



Figure 1: Network graph of a  $(L + 1)$ -layer perceptron with  $D$  input units and  $C$  output units. The  $l^{\text{th}}$  hidden layer contains  $m^{(l)}$  hidden units.