20220507-书&机器学习

1.过程描述

1.1故事

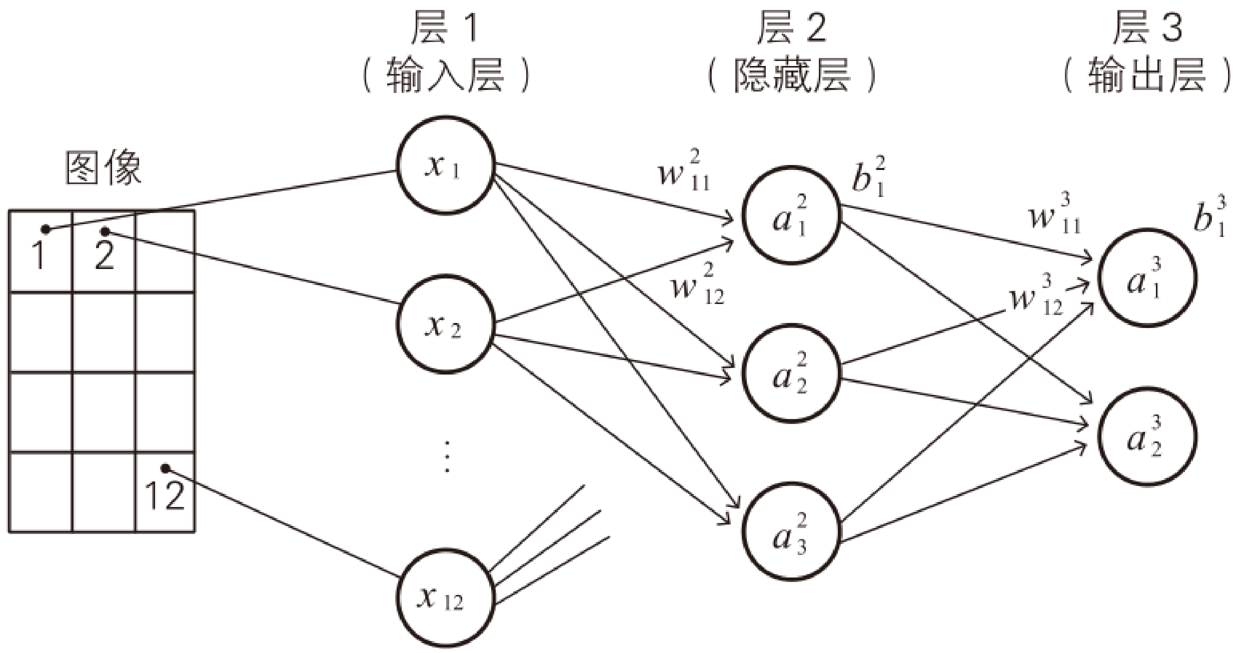
> 故事是生活的比喻。一个讲故事的人即是一个生活诗人，一个艺术家，将日常生活事件、内在生活和外在生活、梦想和现实转化为一首诗，一首以事件而不是以语言作为韵律的诗——一个长达两个小时的比喻，告诉观众：生活就像是这样！因此，故事必须抽象于生活，提取其精华，但又不能成为生活的抽象化，以致失却了实际生活的原味。故事必须像生活，但又不能一成不变地照搬生活，以致除了市井乡民都能一目了然的生活之外便别无深度和意味。

1.2机器学习

简单神经网络的误差反向传播过程

对于这样的一个网络：

要更新网络中的权值与bias，如果暴力求导，将会使得求解过程变得十分复杂。而借助导数的链式法则，可以很轻松地求出误差对于各个权重及bias的偏导，从而实现权值及bias的更新，进而逼近最优。



< 隐藏层 >

$$z_1^2 = w_{11}^2 x_1 + w_{12}^2 x_2 + \cdots + w_{1\,12}^2 x_{12} + b_1^2$$

$$z_2^2 = w_{21}^2 x_1 + w_{22}^2 x_2 + \cdots + w_{2\,12}^2 x_{12} + b_2^2$$

$$z_3^2 = w_{31}^2 x_1 + w_{32}^2 x_2 + \cdots + w_{3\,12}^2 x_{12} + b_3^2$$

$$a_i^2 = a(z_i^2) \quad (i = 1,\ 2,\ 3)$$

< 输出层 >

$$z_1^3 = w_{11}^3 a_1^2 + w_{12}^3 a_2^2 + w_{13}^3 a_3^2 + b_1^3$$

$$z_2^3 = w_{21}^3 a_1^2 + w_{22}^3 a_2^2 + w_{23}^3 a_3^2 + b_2^3$$

$$a_i^3 = a(z_i^3) \quad (i = 1,\ 2)$$

要更新网络中的权值与bias，如果暴力求导，将会使得求解过程变得十分复杂。而借助导数的链式法则，可以很轻松地求出误差对于各个权重及bias的偏导，从而实现权值及bias的更新，进而逼近最优。

