C}{\partial b} \Rightarrow$ gradient descent optimizer, ($\boxed{\text{SGD, Adam}}$, ...)
  - Learning rate $\boxed{\xi, \gamma}$
- Hyperparameters $\boxed{\lambda}$