→ Gradient Descent is used all over the place in ML

Have some function $J(w,b)$ for linear regression or any function

Want $\min_{w,b} J(w,b)$ $\min_{w_1,\ldots,w_n,b} J(w_1,w_2,\ldots,w_n,b)$ ✓

Outline:

Start with some $w,b$ (set $w=0$, $b=0$)

Keep changing $w,b$ to reduce $J(w,b)$

Until we settle at or near a minimum

J not always

may have >1 minimum

Stanford ONLINE DeepLearning.AI Andrew Ng

→ Gradient descent is used to minimize any function. Not just linear func.

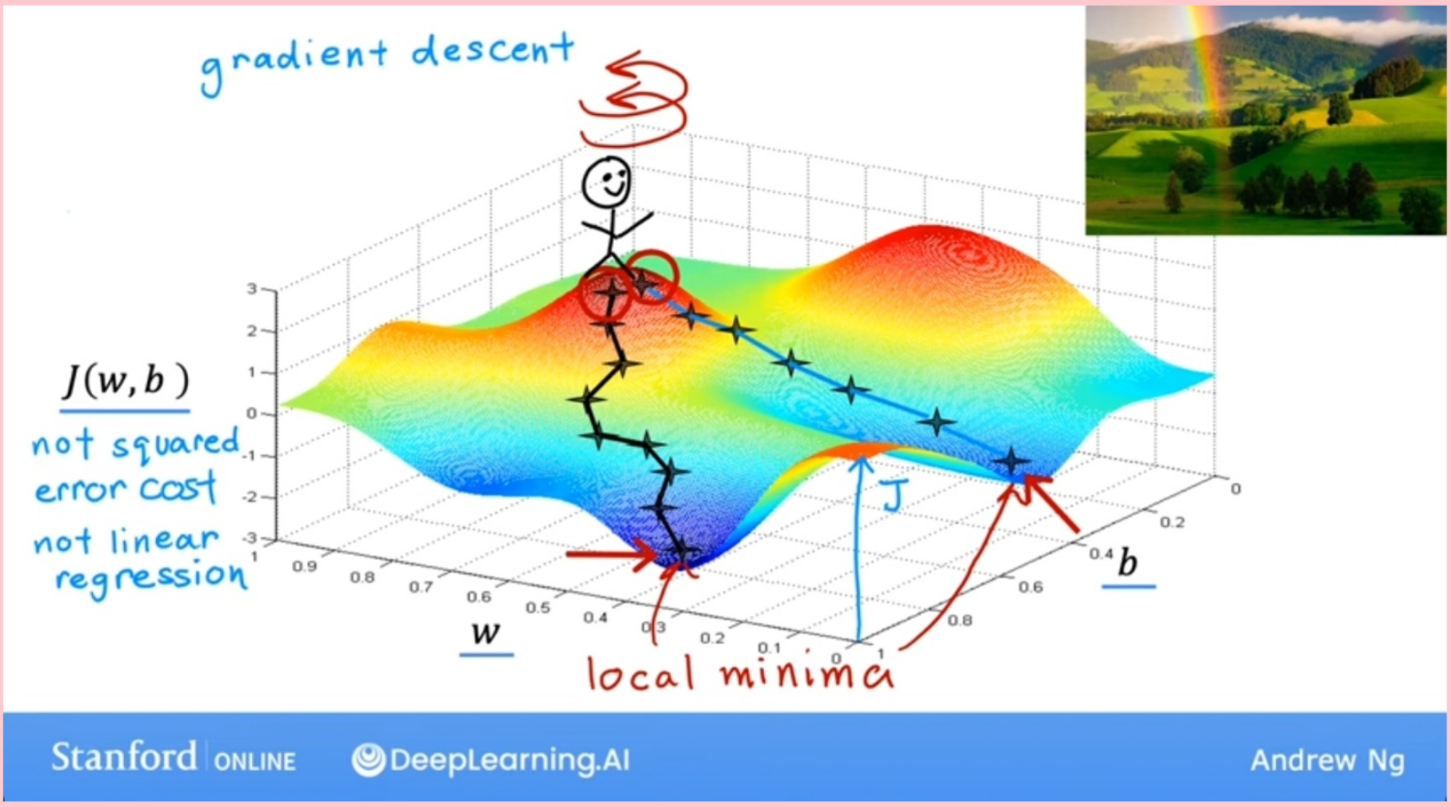
→ It starts with any value of w and b. We assume 0 here. It is a common choise.

→ Now we will change these values to reduce $J(w,b)$

→ Untill we settle at or near a minimum

→ For those function, whose 3D graph version is not bowl shape can have more than 1 minimum.

→ Example: gradient descent.




→ It is not squared error cost func

→ It shows how gradient descent works

→ At first we set the value of w, b

→ Then search for minimum sites.

→ Make a step toward that min point.

→ Then again search again make step.

→ Based on your initial choosing you may get more than 2 solution.

→ Those solutions are local minima for those intial choosing.