$$C=\frac{1}{2}\{(t_1-a_1^3)^2+(t_2-a_2^3)^2\}$$

$$\begin{aligned}&(\Delta w_{11}^2,\ \cdots,\ \Delta w_{11}^3,\ \cdots,\ \Delta b_1^2,\ \cdots,\ \Delta b_1^3,\ \cdots)\\&=-\eta\left(\frac{\partial C_{\mathrm{T}}}{\partial w_{11}^2},\ \cdots,\ \frac{\partial C_{\mathrm{T}}}{\partial w_{11}^3},\ \cdots,\ \frac{\partial C_{\mathrm{T}}}{\partial b_1^2},\ \cdots,\ \frac{\partial C_{\mathrm{T}}}{\partial b_1^3},\ \cdots\right)\end{aligned}$$

在这个过程中，神经节点的误差甚是关键：（可以理解为最后的误差本质上由前面的各个神经元节点的误差构成）

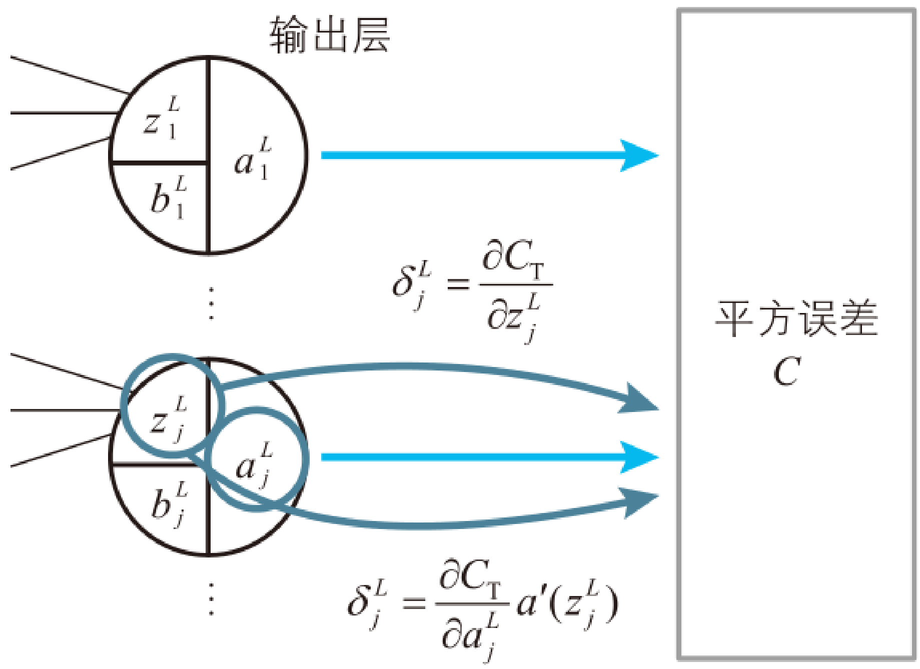
$$\delta_j^l=\frac{\partial C}{\partial z_j^l}\quad (l=2,\ 3,\ \cdots)$$

对于所有神经节点：

$$\frac{\partial C}{\partial w_{11}^3}=\frac{\partial C}{\partial z_1^3}\frac{\partial z_1^3}{\partial w_{11}^3}$$

$$\frac{\partial C}{\partial w_{ji}^l}=\delta_j^l a_i^{l-1},\ \frac{\partial C}{\partial b_j^l}=\delta_j^l\quad (l=2,\ 3,\ \cdots)$$

