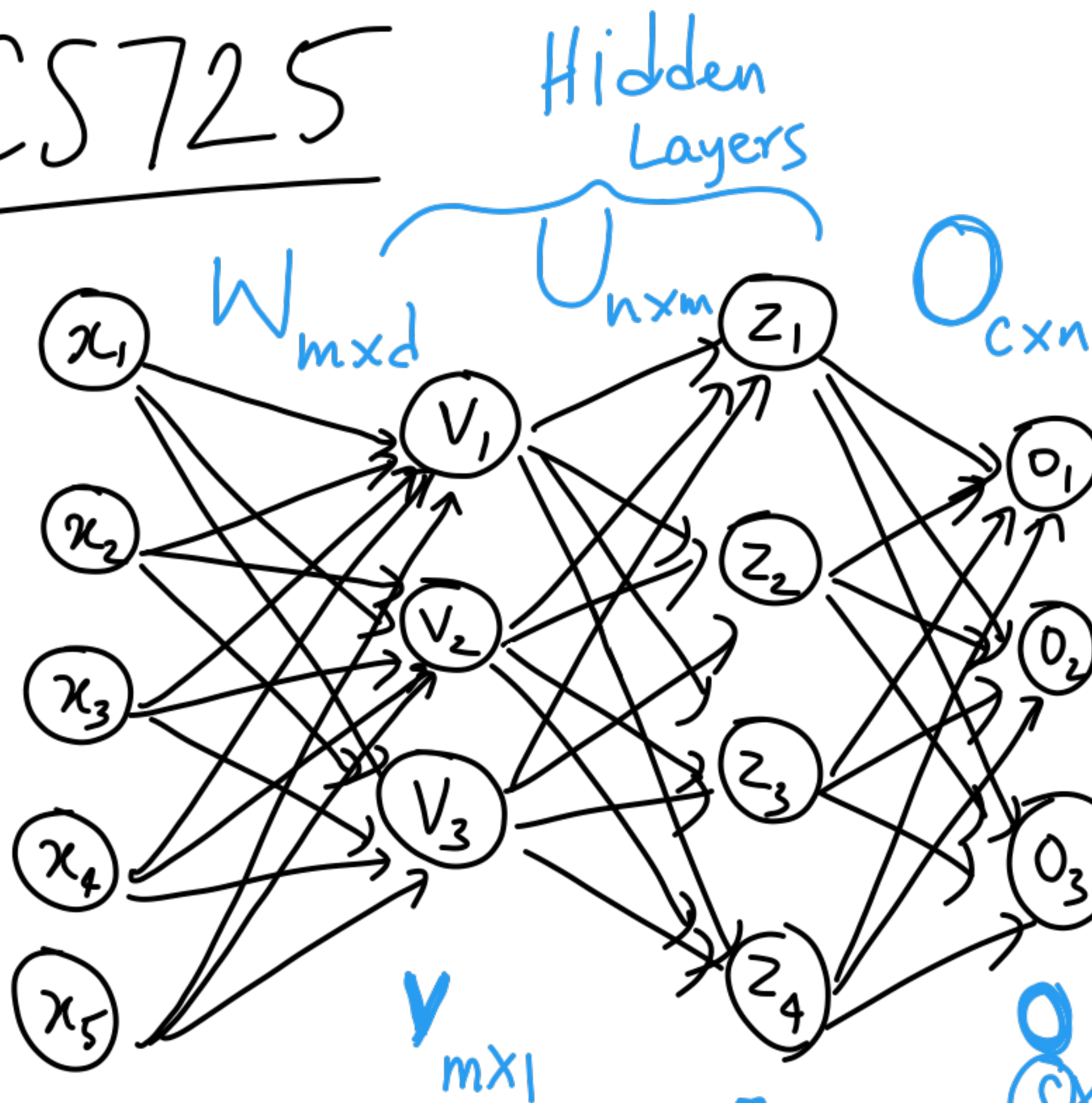


CS725



Parameters  $\theta$  of this neural network are:

$W, U, O, b, b'$   
 $[o_1, \dots, o_c]$  are called LOGITS for a class (task)

$$P(y=j|x) = \text{Softmax}(o_j) = \frac{e^{o_j}}{\sum_i e^{o_i}}$$

INPUT LAYER  
 $x_{d \times 1}$

Where  $C$  is the number of classes

$$Z = g(Uv + b') = g(Ug(Wx + b) + b')$$