Backpropagation in practice

Puteaux, Fall/Winter 2020-2021

```
################################
##                            ##
##  Deep Learning in Python   ##
##                            ##
################################
```

§1 Introduction to Deep Learning in Python

§1.2 Optimizing a neural network with backward propagation

1 Backpropagation in practice

1.1 How to calculate slopes associated with any weight?

- Gradients for weight is the product of:
    1. node value feeding into that weight
    2. slope of activation function for the node being fed into
    3. slope of the loss function with respect to the output node
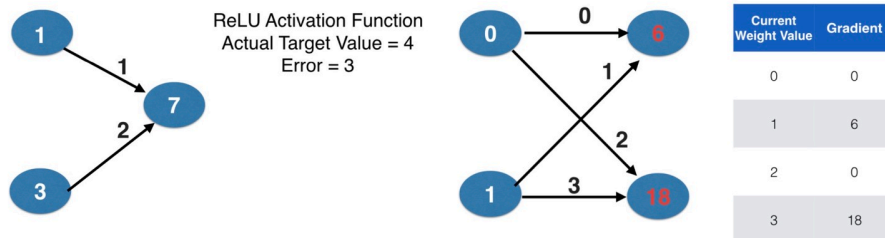
## 1.2 To recap backpropagation, what are the steps?

- Start at some random set of weights.

- Use forward propagation to make a prediction.

- Use backward propagation to calculate the slope of the loss function with respect to each weight.

- Multiply that slope by the learning rate, and subtract from the current weights.

- Keep going with that cycle until got a flat part.

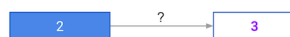## 1.3 What are the principles of stochastic gradient descent?

- What are the principles of stochastic gradient descent?

- It is common to calculate slopes on only a subset of the data (a batch).

- Use a different batch of data to calculate the next update.

- Start over from the beginning once all data is used.

- Each time through the training data is called an epoch.

- When slopes are calculated on one batch at a time, it is so-called stochastic gradient descent.

1.4 Practice question for a round of backpropagation:

- In the network shown below, forward propagation was done, and node values calculated as part of the forward propagation are shown in white. The weights are shown in black. Layers after the question mark show the slopes calculated as part of back-prop, rather than the forward-prop values. Those slope values are shown in purple.

- This network again uses the ReLU activation function, so the slope of the activation function is 1 for any node receiving a positive value as input. Assume the node being examined had a positive value (so the activation function's slope is 1).
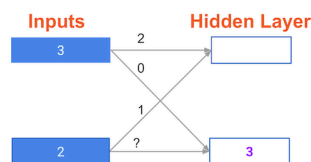
- What is the slope needed to update the weight with the question mark?

☐ 0.

☐ 2.

☒ 6.

☐ Not enough information.