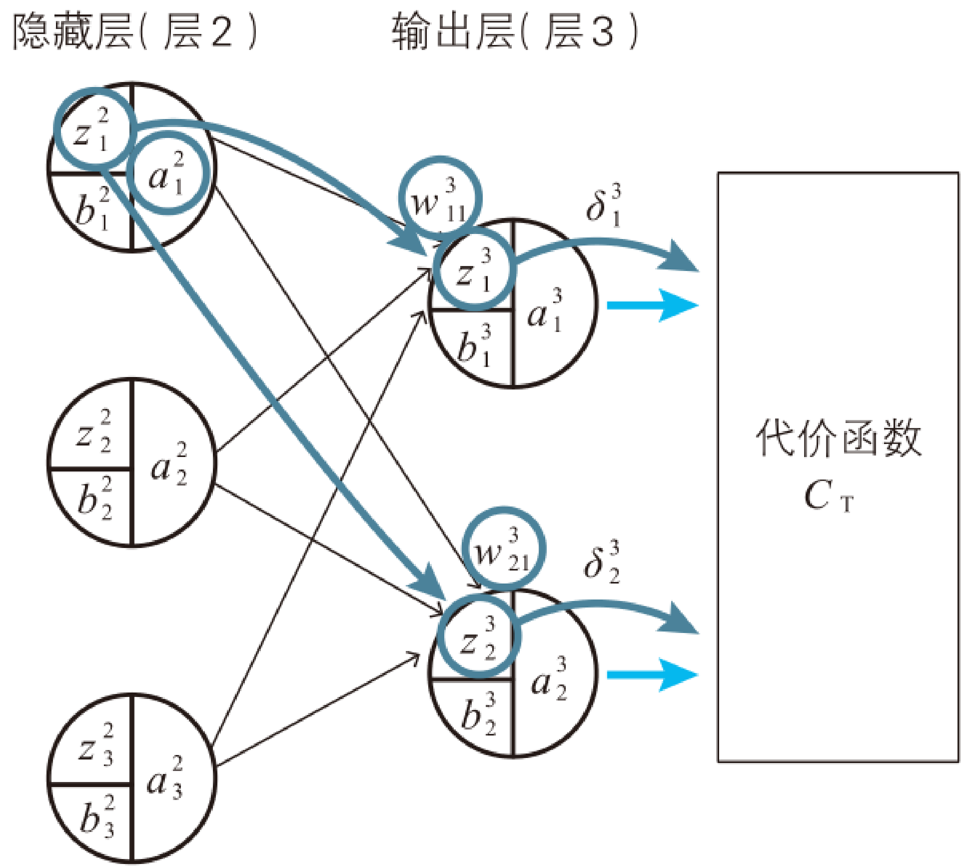
对于输出层，神经元误差可以很方便地求出来：



$$\delta_j^3 = \frac{\partial C}{\partial z_j^3} = \frac{\partial C}{\partial a_j^3}\frac{\partial a_j^3}{\partial z_j^3} = \frac{\partial C}{\partial a_j^3} a'(z_j^3)$$

$$\delta_j^L = \frac{\partial C}{\partial a_j^L} a'(z_j^L)$$

对于隐藏层，借助链式规则可以求出神经元误差：



$$\delta_1^2 = \frac{\partial C}{\partial z_1^2} = \frac{\partial C}{\partial z_1^3}\frac{\partial z_1^3}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2} + \frac{\partial C}{\partial z_2^3}\frac{\partial z_2^3}{\partial a_1^2}\frac{\partial a_1^2}{\partial z_1^2}$$

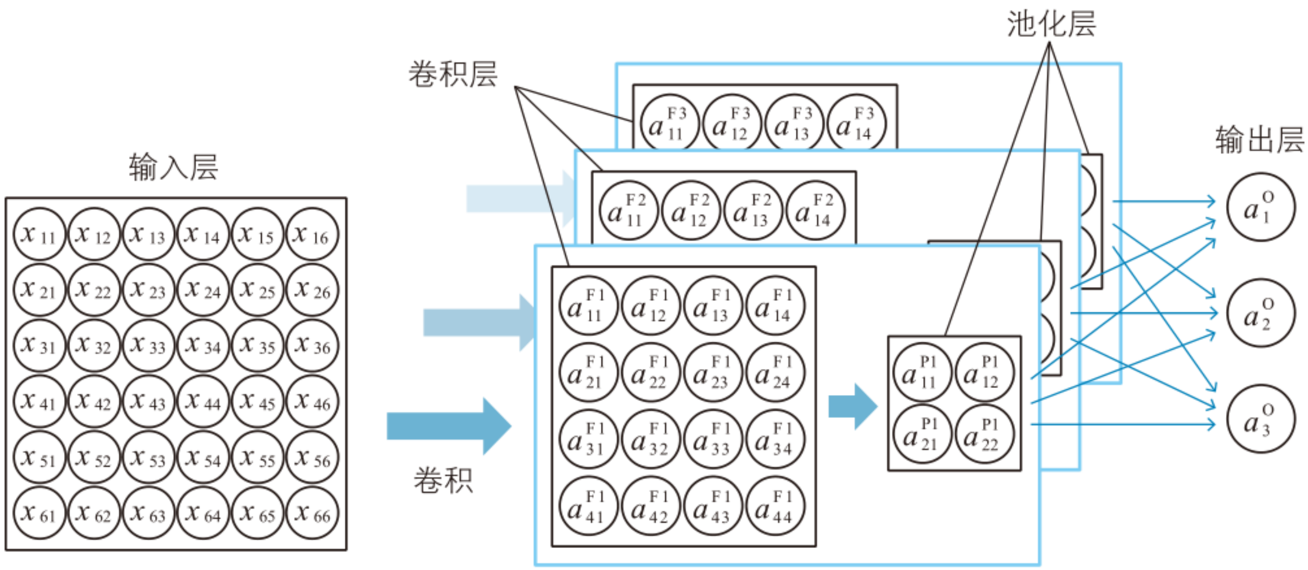
$$\delta_i^2 = (\delta_1^3 w_{1i}^3 + \delta_2^3 w_{2i}^3)a'(z_i^2) \quad (i=1,\ 2,\ 3)$$

$$\delta_i^l = \{\delta_1^{l+1} w_{1i}^{l+1} + \delta_2^{l+1} w_{2i}^{l+1} + \cdots + \delta_m^{l+1} w_{mi}^{l+1}\}a'(z_i^l)$$

