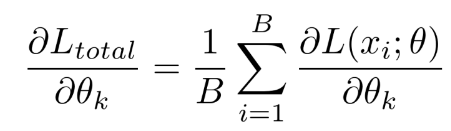
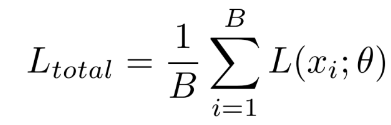
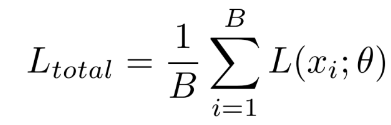
Quiz #3 Hwang Jihyun

**1.What does B mean in the formula below?**



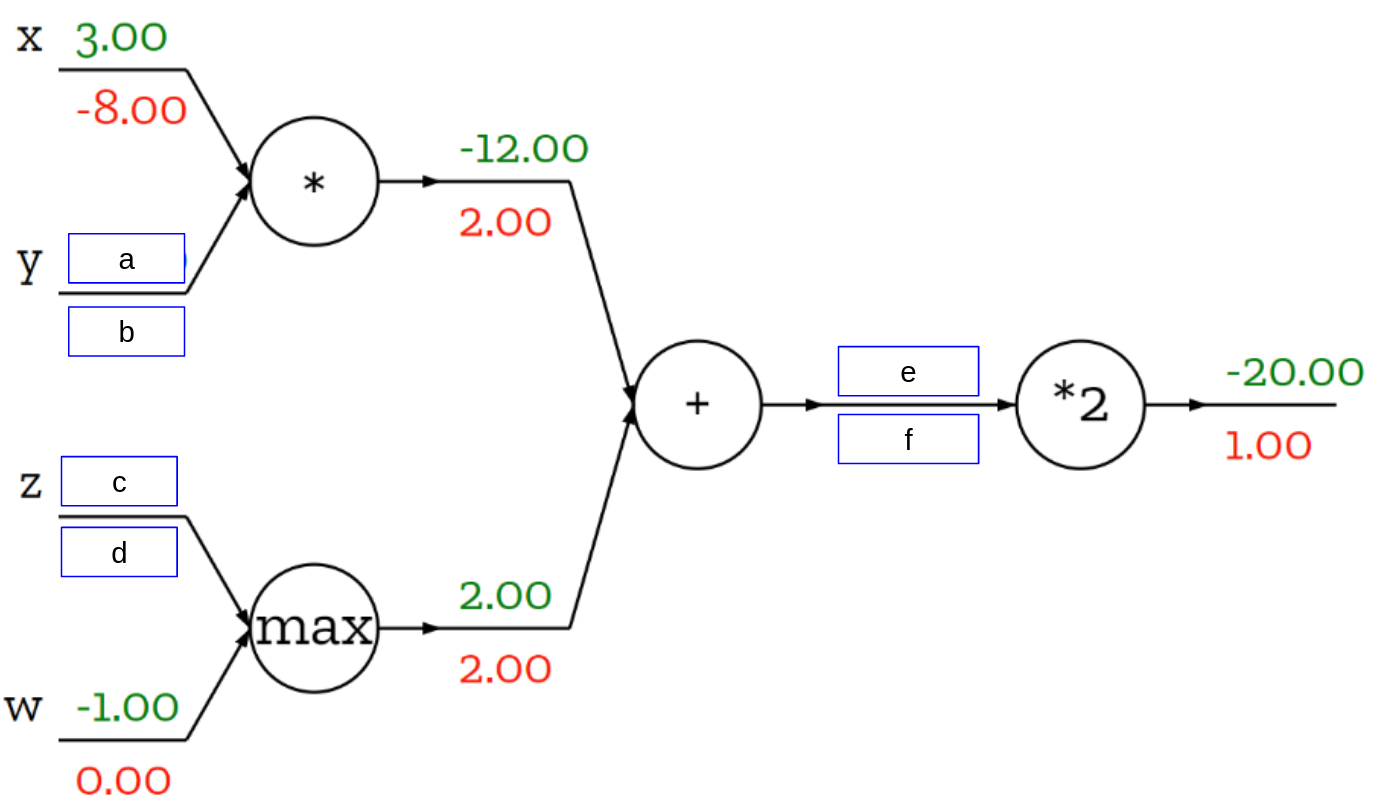
1. Backpropagation weights
2. **Batch(or mini batch) size**
3. Batch-normalization layer size
4. Backward path count

**2.What does**  **not mean in the formula below?**



1. Losses on one parameter
2. **Losses on all data**
3. **Losses on one iteration**
4. **Losses on one data**
5. Losses on one batch(or mini batch)

**3. Fill in the blanks (Backward pass term)**



| **answer** | **a** | **b** | **c** | **d** | **e** | **f** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | -4 | 6 | 2 | 1 | 10 | 1 |
| **2** | 4 | 6 | -2 | 1 | -10 | 2 |
| **3** | 4 | -4 | -2 | 2 | 10 | -1 |
| **4** | -4 | 6 | 2 | 2 | -10 | 2 |
| **5** | -4 | -4 | 2 | 2 | 10 | 1 |

**4. How to set the starting(initial) point in the gradient descent method?**

1. Random selection of points where the slope is expected to be large
2. Random selection of points where the slope is expected to be small
3. Set to start from both ends of the data
4. **Set the starting point at random each time**
5. Set randomly only at the beginning, and then repeatedly executed at the same point

**5. Using mini-batch reduces the time it takes to lengthen or lose data when there is more data, so why not set the batch size to 1?**

1. **If set the batch size to 1, It's hard to reflect loss across the data**
2. If set the batch size to 1, It took too long to complete one iteration
3. **If set the batch size to 1, it's difficult to reduce the effect on the outlier**
4. If set the batch size to 1, The calculation result becomes too small