Se deriva la funcion, para encontrar max y min

Se da un peso inicio, y se busca el punto mas bajo, que da la mejor precision, menor error.

Como alguien que baja la montana y va buscando pasajes para ir descendiendo mas.

Y luego vuelve a intentar otro camino hasta encontrar el punto mas bajo.

En 3D es mas facil verlo, pero cuando hay mas de 3 variables no es posible observarlo.

Cuando learning rate

- Es alto, da pasos mas largos y es menos preciso.

- Es bajo, da pasos cortos, y puede lograr mejor precision, pero toma tiempo y computo.

|  |  |
| --- | --- |
|  | Pero en este caso, si el learning rate es muy bajo encontrara el minimo local y se estancara.  Si el learning rate es muy alto no encontrara ninguno de los dos minimos |