当两类误差都可以很方便地求解出来时，便可以借助其与误差–权重（bias）的偏导的关系求出相应的更新值。不断重复这个过程便可以逐步逼近最优。

卷积神经网络的误差反向传播过程

对于这样一个卷积神经网络：



对于卷积层：

$$
\begin{aligned}
z_{ij}^{Fk} &= w_{11}^{Fk} x_{ij} + w_{12}^{Fk} x_{ij+1} + w_{13}^{Fk} x_{ij+2} \\
&\quad + w_{21}^{Fk} x_{i+1\,j} + w_{22}^{Fk} x_{i+1\,j+1} + w_{23}^{Fk} x_{i+1\,j+2} \\
&\quad + w_{31}^{Fk} x_{i+2\,j} + w_{32}^{Fk} x_{i+2\,j+1} + w_{33}^{Fk} x_{i+2\,j+2} + b^{Fk} \\
a_{ij}^{Fk} &= a(z_{ij}^{Fk})
\end{aligned}
$$

对于池化层：

$$
\begin{aligned}
z_{ij}^{Pk} &= \mathrm{Max}(a_{2i-1\,2j-1}^{Pk}，a_{2i-1\,2j}^{Pk}，a_{2i\,2j-1}^{Pk}，a_{2i\,2j}^{Pk}) \\
a_{ij}^{Pk} &= z_{ij}^{Pk}
\end{aligned}
$$

对于输出层：

$$
\begin{aligned}
z_n^{\mathrm{O}} = {} & w_{1-11}^{\mathrm{O}n} a_{11}^{\mathrm{P1}} + w_{1-12}^{\mathrm{O}n} a_{12}^{\mathrm{P1}} + w_{1-21}^{\mathrm{O}n} a_{21}^{\mathrm{P1}} + w_{1-22}^{\mathrm{O}n} a_{22}^{\mathrm{P1}} \\
& + w_{2-11}^{\mathrm{O}n} a_{11}^{\mathrm{P2}} + w_{2-12}^{\mathrm{O}n} a_{12}^{\mathrm{P2}} + w_{2-21}^{\mathrm{O}n} a_{21}^{\mathrm{P2}} + w_{2-22}^{\mathrm{O}n} a_{22}^{\mathrm{P2}} \\
& + w_{3-11}^{\mathrm{O}n} a_{11}^{\mathrm{P3}} + w_{3-12}^{\mathrm{O}n} a_{12}^{\mathrm{P3}} + w_{3-21}^{\mathrm{O}n} a_{21}^{\mathrm{P3}} + w_{3-22}^{\mathrm{O}n} a_{22}^{\mathrm{P3}} + b_n^{\mathrm{O}} \\
a_n^{\mathrm{O}} = {} & a(z_n^{\mathrm{O}})
\end{aligned}
$$

平方误差：

$$
C = \frac{1}{2}\{(t_1 - a_1^{\mathrm{O}})^2 + (t_2 - a_2^{\mathrm{O}})^2 + (t_3 - a_3^{\mathrm{O}})^2\}
$$

同样地，也是借助于导数的链式法则实现权重以及bias的更新。这个过程比较tricky的是各个上标及下标，很容易混乱。最好还是借助一个简单的网络图形来帮助理清各个符号之前的关系。
最终的目标：

$$
= \left(\underbrace{\frac{\partial C_{\mathrm{T}}}{\partial w_{11}^{\mathrm{F1}}}, \quad \cdots \quad}_{\text{关于过滤器的偏导数}} , \underbrace{\frac{\partial C_{\mathrm{T}}}{\partial w_{1-11}^{\mathrm{O1}}}, \quad \cdots \quad}_{\text{关于输出层神经单元的权重的偏导数}} , \underbrace{\frac{\partial C_{\mathrm{T}}}{\partial b^{\mathrm{F1}}}, \quad \cdots \quad}_{\text{关于卷积层神经单元的偏置的偏导数}} , \underbrace{\frac{\partial C_{\mathrm{T}}}{\partial b_1^{\mathrm{O}}}, \quad \cdots \quad ,}_{\text{关于输出层神经单元的偏置的偏导数}} \right)
$$

