Forums de discussion

Obtenez de l'aide et discutez du contenu du cours avec la communauté.

FORUM DE DISCUSSION DE CETTE SEMAINE

Week 7

Discuss and ask questions about Week 7.

2443 sujets de discussion · Dernière publication an hour ago

Aller au forum de discussion

Forums | **Tous les sujets de discussion**

Search

All Course Discussions

SB

Books for more detailed math derivation (Back propagation)

Sergey Borovikov Week 5 · 9 days ago

This question was asked about two years ago ans answered with only online references to actual derivation. Any up to date changes? I prefer old school paper books for reading serious math.

0 J'aime Reply Suivre cette discussion

Au plus tôt **Sommet** **Plus récent**


min

Neil Ostrove Mentor · 9 days ago · Modifié

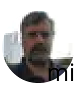
I'm not really aware of anything printed that follows the conventions Prof Ng uses in this course. If you come across something, please post.

There are three main formulas we use here. I'll sketch derivations below. Let me know if you have any questions.

Backpropagation is essentially repeated applications of the chain rule in calculus to compute the derivatives (gradients) used for gradient descent. The $\delta_i^{(l)}$ are actually $\frac{\partial}{\partial z_i^{(l)}} J$. The $\Delta_{j,i}^{(l)}$, after being scaled by m and regularized to get $D_{j,i}^{(l)}$, are actually the $\frac{\partial}{\partial \Theta_{j,i}^{(l)}} J$ that are used to update the $\Theta_{j,i}^{(l)}$ in gradient descent.

The recurrence relations allow you to move backwards from $l = L$, where the derivatives can be calculated directly from the cost function J, to the hidden layers and then compute the gradients.

0 J'aime Masquer 5 répond


min

Neil Ostrove Mentor · 9 days ago

Here's a backpropagation $\delta^{(L)}$ derivation for J, ignoring superscripts ($l = L$) and subscripts (and bias) for clarity.

$$
\begin{aligned}
\sigma(z) &= \tfrac{1}{1+e^{-z}} = (1+e^{-z})^{-1} \implies \\
\tfrac{\partial \sigma}{\partial z} &= \sigma'(z) = (-1)(1+e^{-z})^{-2}\, \tfrac{\partial}{\partial z}\,(1+e^{-z}) \\
&= (-1)(1+e^{-z})^{-2}(e^{-z})(-1) \\
&= \tfrac{1}{1+e^{-z}}\,(\tfrac{1+e^{-z}-1}{1+e^{-z}}) \\
&= \sigma(z)(1-\sigma(z)) \\
a &= \sigma(z) \implies \tfrac{\partial a}{\partial z} = \sigma'(z) = a(1-a)
\end{aligned}
$$

$$
\begin{aligned}
J(\Theta) &= -[y\log(a) + (1-y)\log(1-a)] \implies \\
\tfrac{\partial J}{\partial a} &= -[y(\tfrac{1}{a}) + (1-y)\,\tfrac{(-1)}{1-a}] \\
&= -[(\tfrac{y}{a}) - (\tfrac{1-y}{1-a})] \\
&= -\,\tfrac{1}{a(1-a)}\,[(1-a)y - a(1-y)] \\
&= -\,\tfrac{1}{a(1-a)}\,[y - ay - a + ay] \\
&= \tfrac{1}{a(1-a)}\,[a-y]
\end{aligned}
$$

$$
\begin{aligned}
\delta^{(L)} &= \tfrac{\partial}{\partial z} J = \tfrac{\partial J}{\partial a}\,\tfrac{\partial a}{\partial z} = \tfrac{1}{a(1-a)}\,[a-y]\,a(1-a) \\
&= a - y = a^{(L)} - y
\end{aligned}
$$

3 J'aime

Neil Ostrove Mentor · 9 days ago

Here's a derivation of the $\delta^{(l)}$ recurrence, which is independent of the cost function and depends only on the structure of the neural network:

$$\delta^{(l-1)} = \frac{\partial}{\partial z^{(l-1)}} J = \frac{\partial J}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial z^{(l-1)}} = \delta^{(l)} \frac{\partial}{\partial z^{(l-1)}} z^{(l)}$$
$$= \delta^{(l)} \frac{\partial}{\partial z^{(l-1)}} (\Theta^{(l-1)} g(z^{(l-1)})) = \delta^{(l)} \Theta^{(l-1)} g'(z^{(l-1)})$$

3 J'aime

Neil Ostrove Mentor · 9 days ago

The chain rule gives the derivative with respect to $\Theta^{(l)}$:

$$\frac{\partial}{\partial \Theta^{(l)}} J = \frac{\partial z^{(l+1)}}{\partial \Theta^{(l)}} \frac{\partial J}{\partial z^{(l+1)}} = \frac{\partial(\Theta^{(l)} a^{(l)})}{\partial \Theta^{(l)}} \delta^{(l+1)} = a^{(l)} \delta^{(l+1)}$$

$a^{(l)}$ only depends on $\Theta^{(l-1)}$ and below so, as far as differentiation by $\Theta^{(l)}$ is concerned, it's multiplication by a constant.

2 J'aime

SB Sergey Borovikov · 7 days ago

Some more questions.

1) Why delta(l-1) is a derivative of cost function. What is the formal definition if delta(l-1(

2) Your formula is missing transposition of matrix theta. Why?

1 J'aime

Neil Ostrove Mentor · 7 days ago

1. $\delta^{(l)}$ is defined as the derivative of J with respect to $z^{(l)}$.
2. The derivation is to show the calculus part. The linear algebra part (transpose, order, etc.) depends on particular conventions like column vector vs. matrix and bias considerations. See the week 2 Tips from Mentors.

0 J'aime