Parameters explanation

1. num\_epochs: 1 epoch = 1 entire passing of all training data through the algorithm/model. More epochs mean the algorithm has gone through the training data many times and may potentially learn more information from the data (but need to be aware of overfitting). Each epoch will take approximately the same time so total training time can be estimated by num\_epochs\*time\_per\_epoch

Common values: num\_epochs = 20, 30, 50 ,100, ...

2. batch\_size: the number of training examples utilized in one iteration (within the for loop). A smaller batch size allows the model to learn from each individual example but takes longer time and more memory to train.

Common values: batch\_size = 8, 16, 32, 64, ...

3. learning\_rate: a parameter in an optimization algorithm to determine the step size at each iteration while moving toward a minimum of a loss function. Basically it represents the speed at which a machine learning model "learns". High learning rate will make the learning jump over minima so that accuracy may not be good and a low learning rate will lead to longer training time.

Common values: any number less than 0.1 and greater than 10^-6

4. regularization: a penalty term to the loss function to prevent overfitting. Large regularization may be misleading if regularization overweighs loss function. Small regularization may not be enough to adjust loss function and prevent overfitting.

Common values: any number less than 0.1 and greater than 10^-6

5. smoothing\_factor: a parameter to implement label smoothing in loss function. Label smoothing replaces the original labels with smoothed distributions, which encourages the model to be less certain and makes it more robust to noisy or incorrect labels. The value of smoothing factor should be determined by the data quality, if just a few incorrect labels exist, then small smoothing factor should be adopted.

Common values: 0.05, 0.1, 0.2, ...