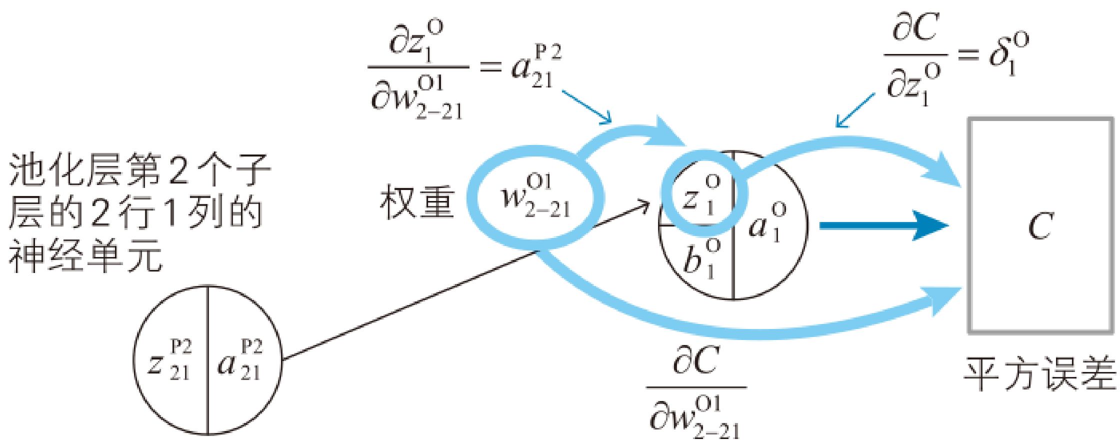
思路跟一般的神经网络比较类似。不同的在于卷积网络的存在共享权值，因此在求偏导时注意考虑到所有与某一自变量相关的所有项。关键的神经元误差：

$$
\delta_{ij}^{\mathrm{F}k} = \frac{\partial C}{\partial z_{ij}^{\mathrm{F}k}}, \quad \delta_n^{\mathrm{O}} = \frac{\partial C}{\partial z_n^{\mathrm{O}}}
$$

输出层神经元误差与权重更新分量的关系：

$$\frac{\partial C}{\partial w_{2-21}^{\mathrm{O1}}} = \frac{\partial C}{\partial z_1^{\mathrm{O}}} \frac{\partial z_1^{O}}{\partial w_{2-21}^{\mathrm{O1}}} = \delta_1^{\mathrm{O}} a_{21}^{\mathrm{P2}}$$

$$\frac{\partial C}{\partial b_1^{\mathrm{O}}} = \frac{\partial C}{\partial z_1^{\mathrm{O}}} \frac{\partial z_1^{\mathrm{O}}}{\partial b_1^{\mathrm{O}}} = \delta_1^{\mathrm{O}}$$



$$\frac{\partial C}{\partial w_{k-ij}^{\mathrm{O}n}} = \delta_n^{\mathrm{O}} a_{ij}^{\mathrm{P}k}, \quad \frac{\partial C}{\partial b_n^{\mathrm{O}}} = \delta_n^{\mathrm{O}}$$

