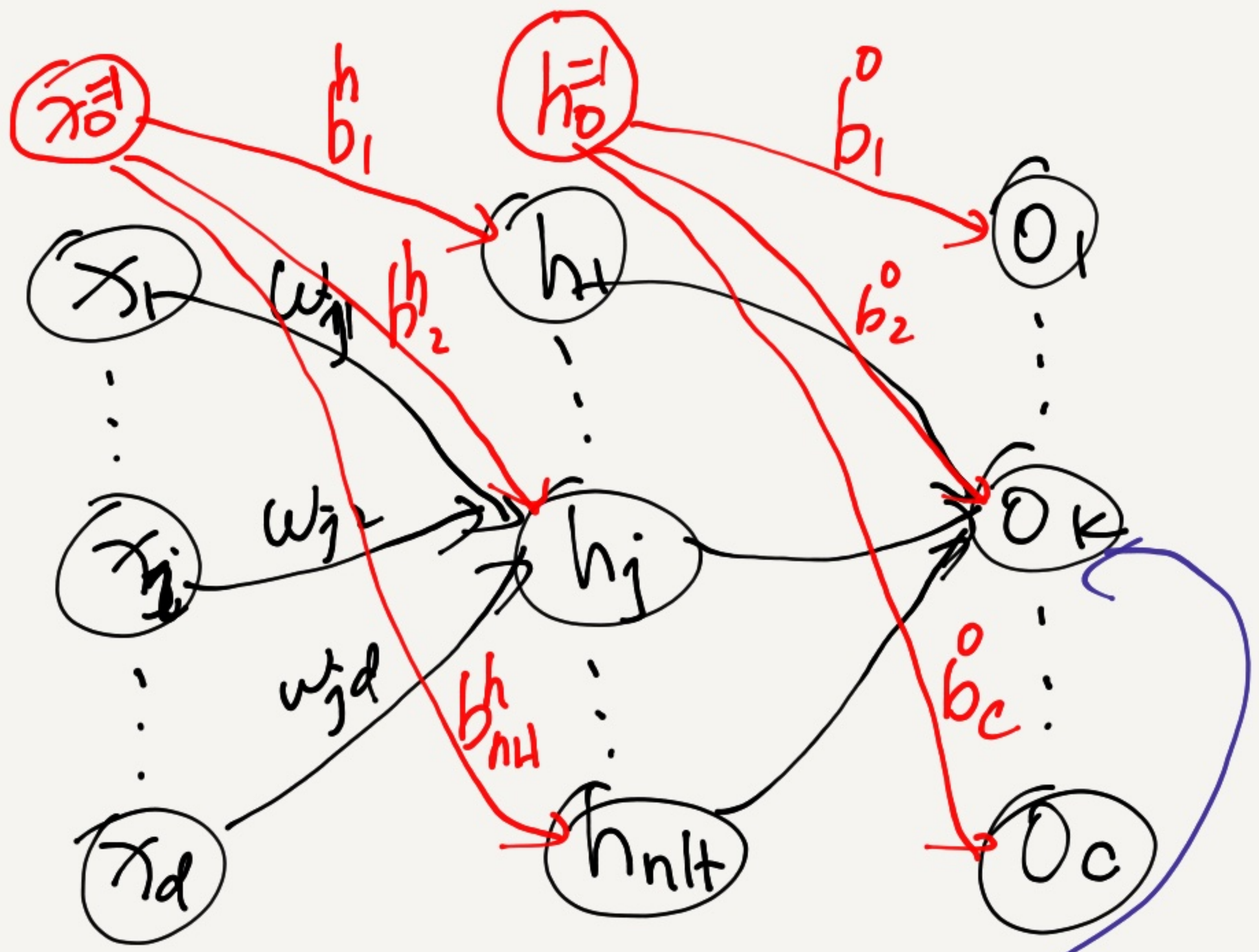
Lecture 11 : NN Implementation

1. Bias nodes in NNs

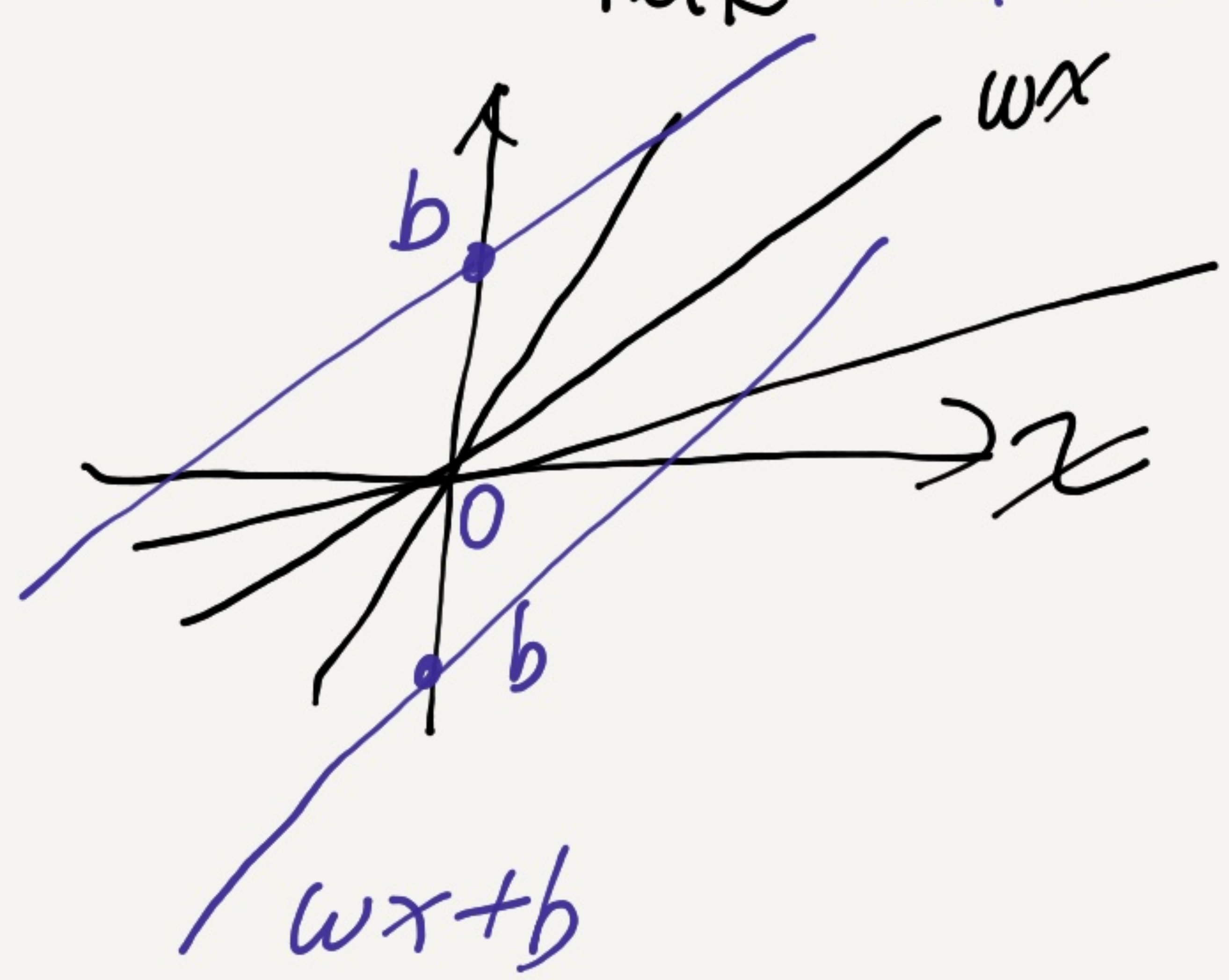
$$\nabla_{u_{kj}} L = \delta_k \cdot h_j \rightarrow \nabla_{b_k^o} L = \delta_k \Leftarrow (h_j = h_0 = 1)$$

$$\nabla_{w_{ji}} L = \delta_j \cdot x_i \rightarrow \nabla_{b_j^h} L = \delta_j \Leftarrow (x_i = x_0 = 1)$$



Net input : $net_j = w_{j1} \cdot x_1 + w_{j2} \cdot x_2 + \cdots + w_{jd} \cdot x_d + b_j$

