JichenDai_HW#1

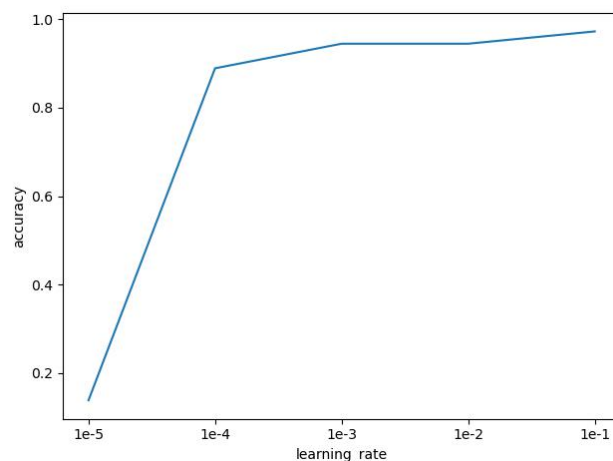## Build a Neural Network on Wind Dataset

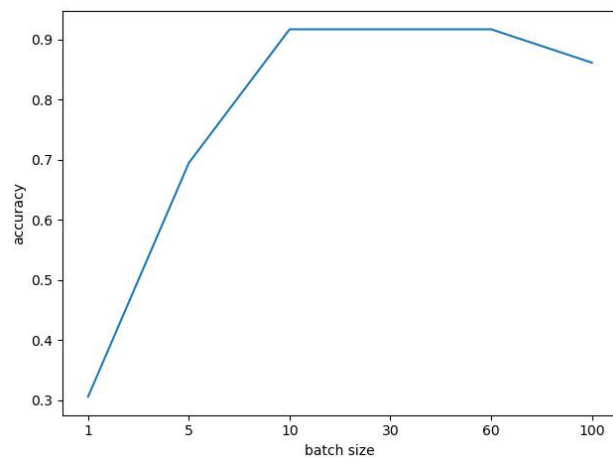## step 3

batch size = 10
learning rate = 1e-3
Accuracy = 0.9444444179534912

Question 1: optimal learning rate and batch size

1.1 optimal learning rate=0.01 (epochs = 200, batch size = 10)
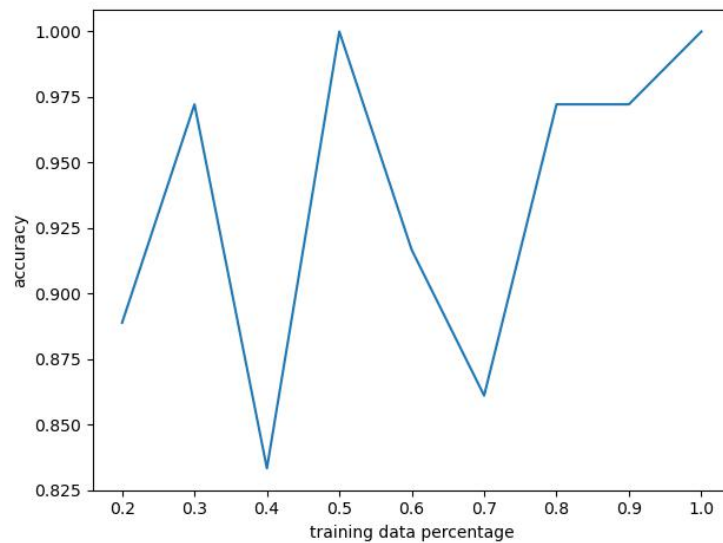