卷积层神经元误差与权重更新分量的关系：

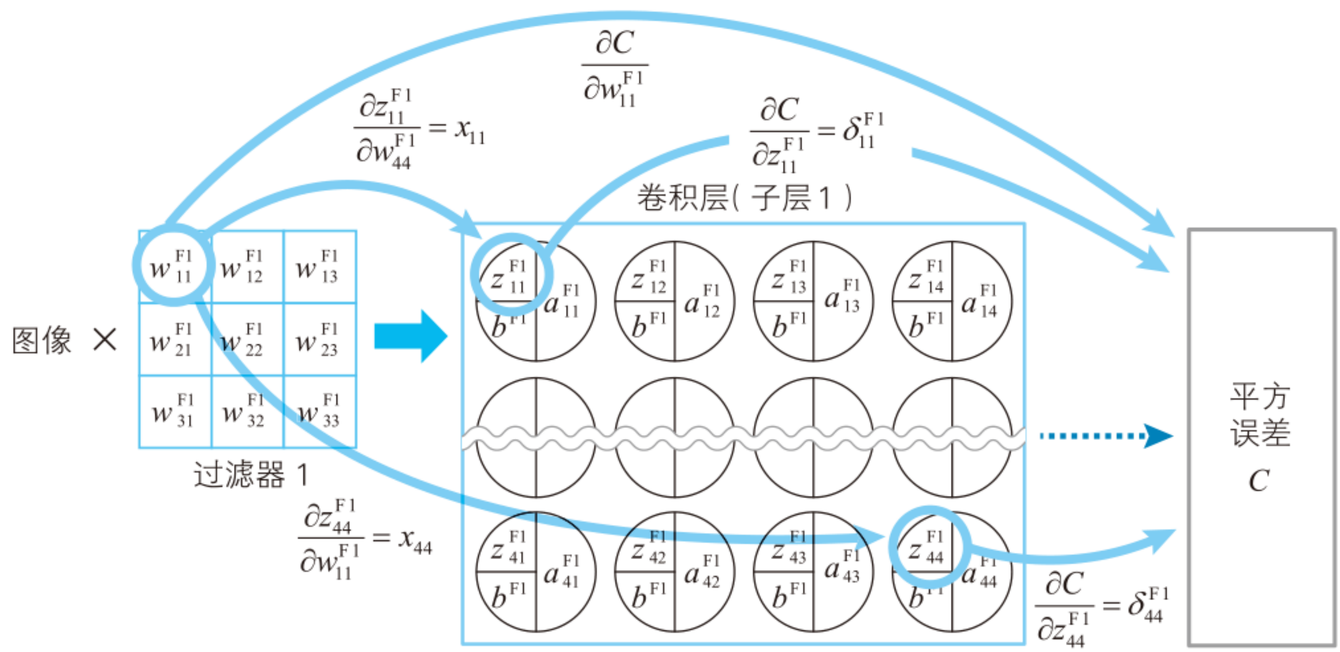
$$\begin{aligned} z_{11}^{\mathrm{F1}} &= w_{11}^{\mathrm{F1}} x_{11} + w_{12}^{\mathrm{F1}} x_{12} + w_{13}^{\mathrm{F1}} x_{13} + w_{21}^{\mathrm{F1}} x_{21} + w_{22}^{\mathrm{F1}} x_{22} + w_{23}^{\mathrm{F1}} x_{23} \\ &\quad + w_{31}^{\mathrm{F1}} x_{31} + w_{32}^{\mathrm{F1}} x_{32} + w_{33}^{\mathrm{F1}} x_{33} + b^{\mathrm{F1}} \\ z_{12}^{\mathrm{F1}} &= w_{11}^{\mathrm{F1}} x_{12} + w_{12}^{\mathrm{F1}} x_{13} + w_{13}^{\mathrm{F1}} x_{14} + w_{21}^{\mathrm{F1}} x_{22} + w_{22}^{\mathrm{F1}} x_{23} + w_{23}^{\mathrm{F1}} x_{24} \\ &\quad + w_{31}^{\mathrm{F1}} x_{32} + w_{32}^{\mathrm{F1}} x_{33} + w_{33}^{\mathrm{F1}} x_{34} + b^{\mathrm{F1}} \end{aligned}$$

……

$$\begin{aligned} z_{44}^{\mathrm{F1}} &= w_{11}^{\mathrm{F1}} x_{44} + w_{12}^{\mathrm{F1}} x_{45} + w_{13}^{\mathrm{F1}} x_{46} + w_{21}^{\mathrm{F1}} x_{54} + w_{22}^{\mathrm{F1}} x_{55} + w_{23}^{\mathrm{F1}} x_{56} \\ &\quad + w_{31}^{\mathrm{F1}} x_{64} + w_{32}^{\mathrm{F1}} x_{65} + w_{33}^{\mathrm{F1}} x_{66} + b^{\mathrm{F1}} \end{aligned}$$

$$\frac{\partial C}{\partial w_{11}^{\mathrm{F1}}} = \frac{\partial C}{\partial z_{11}^{\mathrm{F1}}} \frac{\partial z_{11}^{\mathrm{F1}}}{\partial w_{11}^{\mathrm{F1}}} + \frac{\partial C}{\partial z_{12}^{\mathrm{F1}}} \frac{\partial z_{12}^{\mathrm{F1}}}{\partial w_{11}^{\mathrm{F1}}} + \cdots + \frac{\partial C}{\partial z_{44}^{\mathrm{F1}}} \frac{\partial z_{44}^{\mathrm{F1}}}{\partial w_{11}^{\mathrm{F1}}}$$

$$\frac{\partial C}{\partial w_{11}^{\mathrm{F1}}} = \delta_{11}^{\mathrm{F1}} x_{11} + \delta_{12}^{\mathrm{F1}} x_{12} + \cdots + \delta_{44}^{\mathrm{F1}} x_{44}$$



$$\frac{\partial C}{\partial w_{ij}^{\mathrm{F}k}} = \delta_{11}^{\mathrm{F}k} x_{ij} + \delta_{12}^{\mathrm{F}k} x_{ij+1} + \cdots + \delta_{44}^{\mathrm{F}k} x_{i+3\,j+3}$$

$$\frac{\partial C}{\partial b^{\mathrm{F}k}} = \delta_{11}^{\mathrm{F}k} + \delta_{12}^{\mathrm{F}k} + \cdots + \delta_{33}^{\mathrm{F}k} + \cdots + \delta_{44}^{\mathrm{F}k}$$

