CSC421: Written Homework 2

Due on Monday, Feb 11, 2019

Zhongtian Ouyang
1002341012

Problem 1

Adam
(a)
$(\alpha_A, \beta_1, \beta_2, \epsilon_A) = (\alpha_R, 0, \gamma, \epsilon_R)$
Since Adam is a combination of RMSprop and Momentum. If we don't do the momentum part, it will be identical to RMSprop.
By setting the hyperparameters of Adam as above,
$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t = 0 m_{t-1} + (1 - 0) g_t = g_t$,
$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 = \gamma v_{t-1} + (1 - \gamma) g_t^2$
$\theta_t = \theta_{t-1} - \alpha_R g_t / (\sqrt{(v_t)} + \epsilon_R) = \theta_{t-1} - \alpha_A m_t / (\sqrt{v_t} + \epsilon_A) = \theta_{t-1} - \alpha_R g_t / (\sqrt{v_t} + \epsilon_R)$
which is identical to RMSprop

(b)
$(\alpha_A, \beta_1, \beta_2, \epsilon_A) = (\alpha_S, \mu, 1, 1)$
We can make Adam and SGD with momentum equivalent by satisfying the following conditions: $m_t = -p_t$, $(\sqrt{v_t} + \epsilon_R) = 1$ and $\alpha_A = \alpha_S$

If $v_t = 0$, $\epsilon_A = 1$, $\sqrt{v_t} + \epsilon_A = 1$. We can achieve this by setting β_2 to 1. Since $v_0 = 0$, $v_1, v_2 ... v_t = v_0 = 0$

We can prove by induction that $m_t = -p_t$ when $\beta_1 = \mu$.
$m_0 = 0 = -p_0$,
$m_1 = \beta_1 m_0 + (1 - \beta_1) g_1 = (1 - \beta_1) g_1 = (1 - \mu) \nabla J(\theta_0)$, $p_1 = \mu p_0 - (1 - \mu) \nabla J(\theta_0) = -(1 - \mu) \nabla J(\theta_0)$. $m_1 = -p_1$
Assume $m_{t-1} = -p_{t-1}$, $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t = -\mu p_{t-1} + (1 - \mu) \nabla J(\theta_{t-1}) = -(\mu p_{t-1} - (1 - \mu) \nabla J(\theta_{t-1}))$, $p_t = \mu p_{t-1} - (1 - \mu) \nabla J(\theta_{t-1})$. $m_t = -p_t$

By setting the hyperparameters of Adam as above, $\theta_t = \theta_{t-1} - \alpha_A m_t / (\sqrt{v_t} + \epsilon_A) = \theta_{t-1} + \alpha_S p_t / 1 = \theta_{t-1} + \alpha_S p_t$, which is identical to SGD with momentum.

(c)
$\tilde{L}(y,t) = CL(y,t)$, $\nabla\tilde{J}(\theta_0) = C\nabla J(\theta_0)$, $\tilde{g}_t = Cg_t$

Prove $\tilde{m}_t = Cm_t$ for $t \geq 1$:
Base Case:
$\tilde{m}_1 = \beta_1\tilde{m}_0 + (1-\beta_1)\tilde{g}_1 = C(1-\beta_1)g_1$, $m_1 = \beta_1 m_0 + (1-\beta_1)g_1 = (1-\beta_1)g_1$.
$\tilde{m}_1 = Cm_1$.
Inductive Steps: Assume $\tilde{m}_{t-1} = Cm_{t-1}$
$\tilde{m}_t = \beta_1\tilde{m}_{t-1} + (1-\beta_1)\tilde{g}_t = C\beta_1 m_{t-1} + C(1-\beta_1)g_t = C(\beta_1 m_{t-1} + (1-\beta_1)g_t)$,
$m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t$.
$\tilde{m}_t = Cm_t$.

Prove $\tilde{v}_t = C^2 v_t$ for $t \geq 1$:
Base Case:
$\tilde{v}_1 = \beta_2\tilde{v}_0 + (1-\beta_2)\tilde{g}_1^2 = C^2(1-\beta_2)g_1^2$, $v_1 = \beta_2 v_0 + (1-\beta_2)g_1^2 = (1-\beta_2)g_1^2$.
$\tilde{v}_1 = C^2 v_1$.
Inductive Steps: Assume $\tilde{v}_{t-1} = C^2 v_{t-1}$
$\tilde{v}_t = \beta_2\tilde{v}_{t-1} + (1-\beta_2)\tilde{g}_t^2 = C^2\beta_2 v_{t-1} + C^2(1-\beta_2)g_t^2 = C^2(\beta_2 v_{t-1} + (1-\beta_2)g_t^2)$,
$v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2$.
$\tilde{v}_t = C^2 v_t$.

Prove $\tilde{\theta}_t = \theta_t$ for $t \geq 0$: Base Case: $\tilde{\theta}_0 = \theta_0$ because they initialize to the same value
Inductive Steps: Assume $\tilde{\theta}_{t-1} = \theta_{t-1}$
$\tilde{\theta}_t = \tilde{\theta}_{t-1} - \alpha_A\tilde{m}_t/(\sqrt{\tilde{v}_t} + \epsilon_A) = \theta_{t-1} - \alpha_A Cm_t/(\sqrt{C^2 v_t} + 0) = \theta_{t-1} - \alpha_A Cm_t/C\sqrt{v_t} = \theta_{t-1} - \alpha_A m_t/\sqrt{v_t}$,
$\theta_t = \theta_{t-1} - \alpha_A m_t/(\sqrt{v_t} + \epsilon_A) = \theta_{t-1} - \alpha_A m_t/(\sqrt{v_t} + 0) = \theta_{t-1} - \alpha_A m_t/\sqrt{v_t}$
$\tilde{\theta}_t = \theta_t$