N. KAUSIK
CS21M037

CS6910 Deep Learning

Quiz 1

1. $x^2y^2 + y^2z^2 + z^2x^2 = 0$

diff wrt y, $\Rightarrow 2xy^2\frac{\partial x}{\partial y} + 2x^2y + 2zy^2\frac{\partial z}{\partial y} + 2zy^2 + 0 = 0$

$\Rightarrow 2xy\frac{\partial x}{\partial y} + 2x^2 + 2yz\frac{\partial z}{\partial y} + 2z^2 = 0$

$\Rightarrow x\left(x + \frac{\partial x}{\partial y}\right) = -z\left(z + \frac{\partial z}{\partial y}\right)$

4. we need $E(\hat{m}_t) = E(g)$,

For $m_t = \beta m_{t-1} + (1-\beta)\nabla w_t$,

bias correction done by $\hat{m}_t = \dfrac{m_t}{1-\beta^t}$

If β is taken as $\dfrac{\alpha_1}{\beta_1}$,

$$m_t = \frac{\alpha_1}{\beta_1} * m_{t-1} + \left(1 - \frac{\alpha_1}{\beta_1}\right) * \nabla w_t$$

$$= \frac{\alpha_1}{\beta_1} m_{t-1} + \left(\frac{\beta_1 - \alpha_1}{\beta_1}\right) \nabla w_t \rightarrow \text{required.}$$

$$\therefore \hat{m}_t = \frac{m_t}{1 - \frac{\alpha_1^t}{\beta_1^t}} = \frac{\beta_1^t m_t}{\beta_1^t - \alpha_1^t}$$ //

5. $f(x,y,z) = x^2+y^2+z^2-9$ $\quad (x_0,y_0,z_0) = (1,-2,1)$

G.D, $\quad \eta = 1$

$\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial z} = 2z$, $\frac{\partial f}{\partial y} = 2y$

$\therefore \; x_1 = x_0 - 1\times(2x_0) = x_0 - 2x_0 = -x_0 = -1$

$y_1 = y_0 - 1\times(2y_0) = -y_0 = 2$

$z_1 = z_0 - 1\times(2z_0) = -z_0 = -1$

Updated $\Rightarrow (-1, 2, -1)$

6. $\hat{L} = L + \frac{\lambda}{2} * \sum_{i=1}^{n} w_i^2$

$\nabla_{w_j} \hat{L} = \cancel{\frac{\partial \hat{L}}{\partial}} \quad \frac{\partial \hat{L}}{\partial L} \cdot \cancel{\nabla_{w_j} L}$

$= \nabla_{w_j} L + \frac{\lambda}{2} * \sum_{i=1}^{n} \frac{\partial (w_i^2)}{\partial w_j}$

$= \nabla_{w_j} L + \frac{\lambda}{2} \cdot 2 w_j = \nabla_{w_j} L + \lambda w_j$

update rule $\Rightarrow$ $w_j = w_j - \eta \left(\nabla_{w_j} L + \lambda w_j \right)$

$w_j = w_j - \eta \nabla_{w_j} L - \eta \lambda w_j$

7. For $t=1$,

$$\text{update}_1 = 0 + \eta \nabla w_1 = \eta \nabla w_1$$

For $t=2$,

$$\text{update}_2 = \gamma(\eta \nabla w_1) + \eta \nabla w_2$$

For $t=3$,

$$\text{update}_3 = \gamma^2(\eta \nabla w_1) + \gamma(\eta \nabla w_2) + \eta \nabla w_3$$

$\therefore$ For $t=10$,

$$\text{update}_{10} = \gamma^9(\eta \nabla w_1) + \dots$$

$\therefore$ Fraction $= \eta \gamma^9 = 1 \times (0.85)^9 = (0.85)^{10-1}$ $_{''}$

9. $\Rightarrow \tanh(x)$

~~max (tanh(x)) =~~

deriv of $\tanh(x) = 1 - (\tanh(x))^2$

~~min (tanh(x)) =~~

max of deriv is when $\tanh(x) = 0$

max $= 1 - 0^2 = 1$

for $x = 0$.

3. True $\Rightarrow$ $\left(\overset{\text{Red}}{\frac{40}{100}}, \overset{\text{Green}}{\frac{40}{100}}, \overset{\text{Blue}}{\frac{20}{100}} \right)$

Predicted $\Rightarrow$ $\left(\frac{50}{100}, \frac{25}{100}, \frac{25}{100} \right)$

Cross Entropy $= -\left(\frac{40}{100} \log_2 \frac{50}{100} + \frac{40}{100} \log_2 \frac{25}{100} + \frac{20}{100} \log_2 \frac{25}{100} \right)$

$= -\left(0.4 \log_2 \frac{1}{2} + 0.4 \log_2 \frac{1}{4} + 0.2 \log \frac{1}{4} \right)$

$= -\left(-0.4 - 0.8 - 0.4 \right) = 1.6$,,

9.

i/p $\quad$ n_1 $\quad$ n_2 $\quad$ o/p

4 $\quad$ 3 $\quad$ 4 $\quad$ 4

$4 \times 3 + 3 + 3 \times 4 + 4 + 4 \times 4 + 4$

$= 12 + 3 + 12 + 4 + 16 + 4 = 15 + 16 + 20 = 51$

10. For $k=1$, For $k=2$,

$$b_{j_1} = \frac{e^{a_j}}{\sum_{l=1}^{n} e^{a_l}} \qquad b_{j_2} = \frac{\left(e^{a_j}\right)^2}{\sum_{l=1}^{n} \left(e^{a_l}\right)^2}$$

Now, $$\left(b_{j_1}\right)^2 = \frac{\left(e^{a_j}\right)^2}{\left(\sum_{l=1}^{n} e^{a_l}\right)^2}$$

Since all as are +ve, all e^{a}s are +ve ~~and~~. and ≥ 1

$$\therefore \left(\sum_{l=1}^{n} e^{a_l}\right)^2 > \sum_{l=1}^{n} \left(e^{a_l}\right)^2$$

$$\therefore \quad b_{j_2} > \left(b_{j_1}\right)^2$$

Since we know b_{j_1} and $b_{j_2} \leq 1$, and ≥ 0.

$$0 \leq b_{j_1}, b_{j_2} \leq 1$$

and $b_{j_2} > \left(b_{j_1}\right)^2$