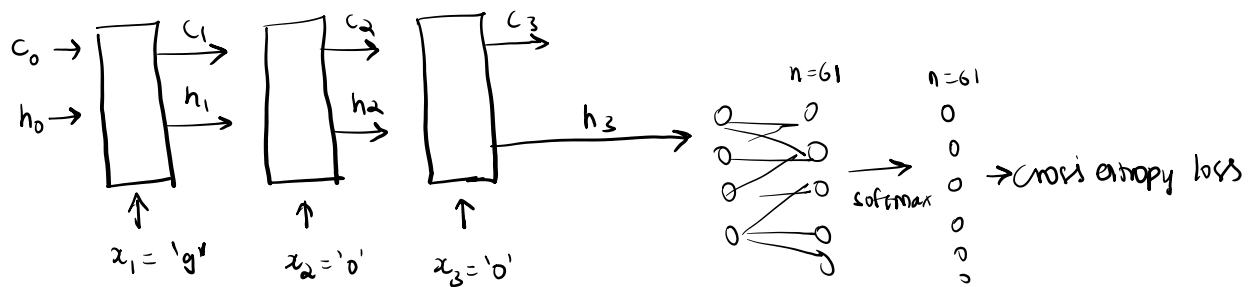


Predict Next character:

Model 1: (Predict next char at last time step)

Training:

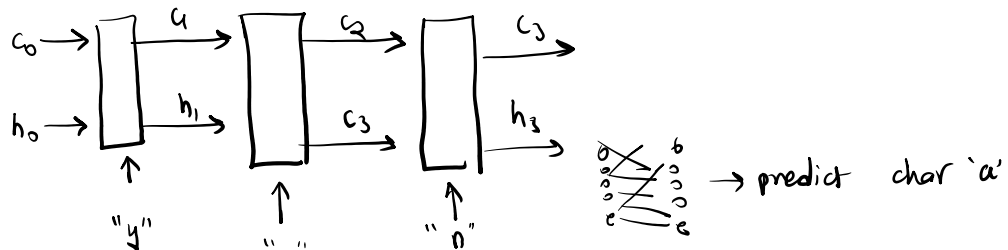
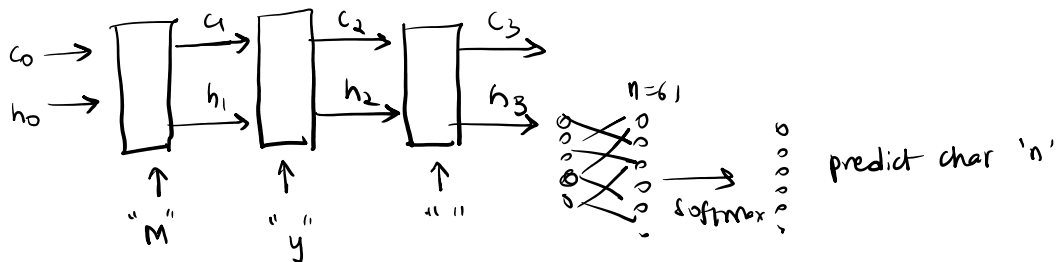
context = ['g', 'o', 'o'], vocab-size = 61, next-char = 'd'



Inference:

prompt = "My "

Type 1:

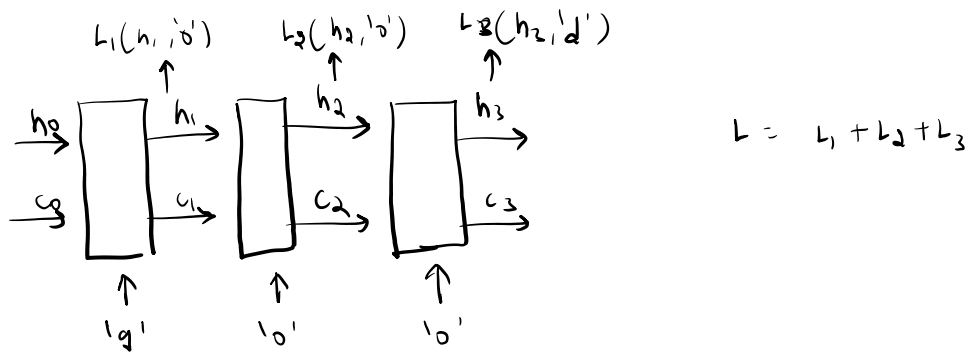


and it follows

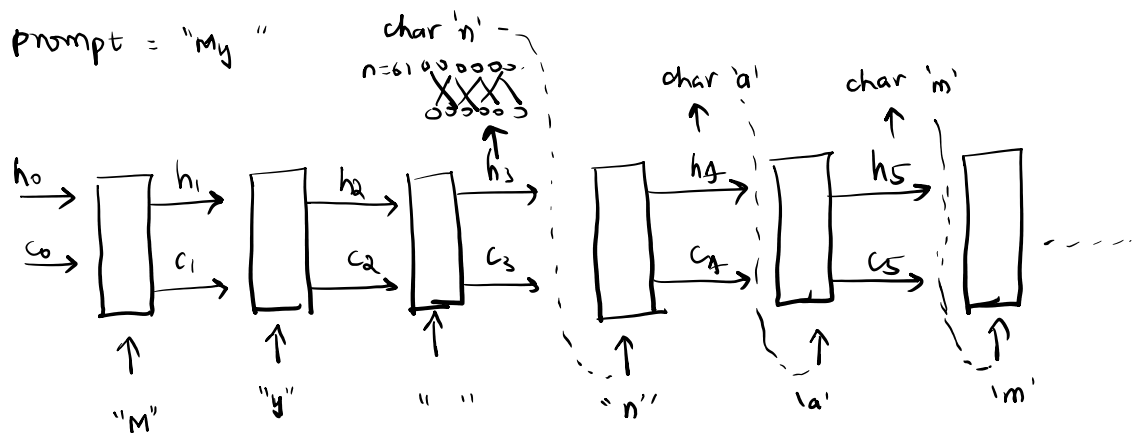
Model 2: (predict the next char at each time step)

Training

$$X = \{ 'g', 'o', 'o' \} \quad Y = \{ 'o', 'o', 'd' \}$$



Inference :



→ This inference can also be used in Model 1

→ In some sense model 2 training is more competitive, because it pushes to learn the char with less context