PadhAI: 6 Jars of Sigmoid Neuron

One Fourth Labs

Deriving the Gradient Descent Update rule

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

1. For ease of notation, let $\Delta\theta$ = u
2. Then from Taylor series,we have:
   a. $L(\theta + \eta u) = L(\theta) + \eta * u^T \nabla_\theta L(\theta)$
   b. Rearranging: $L(\theta + \eta u) - L(\theta) = \eta * u^T \nabla_\theta L(\theta)$
   c. Note,that the move ηu would only be favourable if
      i. $L(\theta + \eta u) - L(\theta) < 0$ (i.e. if the new loss is less than the previous loss)
      ii. This implies $u^T \nabla_\theta L(\theta) < 0$
   d. Now we have $u^T \nabla_\theta L(\theta) < 0$
      i. Let β be the angle between u and $\nabla_\theta L(\theta)$, then we know that,
      ii. $-1 \le cos(\beta) = \frac{u^T \nabla_\theta L(\theta)}{||u|| * ||\nabla_\theta L(\theta)||} \le 1$
      iii. Multiply throughout by k = $||u|| * ||\nabla_\theta L(\theta)||$
      iv. This gives us $-k \le u^T \nabla_\theta L(\theta) \le k$
   e. Thus, $L(\theta + \eta u) - L(\theta) = u^T \nabla_\theta L(\theta) = k * cos(\beta)$ will be most negative when $cos(\beta) = -1$, i.e. when β is 180°
3. Gradient Descent Rule
   a. The direction u that we intend to move in should be at 180° w.r.t, the gradient
   b. In other words, move in a direction opposite to the gradient
4. Parameter Update Rule
   a. $w_{t+1} = w_t - \eta \Delta w_t$
   b. $b_{t+1} = b_t + \eta \Delta b_t$
   c. Where $\Delta w_t = \frac{\partial L(w,b)}{\partial w}$ $at\ w = w_t,\ b = b_t$
   d. Where $\Delta b_t = \frac{\partial L(w,b)}{\partial b}$ $at\ w = w_t,\ b = b_t$