第二章 向量和矩阵基础

第 4 讲 线性映射与线性变换

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

1 4.1 线性映射

2 4.2 线性映射的矩阵表示

3 4.3 线性变换

4 4.4 仿射映射

1 4.1 线性映射

2 4.2 线性映射的矩阵表示

3 4.3 线性变换

4 4.4 仿射映射

线性映射引例：手写数字识别

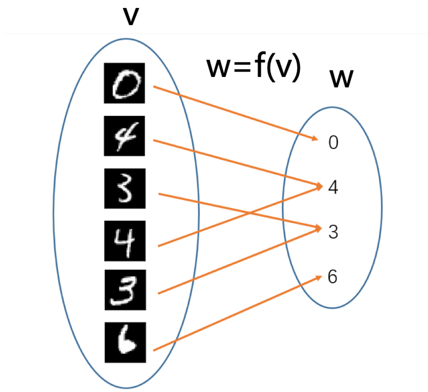

图 1: MNIST 数据集中的图片示例

集合之间的映射

定义 1

设 $\mathbb{V},\mathbb{W}$ 是两个非空集合，如果存在一个法则 f，使得对 $\mathbb{V}$ 中每个元素 v，按法则 f，在 $\mathbb{W}$ 中有唯一确定的元素 w 与之对应，则称 f 为从 $\mathbb{V}$ 到 $\mathbb{W}$ 的映射，记作

$$f:\mathbb{V}\to\mathbb{W},$$

其中 w 称为元素 v（在映射 f 下）的**像**，并记作 $f(v)$，即

$$w=f(v),$$

而元素 v 称为元素 w（在映射 f 下）的一个**原像**；集合 $\mathbb{V}$ 称为映射 f 的**定义域**，记作 D_f，即 $D_f=\mathbb{V}$；$\mathbb{V}$ 中所有元素的像所组成的集合称为映射 f 的**值域**，记作 R_f 或 $f(\mathbb{V})$，即

$$R_f=f(V)=\{f(v)|v\in\mathbb{V}\}.$$

集合之间的特殊映射

定义 2

设 $\mathbb{V}, \mathbb{W}$ 是任意两个集合，$\varphi : \mathbb{V} \to \mathbb{W}$ 是一个映射，如果 φ 满足

1. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V} : \varphi(\boldsymbol{x}) = \varphi(\boldsymbol{y}) \Longrightarrow \boldsymbol{x} = \boldsymbol{y}$，则 φ 称为**单射**；
2. $\varphi(\mathbb{V}) = \mathbb{W}$，则 φ 称为**满射**；
3. 即满足单射又满足满射，则 φ 称为**双射**。

例 1

下列映射中

1. $\Phi_1 : \mathbb{R}^n \to \mathbb{R}^n, \Phi_1(\boldsymbol{x}) = 2\boldsymbol{x}$ 既是单射又是满射，所以是双射。
2. $\Phi_2 : \mathbb{R}^n \to \mathbb{R}, \Phi_2(\boldsymbol{x}) = \max(x_i)$ 只是满射，不是单射。
3. $\Phi_3 : \mathbb{R}^n \to \mathbb{R}, \Phi_3(\boldsymbol{x}) = \max(|x_i|)$ 既不是单射也不是满射。

向量空间之间的线性映射

定义 3

设 $\mathbb{V}, \mathbb{W}$ 是数域 $\mathbb{K}$ 上的两个有限维的向量空间，φ 是 $\mathbb{V}$ 到 $\mathbb{W}$ 的一个映射（$\varphi: \mathbb{V} \to \mathbb{W}$）。如果对任何向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$ 及任意的 $\alpha, \beta \in \mathbb{K}$，有

$$\varphi(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha \varphi(\boldsymbol{x}) + \beta \varphi(\boldsymbol{y}), \tag{1}$$

则称 φ 为 $\mathbb{V}$ 到 $\mathbb{W}$ 的**线性映射**。

例 2

考虑映射 $\epsilon : \mathbb{V} \to \mathbb{V}, \epsilon(\boldsymbol{x}) = \boldsymbol{x}$，我们称这种映射为**恒等映射**

$$\epsilon(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha \boldsymbol{x} + \beta \boldsymbol{y} = \alpha \epsilon(\boldsymbol{x}) + \beta \epsilon(\boldsymbol{y})$$

恒等映射是线性映射。

例 3

考察映射 $\boldsymbol{T}:\mathbb{R}^3\to\mathbb{R}^2$

$$\boldsymbol{T}(\boldsymbol{x})=\begin{bmatrix} x_1+x_2 \\ x_1^2-x_2^2 \end{bmatrix},\quad \boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}}$$

