TF笔记-优化器

2022年7月12日 15:40

## 神经网络参数优化器

待优化参数$w$，损失函数$loss$，学习率$lr$，每次迭代一个batch，t表示当前batch迭代的总次数：

1. 计算t时刻损失函数关于当前参数的梯度$g_t = \nabla loss = \frac{\partial loss}{\partial(w_t)}$
2. 计算t时刻一阶动量$m_t$和二阶动量$V_t$
3. 计算t时刻下降梯度：$\eta_t = lr \cdot m_t / \sqrt{V_t}$
4. 计算t+1时刻参数：$w_{t+1} = w_t - \eta_t = w_t - lr \cdot m_t / \sqrt{V_t}$

一阶动量：与梯度相关的函数

二阶动量：与梯度平方相关的函数

神经网络参数优化器：

待优化参数 w

损失函数 loss

学习率 lr

每次参与迭代一个 batch

t表示当前batch迭代的总次数.

注：一阶动量：梯度相关函数

二阶动量：梯度的相关函数

1. 计算t时刻 loss 对 w 的梯度 $g_t = \nabla loss = \frac{\partial loss}{\partial(w_t)}$

2. 计算t时刻一阶动量 $m_t$，二阶动量 $V_t$

3. 计算t时刻下降梯度：$\eta_t = lr * \frac{m_t}{\sqrt{V_t}}$

4. 计算t+1时刻参数：$W_{t+1} = W_t - \eta_t = W_t - lr * \frac{m_t}{\sqrt{V_t}}$

优化器.

SGD（无动量） base

$m_t = g_t \quad V_t = 1 \quad \eta_t = lr * \frac{m_t}{\sqrt{V_t}} = lr * g_t$

$W_{t+1} = W_t - \eta_t = W_t - lr * g_t$

---

SGDM（SGD+一阶动量）

$m_t = \beta \cdot m_{t-1} + (1-\beta) \cdot g_t \quad V_t = 1 \quad \eta_t = lr * \frac{m_t}{\sqrt{V_t}} = lr * (\beta \cdot m_{t-1} + (1-\beta) \cdot g_t)$

$W_{t+1} = W_t - \eta_t = W_t - lr * (\beta \cdot m_{t-1} + (1-\beta) \cdot g_t)$

---

Adagrad（SGD+二阶动量）

$m_t = g_t \quad V_t = \sum_{\tau=1}^{t} g_\tau^2$ 开始到现在梯度$^2$的累积和. $\eta_t = lr * \frac{m_t}{\sqrt{V_t}} = lr * \frac{g_t}{\sqrt{\sum_{\tau=1}^{t} g_\tau^2}}$

$W_{t+1} = W_t - \eta_t = W_t - lr * \frac{g_t}{\sqrt{\sum_{\tau=1}^{t} g_\tau^2}}$

RMSProp（SGD+二阶动量）

$m_t = g_t \quad V_t = \beta \cdot V_{t-1} + (1-\beta) \cdot g_t^2 \quad \eta_t = lr * \frac{m_t}{\sqrt{V_t}} = lr * \frac{g_t}{\sqrt{\beta \cdot V_{t-1} + (1-\beta) \cdot g_t^2}}$

$W_{t+1} = W_t - \eta_t = W_t - lr * \frac{g_t}{\sqrt{\beta \cdot V_{t-1} + (1-\beta) g_t^2}}$

Adam（同时结合SGDM一阶动量和RMSProp二阶动量）

$$m_t = \beta_1 \cdot m_{t-1} + (1-\beta_1)\cdot g_t \quad \text{修正一阶动量的偏差：} \hat{m}_t = \frac{m_t}{1-\beta_1^t}$$

$$V_t = \beta_2 \cdot V_{t-1} + (1-\beta_2)\cdot g_t^2 \quad \text{修正二阶动量的偏差：} \hat{V}_t = \frac{V_t}{1-\beta_2^t}$$

$$\eta_t = l_r * \frac{\hat{m}_t}{\sqrt{\hat{V}_t}} = l_r * \frac{m_t}{1-\beta_1^t} \Big/ \sqrt{\frac{V_t}{1-\beta_2^t}}$$

$$= l_r * \frac{\beta_1 \cdot m_{t-1} + (1-\beta_1)\cdot g_t}{1-\beta_1^t} \Big/ \sqrt{\frac{\beta_2 \cdot V_{t-1} + (1-\beta_2)\cdot g_t^2}{1-\beta_2^t}}$$

$$W_{t+1} = W_t - \eta_t = W_t - l_r * \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{V_t}{1-\beta_2^t}}}$$