线性网络神经元可视化

Leoeon

2017年04月29日

1 单神经元

对单神经元$\vec{w}$，使其激活最大的输入$\check{x}^*$为

$$\begin{aligned} \check{x}^* &= \mathrm{argmax}_{\check{x}} \vec{w}\check{x} \\ s.t. \quad & ||\check{x}^*||_2 = \mathrm{const} \end{aligned}$$

则

$$\check{x}^* \propto \vec{w}^T$$

2 多层线性神经网络

多隐层网络，则输出

$$o = \begin{bmatrix} \cdots & w_{ok} & \cdots \end{bmatrix} \begin{bmatrix} & \vdots & \\ \cdots & w_{kj} & \cdots \\ & \vdots & \end{bmatrix} \begin{bmatrix} & \vdots & \\ \cdots & w_{ji} & \cdots \\ & \vdots & \end{bmatrix} \begin{bmatrix} \vdots \\ x_i \\ \vdots \end{bmatrix}$$

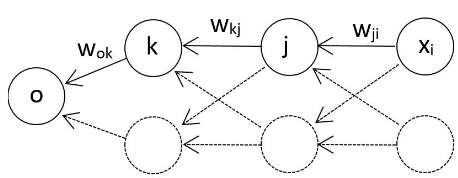

Figure 1: visualization1

则可得

$$
\begin{aligned}
\begin{bmatrix} \vdots \\ x_i^* \\ \vdots \end{bmatrix} &\propto \left(\begin{bmatrix} \cdots & w_{ok} & \cdots \end{bmatrix} \begin{bmatrix} & \vdots & \\ \cdots & w_{kj} & \cdots \\ & \vdots & \end{bmatrix} \begin{bmatrix} & \vdots & \\ \cdots & w_{ji} & \cdots \\ & \vdots & \end{bmatrix} \right)^T \\
&= \begin{bmatrix} & \vdots & \\ \cdots & w_{ji} & \cdots \\ & \vdots & \end{bmatrix}^T \begin{bmatrix} & \vdots & \\ \cdots & w_{kj} & \cdots \\ & \vdots & \end{bmatrix}^T \begin{bmatrix} \cdots & w_{ok} & \cdots \end{bmatrix}^T
\end{aligned}
$$

等价于倒转网络，从o输入定值，乘以各w，最终得到$\check{x}^*$

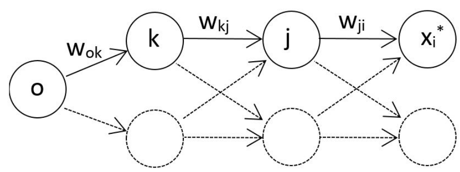

Figure 2: visualization2

3 卷积层

把卷积扩展为全连接形式。

例如，将$\begin{array}{|c|c|}\hline y_{11} & y_{12} \\ \hline y_{21} & y_{22} \\ \hline \end{array} = \begin{array}{|c|c|}\hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline \end{array} * \begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}$扩展为

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{21} \\ x_{22} \\ x_{23} \\ x_{31} \\ x_{32} \\ x_{33} \end{bmatrix}
$$

则

$$\begin{bmatrix} x_{11}^* \\ x_{12}^* \\ x_{13}^* \\ x_{21}^* \\ x_{22}^* \\ x_{23}^* \\ x_{31}^* \\ x_{32}^* \\ x_{33}^* \end{bmatrix} = \begin{bmatrix} w_{11} & 0 & 0 & 0 \\ w_{12} & w_{11} & 0 & 0 \\ 0 & w_{12} & 0 & 0 \\ w_{21} & 0 & w_{11} & 0 \\ w_{22} & w_{21} & w_{12} & w_{11} \\ 0 & w_{22} & 0 & w_{12} \\ 0 & 0 & w_{21} & 0 \\ 0 & 0 & w_{22} & w_{21} \\ 0 & 0 & 0 & w_{22} \end{bmatrix} \begin{bmatrix} y_{11}^* \\ y_{12}^* \\ y_{21}^* \\ y_{22}^* \end{bmatrix}$$

即当x_{ij}在卷积核移动到不同位置时卷积得不同的$y_{kl} = w_{pq}x_{ij} + \cdots$，则$x_{ij}^* = \sum w_{pq}y_{kl}^*$