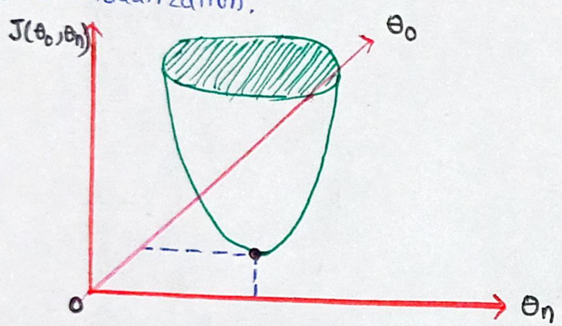
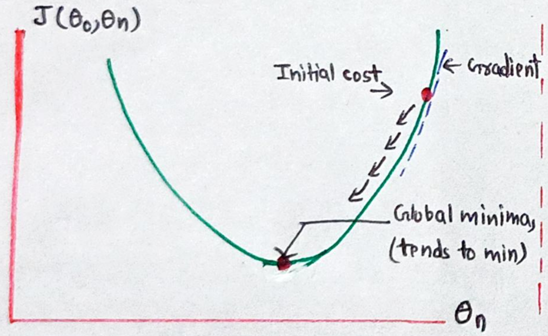
→ Gradient Descent

- Optimization technique to find minimum of any function.
- In Linear regression we want to find minima of our cost function.
- If we plot $J(\theta_n)$, θ_n, θ_0 we get this bowl shape, we can also plot $J(\theta_0,\theta_n)$ vs θ_n for 2D visualization.





- Where the slope of valley is going down, we move towards it, we repeat it till we reach the global minima and see that every contour around us is higher than where we are.

- This process or method of reaching global minima is called convergence.
- Convergence algorithm optimizes the value of slope i.e θ_n

$$\theta_n = \theta_n - \alpha \frac{\partial}{\partial \theta_n} (J(\theta_n,\theta_0))$$

where α is learning rate and is set by us, it basically means by how many data points we should move down the valley or converge.

- If α very high then, we will move out of data point spread or curve or may be jumping here their and not reaching global minima.
- If α too small it will be time consuming.
- Convergence algorithm updates the slope θ_n by a new value calculated by the formula above, which leads to global minima.
- Gradient Descent converges by:

1. We have to know where the slope of valley is going, to get it we partially differentiate the slope of function at that point, thus Gradient differentiates the cost function.

$$\frac{\partial}{\partial \theta_n}(J(\theta_0,\theta_n))$$

Let cost function be mean squared.

$$D\theta_n = \frac{\partial}{\partial \theta_n}\left[\frac{1}{n}\sum_{i=0}^{n}(h_\theta(x)^i - y^i)\right]^2$$

$$\Rightarrow -\frac{2}{n}\sum_{i=0}^{n} x^i (h_\theta(x)^i - y^i)$$

Similarly,

$$D\theta_0 = \frac{\partial}{\partial \theta_0}\left[\frac{1}{n}\sum_{i=1}^{n}(h_\theta(x)^i - y^i)\right]^2$$

$$\Rightarrow -\frac{2}{n}\sum_{i=1}^{n}(h_\theta(x)^i - y^i)$$

2. Now update the value as,

$$\theta_n = \theta_n - \alpha D\theta_n$$

$$\theta_0 = \theta_0 - \alpha D\theta_0$$

same as,

$$\boxed{\theta_n = \theta_n - \alpha \frac{\partial}{\partial \theta_n}(J(\theta_n,\theta_0))}$$

$$\boxed{\theta_0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0}(J(\theta_n,\theta_0))}$$

3. Repeat this untill reaches global minima, ideally 0.