

# Question: Iterative Learning/Backprop

## Contents

Is it fair to assume that the error surface changes depending on the input? The error surface of a model is generally seen as something like the image below in figure 1. Since our goal is to find the ML solution of the weights given data, we need to see as much error as possible and find the correct weights. We do this by using iterative updates in backpropagation. And we say this is okay to do in a mini-batch setting, despite the fact that we aren't using all the data, due to computational efficiency. However, technically each set of data would give a random error surface. All the error surfaces would come from the same distribution of error surfaces, given the model, but each mini-batch would be different. And when we iteratively update the error values, we don't know the "true" minimum... ever. So mini-batches are fine because a full-batch pass would only give a sample of the "true" minimum anyway. Therefore, for the sake of generalization and computational resources, we would use mini-batches.

How correct is the above statement?

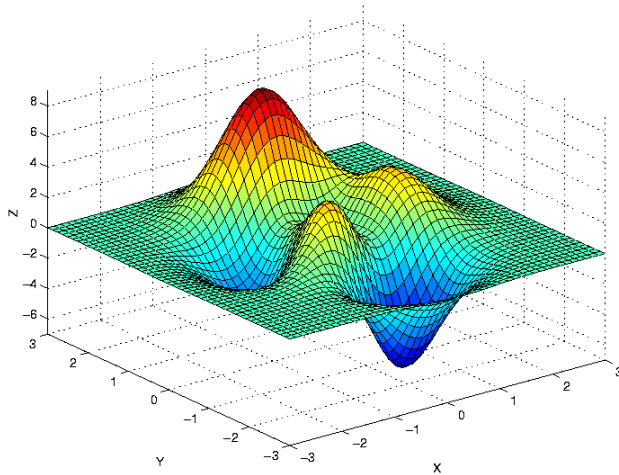


Figure 1: Imagine  $z$ -axis =  $C(w)$ ,  $y$ -axis =  $w_0$ , and  $x$ -axis =  $w_1$ , where  $C(w) = w_0 + x_1 * w_1$ .