1.2 optimal batch size=10 (epochs = 200, learning rate = 0.1)
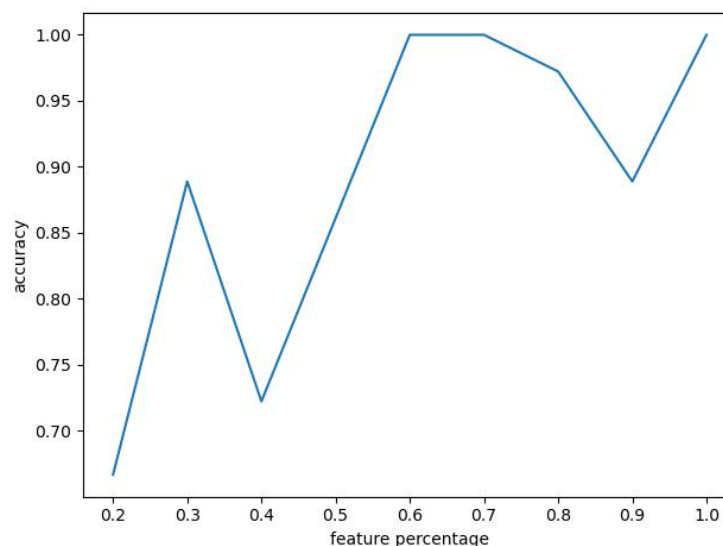
Question 2: different percentage of training data

learning rate = 1e-3, batch size = 30, epochs = 200

Question 3: different percentage of feature

learning rate = 1e-3, batch size = 30, epochs = 200
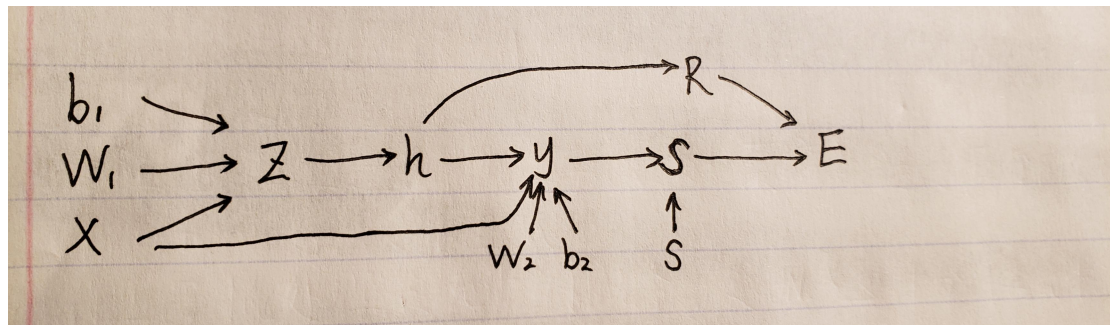

Hard_Coding a Network

1. give weights and bias

W1 = [[-1, 1, 0, 0],
        [0, -1, 1, 0],
         [0, 0, -1, 1]]

b1 = [-1, -1, -1]

W2 = [1, 1, 1]

b2 = -3

2.1 computation graph

2.2 Derive

Derive:

$$\frac{dE}{dx} = \frac{d(R+S)}{dx} = \frac{dR}{dx} + \frac{dS}{dx}$$

Since $\quad \frac{dR}{dx} = \frac{dR}{dh} \cdot \frac{dh}{dz} \cdot \frac{dz}{dx}$

$$= r^T \cdot \delta'(z) \cdot W_1$$

● and $\quad \frac{dS}{dx} = \frac{dS}{dy} \cdot \frac{dy}{dx} = \frac{dS}{dy} \cdot \frac{d(x + W_2 h + b)}{dx} = \frac{dS}{dy} \cdot \left(\frac{dx}{dx} + \frac{dW_2 h}{dx} \right)$

$$= \frac{dS}{dy} \cdot \left(1 + W_2 \frac{dh}{dz} \cdot \frac{dz}{dx} \right)$$

$$= (y-s) \cdot (1 + W_2 \cdot \delta'(z) \cdot W_1)$$

So, $\quad \frac{dE}{dx} = \frac{dR}{dx} + \frac{dS}{dx} = r^T \cdot \delta'(z) \cdot W_1 + (y-s) \cdot [1 + W_2 \cdot W_1 \cdot \delta'(z)]$