

ImageRecognition aiheääritys (tarkentuu)

Simo Korkolainen

30. toukokuuta 2016

Projektin tarkoituksena on tehdä ohjelma, joka opettaa neuroverkon tunnistamaan kuvia backpropagation-algoritmin avulla. Neuroverkon opetuksessa verkon painoja muutetaan liikuttamalla niitä virhefunktion gradientin vastaiseen suuntaan, kunnes virhefunktio on minimoitunut ja neuroverkko on oppinut tunnistamaan kuvat. Derivoinnin ketjusääntöön perustuva backpropagation-algoritmi mahdollistaa gradientin nopean laskemisen.

1 Neuroverkko

Olkoon L neuroverkon kerroksien lukumäärä. Olkoon l_k kerroksen $k = 1, \dots, L$ neuronien lukumäärä. Merkitään kerroksen k aktivaatiota vektorina $z_k \in \mathbb{R}^{l_k}$. Ensimmäisen kerroksen aktivaatio z_1 on neuroverkon syöte ja viimeisen kerroksen aktivaation z_L on neuroverkon antama tuloste.

Jokaisen kerroksen $k > 1$ aktivaation voidaan ajatella laskettavan parametrisoidun funktion $f_k : \mathbb{R}^{l_{k-1}} \times A_k \rightarrow \mathbb{R}^{l_k}$ avulla. Tässä A_k on verkon kerroksien $k-1$ ja k yhteyksien painoina toimivien parametrien joukko. Kerroksen aktivaatio lasketaan kaavan

$$z_k = f(z_{k-1}, a_k) \tag{1}$$

avulla, missä $a_k \in A_k$.

2 Backpropagation-algoritmi

3 Aikavaativuus

Ohjelmassa käytetään vain eteenpäin kytkettyjä neuroverkkoja. Neuroverkko koostuu kerroksista, joissa on neuroneita. Neuroneiden aktivaatio z_k kerroksessa k lasketaan täsmälleen edellisen kerroksen aktivaatioiden perusteella eli $z_k = f(z_{k-1}, a_k)$ missä f on aktivaatiofunktio. Kuten aikasemmin, olkoon L neuroverkon kerroksien lukumäärä ja olkoon l_k kerroksen $k = 1, \dots, L$ neuronien lukumäärä. Jos jokainen kerroksen k neuroni on kytketty kaikkiin edellisen kerroksen solmuihin ja neuronipariin liittyvän laskennan aikavaativuus

on luokkaa $O(1)$, yhden kerroksen k neuronin aktivaation laskemisen aikavaativuus on luokkaa $O(l_{k-1})$. Koska kerroksessa k on l_k neuronia, koko kerrokseen liittyvän laskennan aikavaativuus on $O(l_{k-1}l_k)$. Ensimmäisen kerroksen eli syötekerroksen aktivaatioiden asettamisen aikavaativuus on $O(l_1)$.

Koko neuroverkon aktivaatioiden laskennan aikavaativuus T_{act} on kerrosten aikavaativuuksien summa eli

$$T_{act} = O(l_1 + \sum_{k=2}^L l_{k-1}l_k)$$

Tarkastellaan tapausta, jossa kerrosten neuronien lukumäärä pienenee eksponentiaalisesti eli $l_k = \alpha^{k-1}l_1$, missä $0 < \alpha < 1$. Tällöin

$$\begin{aligned} l_1 + \sum_{k=2}^L l_{k-1}l_k &= l_1 + \sum_{k=2}^L \alpha^{k-2}l_1\alpha^{k-1}l_1 \\ &= l_1 + l_1^2 \sum_{k=2}^L \alpha^{2k-3} \\ &= l_1 + l_1^2 \alpha \sum_{k=0}^{L-2} (\alpha^2)^k \\ &\leq l_1 + l_1^2 \alpha \sum_{k=0}^{\infty} (\alpha^2)^k \\ &= l_1 + l_1^2 \frac{\alpha}{1-\alpha^2} \end{aligned}$$

Saamme, että $T_{act} = O(l_1^2)$, koska $\frac{\alpha}{1-\alpha^2}$ on positiivinen vakio.