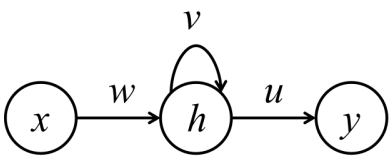
Quiz 6	**Name:** _____
Week 8, Nov/2/2022	**Number:** _____
CS 280: Fall 2022	**On your left:** _____
Instructor: Lan Xu	**On your right:** _____

Problem 1. (10 points) We have a simple recurrent neural network with the following network structure, and for simplicity let us assume all the variables and operations in activation functions are linear:



- Draw the feed-forward network when unfolding the recurrent network through time for 2 time step $t = 1, 2$.
- Given a training sample $(\hat{y}_1, \hat{y}_2)$ as the ground truth for (y_1, y_2), derive $\frac{\partial L}{\partial w}$ using back propagation based on the graph you have built.
- Please explain the two kinds of gradient problems in RNN training, using the above case as an example.

Problem 2. (10 points) *Transformer*.

1. Draw the diagram of Transformer encoder and decoder.
2. Describe the multi-head attention mechanism. (You need to formulate the computation, i.e., K, Q, V)