计算输出层的神经元误差：

$$\delta_n^{\mathrm{O}} = \frac{\partial C}{\partial z_n^{\mathrm{O}}} = \frac{\partial C}{\partial a_n^{\mathrm{O}}}\frac{\partial a_n^{\mathrm{O}}}{\partial z_n^{\mathrm{O}}} = \frac{\partial C}{\partial a_n^{\mathrm{O}}} a'(z_n^{\mathrm{O}})$$

卷积层的神经元误差与输出层的神经元误差的关系：

$$\begin{aligned}\delta_{11}^{\mathrm{F1}} = \frac{\partial C}{\partial z_{11}^{\mathrm{F1}}} &= \frac{\partial C}{\partial z_1^{\mathrm{O}}}\frac{\partial z_1^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}}\frac{\partial a_{11}^{\mathrm{P1}}}{\partial z_{11}^{\mathrm{P1}}}\frac{\partial z_{11}^{\mathrm{P1}}}{\partial a_{11}^{\mathrm{F1}}}\frac{\partial a_{11}^{\mathrm{F1}}}{\partial z_{11}^{\mathrm{F1}}} \\ &+ \frac{\partial C}{\partial z_2^{\mathrm{O}}}\frac{\partial z_2^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}}\frac{\partial a_{11}^{\mathrm{P1}}}{\partial z_{11}^{\mathrm{P1}}}\frac{\partial z_{11}^{\mathrm{P1}}}{\partial a_{11}^{\mathrm{F1}}}\frac{\partial a_{11}^{\mathrm{F1}}}{\partial z_{11}^{\mathrm{F1}}} + \frac{\partial C}{\partial z_3^{\mathrm{O}}}\frac{\partial z_3^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}}\frac{\partial a_{11}^{\mathrm{P1}}}{\partial z_{11}^{\mathrm{P1}}}\frac{\partial z_{11}^{\mathrm{P1}}}{\partial a_{11}^{\mathrm{F1}}}\frac{\partial a_{11}^{\mathrm{F1}}}{\partial z_{11}^{\mathrm{F1}}}\end{aligned}$$

$$\delta_{11}^{\mathrm{F1}} = \left\{\frac{\partial C}{\partial z_1^{\mathrm{O}}}\frac{\partial z_1^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}} + \frac{\partial C}{\partial z_2^{\mathrm{O}}}\frac{\partial z_2^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}} + \frac{\partial C}{\partial z_3^{\mathrm{O}}}\frac{\partial z_3^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}}\right\}\frac{\partial a_{11}^{\mathrm{P1}}}{\partial z_{11}^{\mathrm{P1}}}\frac{\partial z_{11}^{\mathrm{P1}}}{\partial a_{11}^{\mathrm{F1}}}\frac{\partial a_{11}^{\mathrm{F1}}}{\partial z_{11}^{\mathrm{F1}}}$$

$$\frac{\partial z_1^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}} = w_{1-11}^{\mathrm{O1}},\ \ \frac{\partial z_2^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}} = w_{1-11}^{\mathrm{O2}},\ \ \frac{\partial z_3^{\mathrm{O}}}{\partial a_{11}^{\mathrm{P1}}} = w_{1-11}^{\mathrm{O3}}$$

$$a_{11}^{\mathrm{P1}} = z_{11}^{\mathrm{P1}},\ z_{11}^{\mathrm{P1}} = \mathrm{Max}(a_{11}^{\mathrm{F1}},\ a_{12}^{\mathrm{F1}},\ a_{21}^{\mathrm{F1}},\ a_{22}^{\mathrm{F1}})$$

