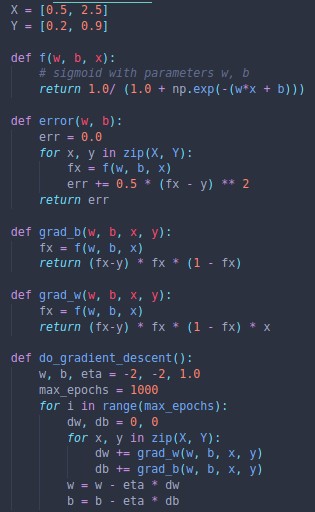
## **Highlighting a limitation of Gradient Descent**

Let’s look at better learning algorithms

1. The gradient descent update rule is as follows
   1. a
   2. The questions we should be asking are: How do we compute the gradients? What data should we use for computing the gradients?
   3. To follow up on those questions: How do we use the gradients? Can we come up with a better update rule?
   4. Here is the Python implementation of Gradient Descent similar to what we have seen earlier
2. Look at the following plot of w, b and error using Gradient descent 