Vedlegg B

I. ADAM

Adam (**?**) er en optimeringsalgoritme som er en tilpasning av SGD hvor læringsraten η varierer gjennom iterasjonene. I Adam inkluderes gradienten fra forrige iterasjon i første og andre potens i tillegg, på en måte som kan betraktes som en tilpasning av læringsraten.

Adam har flere metaparametere, η, ρ_1, ρ_2, og vi ser at denne metoden benytter både gradienten i første potens (m), med minne fra forrige iterasjon, og gradienten i andre potens (s) for å justere steget i $\boldsymbol{w}$:

$$\boldsymbol{m}^{(i)} = \rho_1 \boldsymbol{m}^{(i-1)} + (1-\rho_1)\boldsymbol{g}^{(i)}$$

$$\boldsymbol{s}^{(i)} = \rho_2 \boldsymbol{s}^{(i-1)} + (1-\rho_2)(\boldsymbol{g}^{(i)})^2$$

$$\boldsymbol{m}^{(i)} = \frac{\boldsymbol{m}^{(i)}}{1-\rho_1^i}$$

$$\boldsymbol{s}^{(i)} = \frac{\boldsymbol{s}^{(i)}}{1-\rho_2^i}$$

$$\boldsymbol{w}^{(i+1)} = \boldsymbol{w}^{(i)} - \eta^{(i)} \frac{\boldsymbol{m}^{(i)}}{\sqrt{\boldsymbol{s}^{(i)} + \delta}}.$$

δ er her et lite tall som legges til for å unngå problemer med å dele på 0. Og igjen, vil det være et helt tilsvarende sett med ligninger for $\boldsymbol{b}$, og det er vanlig å benytte de samme metaparameterne for å optimere både $\boldsymbol{w}$ og $\boldsymbol{b}$.