HOMEWORK - 4 GAURAV SHARMA

GXS230001

Q1 $\bar{x}_t = \beta \bar{x}_{t-1} + (1-\beta) x_t$

If $\beta = 0$ then:-

$\bar{A}_t = (1-0) x_t = x_t$

And:-

$\hat{x}_t = \frac{\bar{x}_t}{1-\beta^t} = \frac{x_t}{1-0^t} = \frac{x_t}{1} = x_t$

Therefore

$\hat{x}_t = x_t$ for all t when $\beta = 0$

Q2: $\bar{x}_t = (1-\beta) \sum_{i=1}^{t} \beta^{t-i} x_i$

$\hat{x}_t = \frac{\bar{x}_t}{1-\beta^t} = \frac{(1-\beta) \sum_{i=1}^{t} \beta^{t-i} x_i}{1-\beta^t}$

$\beta = 1$

~~we~~ we need to show

$$\lim_{\beta \to 1} \hat{x}_t = \frac{1}{t} \sum_{i=1}^{t} x_i$$

let

$$f(\beta) = (1-\beta) \sum_{i=1}^{t} \beta^{t-i} x_i$$

$$g(\beta) = 1 - \beta^t$$

$$\hat{x}_t = \frac{f(\beta)}{g(\beta)}$$

$$f'(\beta) = -\sum \beta^{t-i} x_i + (1-\beta) \sum_{i=1}^{t} (t-i) \beta^{t-i-1} x_i$$

Denominator :-

$$g'(\beta) = -t\beta^{t-1}$$

As $\beta \to 1$

$$\lim_{\beta \to 1} \hat{x}_t = \frac{1}{t} \sum_{i=1}^{t} x_i$$

Ques 2

Standard Back propogation update rule:-

$$\theta_{t+1} = \theta_t - \epsilon \nabla_\theta J(\theta_t)$$

Adam optimizer update Rule:-

Initialize :-

$$m_0 = 0 , v_0 = 0$$

Step 1:- Compute gradient

$$g_t = \nabla_\theta J(\theta_t)$$

Step 2:- First moment estimate

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$$

Step 3:- Second moment estimate

$$v_t = \beta_2 \cdot v_{t-1} + (1-\beta_2) \cdot g_t^2$$

Step 4:- Bias Estimation

$$\hat{m_t} = \frac{m_t}{1-\beta^t} , \hat{v_t} = \frac{v_t}{1-\beta_2^t}$$

Step 5:- Parameter update

$$\theta_{t+1} = \theta_t - \frac{\alpha \cdot \hat{m_t}}{\sqrt{v_t} + \epsilon}$$

Reducing adam to SGD

Let

$\beta_1 = 0 \Rightarrow m_t = g_t$

$\beta_2 = 1 \Rightarrow v_t = v_{t-1}$

$\alpha = \epsilon$

$\epsilon \to 0$

plug values in

$$\theta_{t+1} = \theta_t - \frac{\alpha \cdot g_t}{\sqrt{v_t} + \epsilon} \approx \theta_t - \alpha \cdot g_t$$

Since $\alpha = \epsilon$ we get

$$\theta_{t+1} = \theta_t - \epsilon \nabla_\theta J(\theta_t)$$

To make Adam $\approx$ SGD:

$\alpha = \epsilon$

$\beta_1 = 0$

$\beta_2 = 1$

$\epsilon = 10^{-8}$

$\rightarrow$ same as the standard back propagation

Ques 3:

1.) Running time of a single training iteration :-

A_1 :- Increase
A_2 :- Increase
B_1 :- Increase
B_2 :- Increase
C :- Increase
D :- Increase
E :- No - effect
F :- Increase

2.) Running time of the learned mode on a single testing example:-

A1: No-effect
A2 :- No effect
~~A3~~ B1 :- Increase
B2 :- No - effect
C :- Increase
D :- Increase
E :- No-effect
F :- No-effect

3) A1 :- Decrease
A2 :- Decrease
B1 :- Decrease
B2 :- Decrease
C :- Increase
D :- Increase
E :- No-effect
F :- Increase

4) A1 :- Decrease
A2 :- Decrease
B1 :- Decrease
B2 :- Decrease
C :- Increase
D :- Decrease
E :- Increase
F :- Increase

5) A1 :- Increase
A2 :- Increase
B1 :- Increase
B2 :- Increase
C :- Increase
D :- Increase
E :- No-effect
F :- No-effect