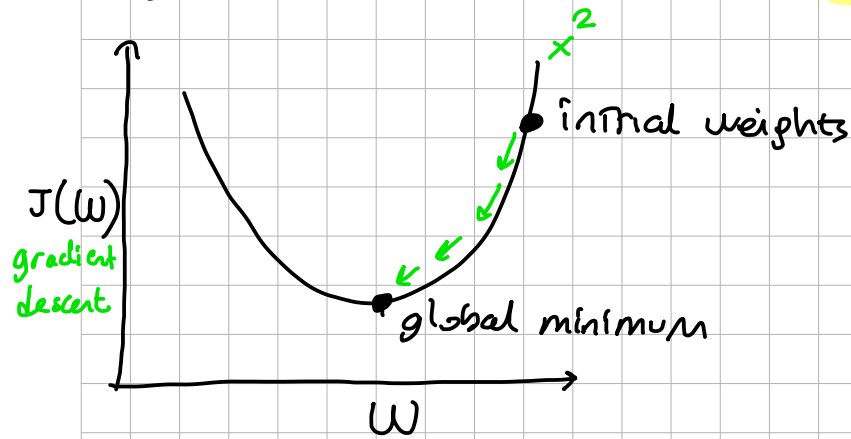
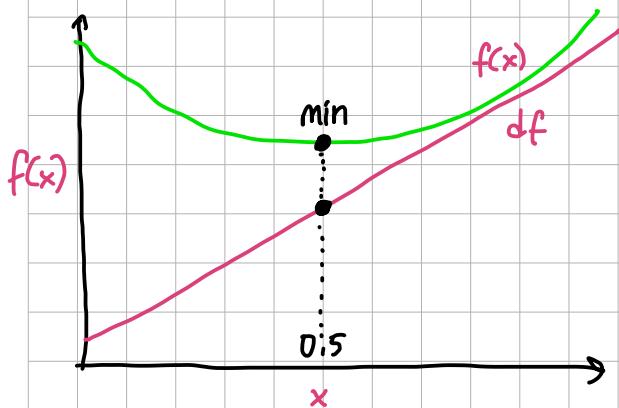


1) Gradient Descent

"get the smallest errors as possible."

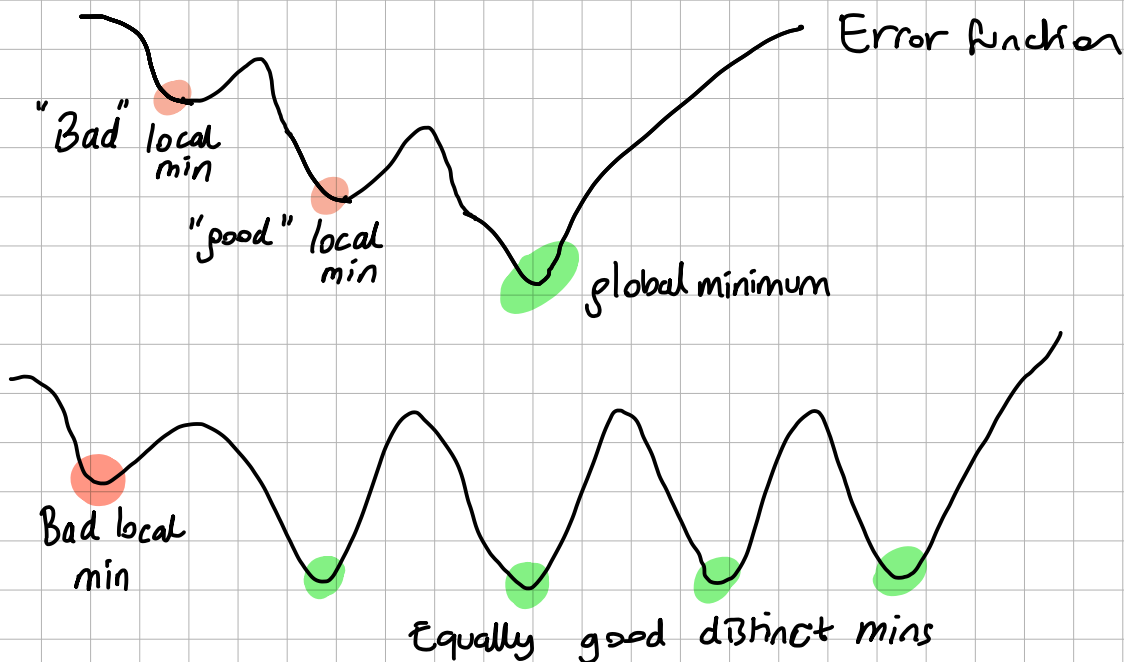


$$f(x) = 3x^2 - 3x + 4, \quad \frac{df}{dx} = 6x - 3 \quad (\text{derivative})$$



- gradient descent algorithm
- ① initialize random guess of minimum.
 - ② loop over training iterations
 - compute derivative at guess min.
 - updated guess min is itself minus derivative scaled by learning rate.

Gradient Descent algorithm is not guaranteed to give us the right solution.



Saddle point = Min in one direction & Max in other direction

When model performance is good, don't worry about "local minima".

- 1) Re-train the model many times using different random weights.
- 2) Increase the dimensionality of the model to have fewer local minima.