Lecture 11

§1 Affine transformation as fully-connected feedforward ANNs

1. **Definition:** Fully-connected feedforward affine transformation ANNs (2.3.1)

令 ① $m, n \in N$

② $W \in R^{m\times n}$, $B \in R^m$

则记 fully-connected feedforward affine transformation ANNs 为以下 FNN:

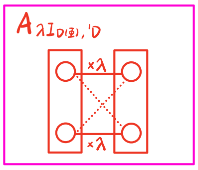
$$A_{W,B} = (W, B) \in (R^{m\times n} \times R^m) \subseteq N$$

注: $H(A_{W,B}) = 0$

2. **Lemma:** Fully-connected feedforward affine transformation ANNs 的 realizations (2.3.2)

令 ① $m, n \in N$

② $W \in R^{m\times n}$, $B \in R^m$



则 ① $D(A_{W,B}) = (n, m) \in N^2$

② 对 $\forall\, a \in C(R, R)$, 有

$$R_a^N(A_{W,B}) \in C(R^n, R^m)$$

③ 对 $\forall\, a \in C(R, R)$, $x \in R^n$, 有

$$(R_a^N(A_{W,B}))(x) = Wx + B$$ ($H(A_{W,B}) = 0$, 因此 activation 并没有用上)

证明:

Proof of Lemma 2.3.2. Note that the fact that $\mathbf{A}_{W,B} \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ shows that

$$\mathcal{D}(\mathbf{A}_{W,B}) = (n, m) \in \mathbb{N}^2. \tag{2.119}$$

This proves item (i). Furthermore, observe that the fact that

$$\mathbf{A}_{W,B} = (W, B) \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m) \tag{2.120}$$

and (1.91) ensure that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a^{\mathbf{N}}(\mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$(\mathcal{R}_a^{\mathbf{N}}(\mathbf{A}_{W,B}))(x) = Wx + B. \tag{2.121}$$

This establishes items (ii) and (iii). ~~The proof of Lemma 2.3.2 is thus complete.~~ The proof of Lemma 2.3.2 is thus complete. □

3. **Lemma:** Fully-connected feedforward affine transformation ANNs 的 compositions (2.3.3)

令 ① $\Phi \in N$

则 ① 对 $\forall\, m \in N$, $W \in R^{m\times D(\Phi)}$, $B \in R^m$, 有

$$D(A_{W,B} \cdot \Phi) = (D_0(\Phi), D_1(\Phi), \cdots, D_{H(\Phi)}(\Phi), m)$$

② 对 $\forall\, a \in C(R, R)$, $m \in N$, $W \in R^{m\times D(\Phi)}$, $B \in R^m$, 有

$$R_a^N(A_{W,B} \cdot \Phi) \in C(R^{I(\Phi)}, R^m)$$

③ 对 $\forall\, a \in C(R, R)$, $m \in N$, $W \in R^{m\times D(\Phi)}$, $B \in R^m$, 有

$$(R_a^N(A_{W,B} \cdot \Phi))(x) = W((R_a^N(\Phi))(x)) + B$$

④ 对 $\forall\, n \in N$, $W \in R^{I(\Phi)\times n}$, $B \in R^{I(\Phi)}$, 有

$$D(\Phi \cdot A_{W,B}) = (n, D_1(\Phi), D_2(\Phi), \cdots, D_{L(\Phi)}(\Phi))$$

⑤ 对 $\forall\ a\in C(\mathbb{R},\mathbb{R}),\ n\in\mathbb{N},\ W\in\mathbb{R}^{\mathcal{I}(\Phi)\times n},\ B\in\mathbb{R}^{\mathcal{I}(\Phi)}$，有

$$\mathcal{R}_a^{\mathbf{N}}(\Phi\cdot \mathbf{A}_{W,B})\in C(\mathbb{R}^n,\mathbb{R}^{\mathcal{O}(\Phi)})$$

⑥* 对 $\forall\ a\in C(\mathbb{R},\mathbb{R}),\ n\in\mathbb{N},\ W\in\mathbb{R}^{\mathcal{I}(\Phi)\times n},\ B\in\mathbb{R}^{\mathcal{I}(\Phi)}$，有

$$(\mathcal{R}_a^{\mathbf{N}}(\Phi\cdot \mathbf{A}_{W,B}))(x)=(\mathcal{R}_a^{\mathbf{N}}(\Phi))(Wx+B)$$

证明:

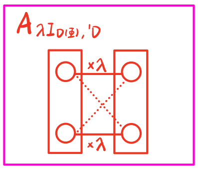
Proof of Lemma 2.3.3. Note that Lemma 2.3.2 implies that for all $m,n\in\mathbb{N}$, $W\in\mathbb{R}^{m\times n}$, $B\in\mathbb{R}^m$, $a\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^n$ it holds that $\mathcal{R}_a^{\mathbf{N}}(\mathbf{A}_{W,B})\in C(\mathbb{R}^n,\mathbb{R}^m)$ and

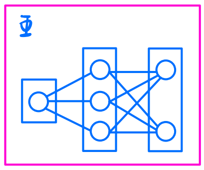
$$(\mathcal{R}_a^{\mathbf{N}}(\mathbf{A}_{W,B}))(x)=Wx+B \tag{2.126}$$

