## **More intuitions about Taylor Series**

Can we get the answer from some basic mathematics?

1. The real aim is:
   1. w ⇒ w + η𝚫w
   2. b ⇒ b + η𝚫b
   3. Loss(w) > Loss(w + η𝚫w)
   4. Loss(b) > Loss(b + η𝚫b)
   5. Loss(w, b) > Loss(w + η𝚫w, b + η𝚫b)
   6. Loss(𝜃) > Loss(𝜃 + η𝚫𝜃) (where 𝜃 = [w, b])
2. Vectorized Taylor Series:
3. Where, u = 𝚫𝜃
4. Here, we know that in practice, η is very small ie (0.001) etc
5. So η2 ,η3… all end up being negligible, so remove those corresponding terms
6. New Vectorized Taylor Series:
   1. Here, refers to Gradient w.r.t 𝜃 and it consists of the partial derivatives of L(𝜃) w.r.t w and b, stacked up into a vector
   2. ∈ ℝ
   3. ∈ ℝ
   4. ∈ ℝ
   5. uT() = Dot product of 𝚫𝜃 transposed and the partial derivative vector (∈ ℝ)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | * 1. [𝚫w | * 1. 𝚫b] | | |  | | --- | |  | |  | |