作业3：

1. (1)：① $\frac{32-5}{1}+1=28$，故为 $8\times28\times28$

② $\frac{28-2}{2}+1=14$，故为 $8\times14\times14$

③ $\frac{14-5}{1}+1=10$，故为 $16\times10\times10$

④ $\frac{10-2}{2}+1=5$，故为 $16\times5\times5$

(2)：① 5个　② 2个

2、~~$q_1=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}$ $q_2=\begin{pmatrix}3\\3\\3\end{pmatrix}$ $k_1=27$~~

~~$q_2=\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}$~~

$(q_1, q_2, q_3, q_4)=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}(a_1\ a_2\ a_3\ a_4)=\begin{pmatrix}3&6&9&12\\3&6&9&12\\3&6&9&12\end{pmatrix}$

~~$k_1=k_2=k_3=k_4=27$~~ 且有 $q=k=27$

$\therefore \alpha_{1,1}=\frac{q_1^T k_1}{\sqrt{3}}=\frac{27}{\sqrt{3}}=9\sqrt{3}$，

$\alpha_{1,2}=\frac{q_1^T k_2}{\sqrt{3}}=\frac{54}{\sqrt{3}}=18\sqrt{3}$，

$\alpha_{1,3}=\frac{q_1^T k_3}{\sqrt{3}}=\frac{81}{\sqrt{3}}=27\sqrt{3}$，

$\alpha_{1,4}=\frac{q_1^T k_4}{\sqrt{3}}=\frac{108}{\sqrt{3}}=36\sqrt{3}$ . . . . . .

date. page.

$\sum_{\vec{v}} \exp(\alpha_{1,1\vec{v}}) = e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}} + e^{36\sqrt{3}} = e^{9\sqrt{3}}(1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}})$

$\therefore \hat{\alpha}_{1,1} = \frac{\alpha_{1,1}}{\sum} = \frac{1}{1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}}}$

$\hat{\alpha}_{1,2} = \frac{\alpha_{1,2}}{\sum} = \frac{e^{9\sqrt{3}}}{1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}}}$

$\hat{\alpha}_{1,3} = \frac{\alpha_{1,3}}{\sum} = \frac{e^{18\sqrt{3}}}{1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}}}$

$\hat{\alpha}_{1,4} = \frac{e^{27\sqrt{3}}}{1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}}}$

$\therefore b_1 = \sum \hat{\alpha}_{1,\vec{v}} \vec{v}_{\vec{v}} = \frac{1}{1 + e^{9\sqrt{3}} + e^{18\sqrt{3}} + e^{27\sqrt{3}}} \begin{pmatrix} 3 + 6e^{9\sqrt{3}} + 9e^{18\sqrt{3}} + 12e^{27\sqrt{3}} \\ 3 + 6e^{9\sqrt{3}} + 9e^{18\sqrt{3}} + 12e^{27\sqrt{3}} \\ 3 + 6e^{9\sqrt{3}} + 9e^{18\sqrt{3}} + 12e^{27\sqrt{3}} \end{pmatrix}$

3、1) 相 LSTM 通过在系统中增加 gates 的方式来控制信息的流量（增加/删除），从而使得在回溯前时候不会出现 gradient 增长过快而 explode 或过小而 vanish 的情况。

2) ① 大小为 1，减小了 CNN 模型的参数个数

② 不会丢失图片信息，即不会像 1 的 kernel 那样出现图片边缘 pixels 缺失，即尺寸不变

③ 无论再维护长、宽信息，只需修改卷积核个数即可实现进行升维/降维。