THE STEP SIZE (LEARNING RATE α)

The learning rate is the size of each step the network takes when it descends the error

mountain, and it is usually denoted by the Greek letter alpha (α). It is one of the most

important hyperparameters that you tune when you train your neural network (more

on that later). A larger learning rate means the network will learn faster (since it is

descending the mountain with larger steps), and smaller steps mean slower learning.

Well, this sounds simple enough. Let’s use large learning rates and complete the neural network training in minutes instead of waiting for hours. Right? Not quite. Let’s

take a look at what could happen if we set a very large learning rate value.

In figure 2.32, you are starting at point A. When you take a large step in the direction of the arrow, instead of descending the error mountain, you end up at point B,

on the other side. Then another large step takes you to C, and so forth. The error will

keep oscillating and will never descend. We will talk more later about tuning the learning rate and how to determine if the error is oscillating. But for now, you need to

know this: if you use a very small learning rate, the network will eventually descend the

metin, diyagram, grafik, harita içeren bir resim

Açıklama otomatik olarak oluşturuldu

mountain and will get to the minimum error. But this training will take longer (maybe weeks or months). On the other hand, if you use a very large learning rate, the network might keep oscillating and never train. So we usually initialize the learning rate value to 0.1 or 0.01 and see how the network performs, and then tune it further