Extra afleidingen papers

October 26, 2020

Niet alle stappen zijn in detail beschreven, maar voldoende om te begrijpen hoe men aan het resultaat komt.

1 Train faster, generalize better:

Pagina 3: Lemma 2.4

$$\delta_{t+1} \leq \begin{cases} \eta\delta_t & \text{indien } G_t = G'_t \text{ en beide } \eta \text{ expansief} \\ \min(\eta, 1)\delta_t + 2\sigma_t & \text{indien } G_t \text{ en } G'_t\ \sigma\text{-bounded zijn en } G_t \text{ is } \eta \text{ expansief} \end{cases}$$

Bewijs/afleiding:

$$\begin{aligned} \delta_{t+1} &= \|w'_{t+1} - w_{t+1}\| \\ &= \|G'_t(w'_t) - G_t(w_t)\| \qquad (1) \\ &= \|(w_t - G_t(w_t)) - (w'_t - G'_t(w'_t)) + (w'_t - w_t)\| \qquad (2) \end{aligned}$$

$$\begin{aligned} \delta_{t+1} &= \|w'_{t+1} - w_{t+1}\| \\ &= \|G'_t(w'_t) - G_t(w_t)\| \\ &= \|(G'_t(w'_t) - G_t(w_t)) + (G_t(w'_t) - G_t(w_t) + (G_t(w_t) - G_t(w'_t))\| \\ &= \|(G'_t(w'_t) - G_t(w'_t)) + (G_t(w'_t) - G_t(w_t))\| \qquad (3) \end{aligned}$$

Indien $G_t = G'_t$ en beide η expansief:

Gebruikmakend van (1):

$$\begin{aligned} G_t = G'_t \Rightarrow \|G'_t(w'_t) - G_t(w_t)\| &= \|G_t(w'_t) - G_t(w_t)\| \\ &\leq \eta\|w'_t - w_t\| \\ &= \eta\delta_t \end{aligned}$$

Indien G_t en G'_t σ-bounded zijn en G_t is η expansief:

Gebruikmakend van (2):

$$\begin{aligned} \|(w_t - G_t(w_t)) - (w'_t - G'_t(w'_t)) + (w'_t - w_t)\| &\leq \|w_t - G_t(w_t)\| + \|w'_t - G'_t(w'_t)\| + \|w'_t - w_t\| \\ &\leq 2\sigma_t + \delta_t \end{aligned} \tag{4}$$

Gebruikmakend van (3):

$$\begin{aligned} &\|(G'_t(w'_t) - G_t(w'_t)) + (G_t(w'_t) - G_t(w_t))\| \\ &\leq \|G'_t(w'_t) - G_t(w'_t)\| + \|G_t(w'_t) - G_t(w_t)\| \\ &\leq \|G'_t(w'_t) - G_t(w'_t) + w'_t - w'_t\| + \|G_t(w'_t) - G_t(w_t)\| \\ &\leq \|w'_t - G'_t(w'_t)\| + \|w'_t - G_t(w'_t)\| + \|G_t(w'_t) - G_t(w_t)\| \\ &\leq 2\sigma_t + \eta\delta_t \end{aligned} \tag{5}$$

Indien we (4) en (5) vergelijken zien we dat de uiteindelijke laagste bovengrens zal gegeven worden door het feit dat η kleiner is dan 1 of niet. Dit leidt tot wat we zoeken:

$$\delta_{t+1} \leq 2\sigma_t + min(\eta, 1)\delta_t$$

□

Pagina 4: Eerste alinea sectie 3.1 Proof idea

Indien de verlies-functie f, L-lipschitz is geldt:

$$\mathbb{E}[|f(w;z) - f(w';z)|] \leq L\,\mathbb{E}[\|w - w'\|]$$

Bewijs/afleiding:

$$\mathbb{E}[|f(w;z) - f(w';z)|] \leq \mathbb{E}[L\|w - w'\|] = L\,\mathbb{E}[\|w - w'\|]$$

□

Pagina 4: Lemma 3.3

Indien de verlies-functie f, L-lipschitz is dan is de gradient-update $G_{f,\alpha}$, (αL)-bounded.

Bewijs/afleiding:

$$\|w - G_{f,\alpha}(w)\| = \|\alpha\nabla f(w)\| = \alpha\|\nabla f(w)\| \leq \alpha L$$

Er geldt dus $\|w - G_{f,\alpha}(w)\| \leq \alpha L \Rightarrow G_{f,\alpha}$ is (αL)-bounded.

$\square$

Pagina 4: Lemma 3.6

Indien de verlies-functie f, β-smooth is dan is de gradient-update $G_{f,\alpha}$, $(1 + \alpha\beta)$-expansief.

Bewijs/afleiding:

$$\begin{aligned}
\|G_{f,\alpha}(v) - G_{f,\alpha}(w)\| = \|v - w + \alpha(\nabla f(w) - \nabla f(v))\| &\leq \|v - w\| + \alpha\|\nabla f(v) - \nabla f(w)\| \\
&\leq \|v - w\| + \alpha\beta\|v - w\| \\
&= (1 + \alpha\beta)\|v - w\|
\end{aligned}$$

Er geldt dus $\|G_{f,\alpha}(v) - G_{f,\alpha}(w)\| \leq (1 + \alpha\beta)\|v - w\| \Rightarrow G_{f,\alpha}$ is $(1 + \alpha\beta)$-expansief.

$\square$