$$\frac{\partial a_{11}^{\mathrm{P1}}}{\partial z_{11}^{\mathrm{P1}}} = 1$$

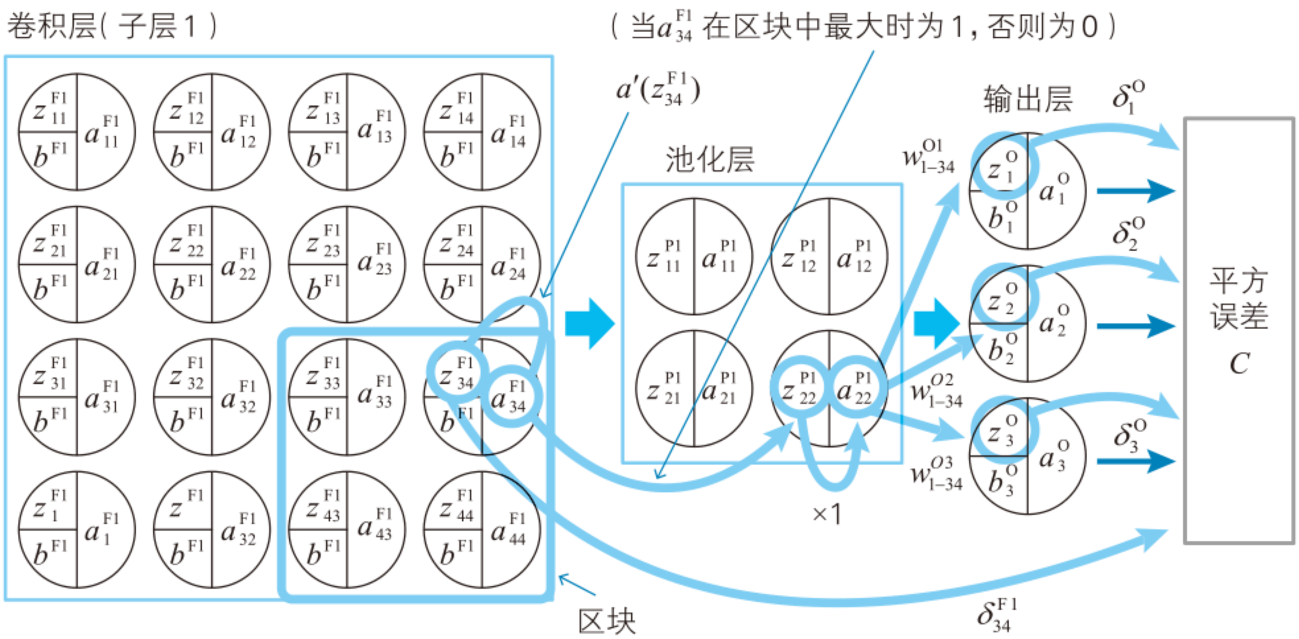
$$\frac{\partial z_{11}^{\mathrm{P1}}}{\partial a_{11}^{\mathrm{F1}}}=\begin{cases}1 & （在区块中 a_{11}^{\mathrm{F1}} 是最大时）\\ 0 & （在区块中 a_{11}^{\mathrm{F1}} 不是最大时）\end{cases}$$

$$\delta_{11}^{\mathrm{F1}}=\{\delta_1^{\mathrm{O}}w_{1-11}^{\mathrm{O1}}+\delta_2^{\mathrm{O}}w_{1-11}^{\mathrm{O2}}+\delta_3^{\mathrm{O}}w_{1-11}^{\mathrm{O3}}\}\times 1$$

$$\times（当 a_{11}^{\mathrm{F1}} 在区块中最大时为 1，否则为 0）\times a'(z_{11}^{\mathrm{F1}})$$

$$\delta_{34}^{\mathrm{F1}}=\{\delta_1^{\mathrm{O}}w_{1-22}^{\mathrm{O1}}+\delta_2^{\mathrm{O}}w_{1-22}^{\mathrm{O2}}+\delta_3^{\mathrm{O}}w_{1-22}^{\mathrm{O3}}\}$$

$$\times（当 a_{34}^{\mathrm{F1}} 在区块中最大时为 1，否则为 0）\times a'(z_{34}^{\mathrm{F1}})$$



$$\delta_{ij}^{\mathrm{F}k}=\{\delta_1^{\mathrm{O}}w_{k-i'j'}^{\mathrm{O1}}+\delta_2^{\mathrm{O}}w_{k-i'j'}^{\mathrm{O2}}+\delta_3^{\mathrm{O}}w_{k-i'j'}^{\mathrm{O3}}\}$$

$$\times（当 a_{ij}^{\mathrm{F}k} 在区块中最大时为 1，否则为 0）\times a'(z_{ij}^{\mathrm{F}k})$$

2.结果输出

今天把《深度学习中的数学》一书看完了，不禁感慨，一本好的教材对于激发一个人的学习热情、引领读者进入一个知识领域的重要性。整本书没有任何装腔作势，用最平实朴素的语言，从最基本的数学出发，把神经网络这个看起来就很艰深的话题将的清清楚楚明明白白，实现了让读者“看得懂”这样一个很多教材都达不到的目标。如果我几年前看到这本书，后续现在对机器学习的认知已经是另一个层次了。明天开始看李牧的《动手学深度学习》，主要用python实现，之前似乎配置环境的时候碰到了一些问题。当这本书看完便要着手用C++设计一个深度学习的API。争取在5月15号前完成这两项任务。之后可能还是得回归到计算机网络的学习，毕竟这个跟工作的相关性比较大。后续可能还有基于QT的软件开发，希望能达到随心所欲开发一些小工具的水平。

时间过得飞快，转眼在宿舍快待一个多月了。回顾这一个多月，学习成果堪忧，只在C++的学习上投入了一点点精力，包括计算机网络、数学、基础大件等的学习全面停滞，C++也远没达到得心应手的地步，实战太少太少。是该好好反思一下自己的学习态度跟学习方法。另外，疫情依旧烦人，学校隔断时间便传出阳性案例，也不知道何年何月才能解封。接下来除了把相关的知识学起来外，更重要的是把体重降下来，以应付好入职体检。