(cf. Definitions 1.3.4 and 2.3.1). Combining this and Proposition 2.1.2 proves items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 2.3.3 is thus complete. □

§2 Fully-connected feedforward ANNs 的 scalar multiplication

1. Definition: FNN 的 scalar multiplication (2.3.4)

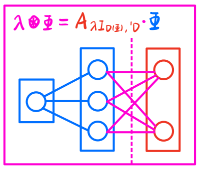




记以下function为 FNN 的 scalar multiplication:

$$(\cdot)\circledast(\cdot):\mathbb{R}\times\mathbf{N}\to\mathbf{N}$$

其满足: 对 $\forall\ \lambda\in\mathbb{R},\ \Phi\in\mathbf{N}$，有



$$\lambda\circledast\Phi=\mathbf{A}_{\lambda I_{\mathcal{O}(\Phi)},0}\cdot\Phi$$

注: $\mathbf{A}_{\lambda I_{\mathcal{O}(\Phi)},0}$ 中的 $\lambda I_{\mathcal{O}(\Phi)}\in\mathbb{R}^{\mathcal{O}(\Phi)\times\mathcal{O}(\Phi)}$，$0\in\mathbb{R}^{\mathcal{O}(\Phi)}$

e.g. $(\mathcal{R}_r^{\mathbf{N}}(\mathbf{A}_{3I_4,0}))(7,7,\dots,7)=\begin{bmatrix}3&0&0&0\\0&3&0&0\\0&0&3&0\\0&0&0&3\end{bmatrix}\begin{bmatrix}7\\7\\7\\7\end{bmatrix}=[21,21,21,21]$

2. Lemma: FNN 的 scalar multiplication 的性质 (2.3.5)

令 ① $\lambda\in\mathbb{R},\ \Phi\in\mathbf{N}$

则 ① $\mathcal{D}(\lambda\circledast\Phi)=\mathcal{D}(\Phi)$

② 对 $\forall\ a\in C(\mathbb{R},\mathbb{R})$，有

$$\mathcal{R}_a^{\mathbf{N}}(\lambda\circledast\Phi)\in C(\mathbb{R}^{\mathcal{I}(\Phi)},\mathbb{R}^{\mathcal{O}(\Phi)})$$

③ 对 $\forall\ a\in C(\mathbb{R},\mathbb{R}),\ x\in\mathbb{R}^{\mathcal{I}(\Phi)}$，有

$$(\mathcal{R}_a^{\mathbf{N}}(\lambda\circledast\Phi))(x)=\lambda\cdot\big((\mathcal{R}_a^{\mathbf{N}}(\Phi))(x)\big)$$

证明:

Proof of Lemma 2.3.5. Throughout this proof, let $L\in\mathbb{N}$, $l_0,l_1,\dots,l_L\in\mathbb{N}$ satisfy

$$L=\mathcal{L}(\Phi)\quad\text{and}\quad(l_0,l_1,\dots,l_L)=\mathcal{D}(\Phi). \tag{2.129}$$

Observe that item (i) in Lemma 2.3.2 demonstrates that

$$\mathcal{D}(\mathbf{A}_{\lambda\mathrm{I}_{\mathcal{O}(\Phi)},0})=(\mathcal{O}(\Phi),\mathcal{O}(\Phi)) \tag{2.130}$$

(cf. Definitions 1.5.5 and 2.3.1). Combining this and item (i) in Lemma 2.3.3 shows that

$$\mathcal{D}(\lambda\circledast\Phi)=\mathcal{D}(\mathbf{A}_{\lambda\mathrm{I}_{\mathcal{O}(\Phi)},0}\bullet\Phi)=(l_0,l_1,\dots,l_{L-1},\mathcal{O}(\Phi))=\mathcal{D}(\Phi) \tag{2.131}$$

(cf. Definitions 2.1.1 and 2.3.4). This establishes item (i). Note that items (ii) and (iii) in Lemma 2.3.3 ensure that for all $a\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^{\mathcal{I}(\Phi)}$ it holds that $\mathcal{R}_a^{\mathbf{N}}(\lambda\circledast\Phi)\in C(\mathbb{R}^{\mathcal{I}(\Phi)},\mathbb{R}^{\mathcal{O}(\Phi)})$ and

$$
\begin{aligned}
\big(\mathcal{R}_a^{\mathbf{N}}(\lambda \circledast \Phi)\big)(x) &= \big(\mathcal{R}_a^{\mathbf{N}}(\mathbf{A}_{\lambda\,\mathrm{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi)\big)(x) \\
&= \lambda\,\mathrm{I}_{\mathcal{O}(\Phi)}\big((\mathcal{R}_a^{\mathbf{N}}(\Phi))(x)\big) \\
&= \lambda\big((\mathcal{R}_a^{\mathbf{N}}(\Phi))(x)\big)
\end{aligned}
\tag{2.132}
$$

(cf. Definition 1.3.4). This proves items (ii) and (iii). The proof of Lemma 2.3.5 is thus complete. $\square$