## **Deriving the Gradient Descent Update rule**

How does Taylor series help us arrive at the right answer?

1. For ease of notation, let 𝚫𝜃 = u
2. Then from Taylor series,we have:
   1. Rearranging:
   2. Note,that the move would only be favourable if
      1. (i.e. if the new loss is less than the previous loss)
      2. This implies
   3. Now we have
      1. Let be the angle between u and , then we know that,
      2. Multiply throughout by k =
      3. This gives us
   4. Thus, will be most negative when , i.e. when β is 180o
3. Gradient Descent Rule
   1. The direction u that we intend to move in should be at 180o w.r.t, the gradient
   2. In other words, move in a direction opposite to the gradient
4. Parameter Update Rule
   1. wt+1 = wt - η𝚫wt
   2. bt+1 = bt + η𝚫bt
   3. Where 𝚫wt =
   4. Where 𝚫bt =