IA353A - Neural Networks
EC2

Rafael Claro Ito
(R.A.: 118430)

August 2020

Question 7

7.1 Transferência negativa

7.2 Camadas compartilhadas

7.3 MALSAR

Question 8

Considerando a camada q de uma rede neural, temos:

$$x^{[q]} = f(W^{[q]}x^{[q-1]} + b)$$

Onde:

- $x^{[q]}$ é saída da camada q após função de ativação;
- $W^{[q]}$ são os pesos dos neurônios da camada q;
- $x^{[q-1]}$ é a entrada camada q (saída da camada $q-1$);
- $f(\cdot)$ é a função de ativação da camada q;

Calculando a variância de ambos lados da equação anterior, temos:

$$x^{[q]} = f(W^{[q]}x^{[q-1]} + b)$$
$$Var(x^{[q]}) = Var(f(W^{[q]}x^{[q-1]} + b))$$

Devemos agora fazer algumas considerações:

(i) função de ativação $f(\cdot) = tanh$, sendo que no início do treinamento trabalha-se com os pesos próximos à região linear (próximo de zero), evitando neurônios operando na região de saturação e favorecendo o aprendizado nas primeiras iterações. Desta forma, considerando a região linear, podemos aproximar a função $tanh$ para uma função identidade;

(ii) W e x são independentes entre si;

(iii) W é uma variável aleatória i.i.d. (independente e identicamente distribuída). Isso geralmente é verdade para a inicialização.;

(iv) x é uma variável aleatória i.i.d. (independente e identicamente distribuída). Embora nem sempre isso seja verdade, por exemplo, pixels de uma imagem geralmente apresentam alta correlação entre pixels ao redor, faremos essa consideração;

Continuando o desenvolvimento da equação anterior, temos:

$$Var(x^{[q]}) = Var(f(W^{[q]}x^{[q-1]} + b))$$

A partir de i), temos:

$$Var(x^{[q]}) = Var(W^{[q]}x^{[q-1]} + b)$$

Se duas variáveis são independentes entre si, temos a igualdade: https://en.wikipedia.org/wiki/Variance#Product_of_independent_variables:

$$Var(XY) = [\mathbb{E}(X)]^2 Var(Y) + [\mathbb{E}(Y)]^2 Var(X) + Var(X)Var(Y)$$

Vamos agora abrir o produto das matrizes W e x em uma soma dos produtos de seus termos w_i e x_i. Usando também a consideração dada por ii) e sabendo que b é uma constante (e portanto sua variância é zero), temos que:

$$Var(x^{[q]}) = Var(W^{[q]}x^{[q-1]} + b)$$
$$Var(x^{[q]}) = Var(\sum_{i=1}^{n^{[q-1]}} w_i^{[q]}x_i^{[q-1]})$$
$$Var(x^{[q]}) = \sum_{i=1}^{n^{[q-1]}} Var(w_i^{[q]}x_i^{[q-1]})$$
$$Var(x^{[q]}) = \sum_{i=1}^{n^{[q-1]}} [\mathbb{E}(w_i^{[q]})]^2 Var(x_i^{[q-1]}) + [\mathbb{E}(x_i^{[q-1]})]^2 Var(w_i^{[q]}) + Var(w_i^{[q]})Var(x_i^{[q-1])}$$

A partir de iii) e iv), temos que $\mathbb{E}(w_i) = 0$ e $\mathbb{E}(x_i) = 0$. Sendo ambos W e x variáveis i.i.d. Assim:

$$Var(x^{[q]}) = \sum_{i=1}^{n^{[q-1]}} Var(w_i^{[q]}) Var(x_i^{[q-1]})$$

$$\boxed{Var(x^{[q]}) = n^{[q-1]} Var(w^{[q]}) Var(x^{[q-1]})}$$

Queremos provar que $b = \sqrt{\frac{3}{n^{[q-1]}}}$ para que a variância da entrada da camada q seja igual a variância da camada $q-1$.

Sabendo que a variância de uma variável aleatória que segue uma distribuição uniforme entre a e b, isto é, $X \sim \mathbb{U}[a,b]$, é dada por: $Var(X) = \frac{(b-a)^2}{12}$ (prova). Temos:

$$Var(x^{[q]}) = n^{[q-1]} Var(w^{[q]}) Var(x^{[q-1]})$$

$$Var(x^{[q]}) = n^{[q-1]} \cdot \frac{(b-(-b))^2}{12} \cdot Var(x^{[q-1]})$$

$$Var(x^{[q]}) = n^{[q-1]} \cdot \frac{(2b)^2}{12} \cdot Var(x^{[q-1]})$$

$$Var(x^{[q]}) = n^{[q-1]} \cdot \frac{4b^2}{12} \cdot Var(x^{[q-1]})$$

Como queremos $Var(x^{[q]}) = Var(x^{[q-1]})$, temos:

$$n^{[q-1]} \cdot \frac{4b^2}{12} = 1$$

$$b^2 = \frac{12}{4 \cdot n^{[q-1]}}$$

$$\boxed{b = \sqrt{\frac{3}{n^{[q-1]}}}}$$

Question 9

9.1 Principais seções do padrão de documentação

9.2 Artigos com propósitos similares

Artigo 1

Artigo 2

Question 10

10.1 EfficientNet

10.2 FixEfficientNet