$net_k = u_{k1} \cdot h_1 + u_{k2} \cdot h_2 + \cdots + u_{kn_H} h_{n_H} + b_k$



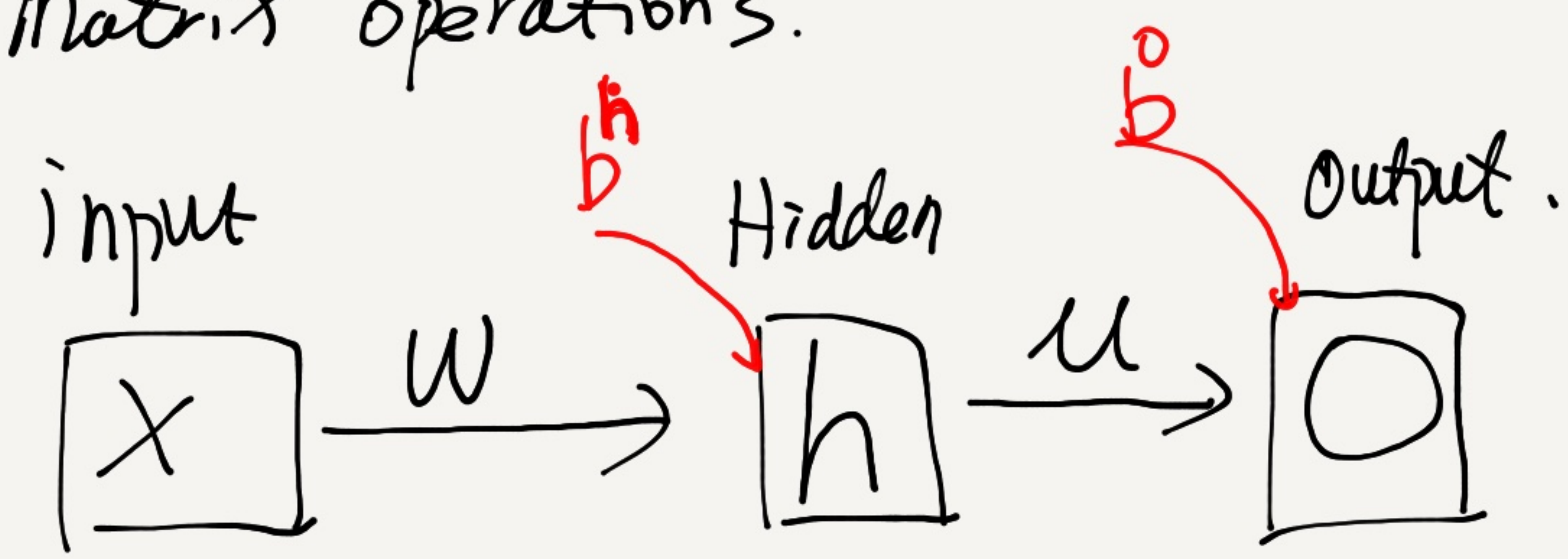
without bias nodes:

$d \times n_H + n_H \times C$

with the bias nodes:

$(d+1) \times n_H + (n_H+1) \times C$

2 Use Matrix operations.



first sample.

first hidden node.

$$X = \begin{bmatrix} x_{11}, x_{12}, \cdots x_{1d} \\ \vdots \quad\quad \vdots \\ x_{n1}, x_{n2}, \cdots, x_{nd} \end{bmatrix} n \times d \qquad W = \begin{bmatrix} w_{11}, w_{12}, \cdots w_{1d} \\ \vdots \\ w_{n_H 1}\, w_{n_H 2}, \cdots w_{n_H d} \end{bmatrix} n_H \times d \qquad b^h = \begin{bmatrix} b^h_1 \\ b^h_2 \\ \vdots \\ b^h_{n_H} \end{bmatrix} \qquad b^o = \begin{bmatrix} b^o_1 \\ b^o_2 \\ \vdots \\ b^o_c \end{bmatrix}$$

n: # of data samples $\qquad$ n_H: # of hidden nodes.

d: # of features. $\leftarrow d$

Output of the hidden layer: $h = g_h(net_h)$ $\quad net_h = X \cdot W^T + (b^h)^T$

Output of the last layer: $O = g_o(net_o)$ $\quad net_o = h \cdot u^T + (b^o)^T$

3. Two settings

1) Regression problem.

universal.

determined by the loss function

$$\mathcal{L} = \frac{1}{2}\sum_{k=1}^{C}(y_k - o_k)^2 \longrightarrow \boxed{\begin{aligned}\nabla_{u_{kj}}\mathcal{L} &= \delta_k \cdot h_j \\ \nabla_{w_{ji}}\mathcal{L} &= \delta_j \cdot x_i\end{aligned}}$$

$$\underline{\underline{\delta_k}} = \boxed{(o_k - y_k)} \cdot g_o'$$

$$\underline{\underline{\delta_j}} = \sum_{k=1}^{C} \underline{\underline{\delta_k}} \cdot u_{kj} \cdot g_h'$$

2) classification problem

$$\mathcal{L} = -\sum_{k=1}^{C} y_k \cdot \log o_k \quad (\text{cross-entropy})$$

g_o: softmax

$$\boxed{\begin{aligned}\delta_k &= (o_k - y_k). \\ \delta_j &= \left(\sum_{k=1}^{C} \delta_k \cdot u_{kj}\right) \cdot g_h'\end{aligned}}$$