Gradient methods for parameter optimization

Exercise sheet 4

2 Seriously Cool Guys

Exercise H4.1

(a)

$$T_2(x|_{x_0}) \approx f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)}{2} f''(0)$$

$$T_2(E^T(\omega_{t+1})|_{\omega_t}) \approx E^T(\omega_t) + \frac{(\omega_{t+1} - \omega_t)^2}{2} \frac{\partial^2 E^T(\omega_t)}{\partial^2 w_t} + (\omega_{t+1} - \omega_t) \frac{\partial E}{\partial w}$$

assuming:

$$\omega_{t+1} = \omega_t - \eta_t d_t \quad \text{and } d_t = g_t \Rightarrow \omega_{t+1} - \omega_t = -\eta_t g_t$$

$$\text{and also: } \frac{\partial^2 E^T}{\partial^2 \omega_t} = H(t) \quad \text{and} \quad \frac{\partial E^T}{\partial \omega_t} = g_t$$

$$\Rightarrow T_2(E^T(\omega_{t+1})|_{\omega_t}) \approx E^T(\omega_t) + \Delta\omega^T g_t + \frac{1}{2} \Delta\omega^T H_t \Delta\omega$$

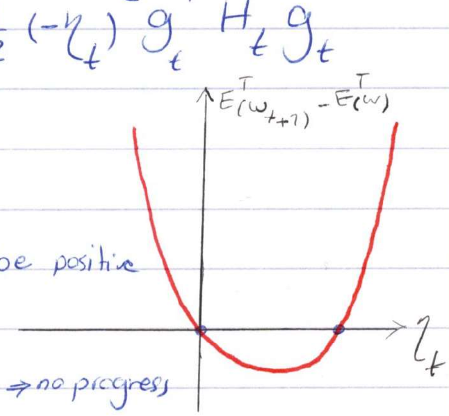
(b) as from (a): $\omega_{t+1} - \omega_t = -\eta_t g_t$

$$E^T(\omega_{t+1}) \overset{!}{\leq} E^T(\omega_t) \Rightarrow E^T(\omega_{t+1}) - E^T(\omega_t) \overset{!}{\leq} 0$$

using the 2nd order taylor series:

$$E^T(\omega_{t+1}) - E^T(\omega_t) \approx g_t^T(-\eta_t) g_t + \frac{1}{2}(-\eta_t)^2 g_t^T H_t g_t$$

$$\Rightarrow \eta_t^2 g_t^T H_t g_t - 2\eta_t g_t^T g_t \leq 0$$



$\Rightarrow$ root 1 $\leq \eta_t \leq$ root 2 (assuming H to be positive difinit)

$$0 < \eta_t \leq \frac{2 g_t^T g_t}{g_t^T H g_t}$$

* if $\eta = 0 \Rightarrow$ no progress

C) using what we got from 1. b)

$$E^T(w_{t+1}) - E^T(w_t) \approx \frac{1}{2}\eta_t^2 g_t^T H_t g_t - \eta_t g_t^T g_t$$

where $g_t = \nabla E^T_{w_t}$

our goal is to minimize $E^T(w_{t+1}) - E^T(w_t)$ with respect to η_t

$$\Rightarrow \frac{\partial}{\partial \eta_t}\left(E^T(w_{t+1}) - E^T(w_t)\right) = 0 \quad \text{where } w_{t+1} = w^*$$

$$\frac{\partial}{\partial t_t}\left(E^T(w^*) - E(w)\right) = 2 \cdot \frac{1}{2}\eta_t g_t^T H_t g_t - g_t^T g_t = 0$$

$$\Rightarrow \eta_t = g_t^T g_t / (g^T H_t g_t)$$

(minimizing the cost function)

replacing g_t with $\nabla E^T(w)$

$$\Rightarrow \eta_{t\,(optimal)} = \frac{\underline{\nabla} E^{T^T}(w)\, \underline{\nabla} E^T(w)}{\underline{\nabla} E^{T^T}(w)\, \underline{H}\, \underline{\nabla} E^T(w)}$$

retry

if $E^T(w) = \frac{1}{2}(\underline{w} - \underline{w}^*)^T H (\underline{w} - \underline{w}^*)$ $\quad w^* = \underline{w} - \eta_t d_t$

$$\frac{\partial E^T(w)}{\partial \eta_t} = \frac{\partial E^T(w)}{\partial(w)} \cdot \frac{\partial(w)}{\partial \eta_t} = \underbrace{(H\underline{w} + \underline{b}) \times (d_t)}_{\text{used in 1.d to show } \perp} = \frac{1}{2}(\eta_t^2 d_t^T H d_t)'$$

$$\Rightarrow \frac{1}{2} 2\eta_t d_t^T H d_t + \frac{1}{2}\eta_t^2 (d_t H d_t)' = 0$$

$$\Rightarrow \eta_t^2 (d_t H d_t)' + 2\eta_t d_t^T H d_t = 0 \qquad \eta_t \neq 0$$

$$\Rightarrow \eta_t = 2\frac{(d_t^T H d_t)}{(d_t^T H d_t)'} \;?$$

Exercise H4.1

d)

if $\underline{d}_t$ is optimized, it must be due to an optimal $\eta_t = \eta_t^*$

Assuming this, and that $g_t = E^T(w_t - \eta_t d_t)$

with $g(t) = (d_t^T H d_t)\eta_t + d_t^T H w_t + b d^T$ b is scaler coeff.

then the minimum of g_t occurs when $g_t' = 0$

when $\eta_t^* = \dfrac{d_t^T (H w_t + b)}{d_t^T H d_t} = -\dfrac{d_t^T \nabla E^T(w)}{d_t^T H d_t}$

considering $p := w_t - \eta_t^* d_t$ where p is minimum point on line $w_t - \eta_t d_t$ we see:

$$\nabla E(p) \cdot d_t = 0$$

Knowing that $\nabla E(p) \neq 0$ and $d_t \neq 0$ $\Rightarrow \nabla E(p) \perp d_t$