令 $\boldsymbol{x}_1=(1,0,0)^{\mathrm{T}},\boldsymbol{x}_2=(2,0,0)^{\mathrm{T}}$,

$$\boldsymbol{T}(\boldsymbol{x}_1)=(1,1)^{\mathrm{T}},\ \boldsymbol{T}(\boldsymbol{x}_2)=(2,4)^{\mathrm{T}}$$

而

$$\boldsymbol{T}(\boldsymbol{x}_1+\boldsymbol{x}_2)=(3,9)^{\mathrm{T}}$$

映射 $\boldsymbol{T}:\mathbb{R}^3\to\mathbb{R}^2$ 不满足线性关系式，故不是线性映射。

例 4

考虑映射 $\boldsymbol{T}_{\boldsymbol{Q}}(\boldsymbol{A})=\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 是可逆矩阵。

$$\begin{aligned}\boldsymbol{T}_{\boldsymbol{Q}}(\alpha\boldsymbol{A}+\beta\boldsymbol{B})&=\boldsymbol{Q}^{-1}(\alpha\boldsymbol{A}+\beta\boldsymbol{B})\boldsymbol{Q}\\&=\alpha\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}+\beta\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}\\&=\alpha\boldsymbol{T}_{\boldsymbol{Q}}(\boldsymbol{A})+\beta\boldsymbol{T}_{\boldsymbol{Q}}(\boldsymbol{B})\end{aligned}$$

向量空间之间的特殊映射

接下来，我们就可以给出一些两个向量空间之间的特殊的映射。

定义 4

设 $\mathbb{V}, \mathbb{W}$ 是数域 $\mathbb{K}$ 上的两个有限维的向量空间：

1. $\varphi : \mathbb{V} \to \mathbb{W}$ 是线性映射，则 $\mathbb{V}, \mathbb{W}$ **同态** *(Homomorphism)*，φ 称为同态映射；
2. $\varphi : \mathbb{V} \to \mathbb{W}$ 是线性映射且是双射，则 $\mathbb{V}, \mathbb{W}$ **同构** *(Isomorphism)*，φ 称为同构映射；
3. $\varphi : \mathbb{V} \to \mathbb{V}$ 是线性映射，则 $\mathbb{V}$ **自同态** *(Endomorphism)*，φ 称为自同态映射；
4. $\varphi : \mathbb{V} \to \mathbb{V}$ 是线性映射且是双射，则 $\mathbb{V}$ **自同构** *(Automorphism)*，φ 称为自同构映射。

例 5

映射 $\varphi : \mathbb{R}^2 \to \mathbb{C}, \varphi(\boldsymbol{x}) = x_1 + ix_2$ 是同态映射，因为

$$\varphi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = (x_1 + y_1) + i(x_2 + y_2) = x_1 + ix_2 + y_1 + iy_2$$
$$= \varphi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \varphi\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$
$$\varphi\left(\lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \lambda x_1 + \lambda i x_2 = \lambda(x_1 + ix_2) = \lambda\varphi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right).$$

映射 $\varphi : \mathbb{R}^2 \to \mathbb{C}, \varphi(\boldsymbol{x}) = x_1 + ix_2$ 也是同构映射，因为它是一个双射。

定理 1

设 $\mathbb{X}_1, \mathbb{X}_2$ 是数域 $\mathbb{K}$ 上的两个有限维向量空间，$\mathbb{X}_1, \mathbb{X}_2$ 同构，当且仅当 $\dim(\mathbb{X}_1) = \dim(\mathbb{X}_2)$。

定理1表明了两个维数相同的向量空间之间存在着一个满足双射的线性映射，从这个观点看，同构的向量空间是可以不加区别的，维数是有限维向量空间的唯一本质特征。

也就是说对于 $\mathbb{R}^{m\times n}$(一个 $m \times n$ 维的矩阵的线性空间) 和 $\mathbb{R}^{mn}$(一个长度为 mn 的向量的线性空间) 其本质是一样的。我们可以通过一个同构映射建立起它们之间的联系。

定义 5

设 $\mathbb{V}, \mathbb{W}$ 是数域 $\mathbb{K}$ 上的两个有限维的向量空间，如果 $\varphi : \mathbb{V} \to \mathbb{W}$ 是一个双射，则可以定义它的**逆映射**，记作 $\varphi^{-1} : \mathbb{W} \to \mathbb{V}$，对于 $\forall \boldsymbol{x} \in \mathbb{V}$ 和 $\forall \boldsymbol{y} \in \mathbb{W}$ 使得

$$\varphi^{-1}(\varphi(\boldsymbol{x})) = \mathcal{E}(\boldsymbol{x}) = \boldsymbol{x}, \varphi(\varphi^{-1}(\boldsymbol{y})) = \mathcal{E}(\boldsymbol{y}) = \boldsymbol{y}$$

例 6

- 恒等映射的逆映射是其本身
- 在例4中定义了映射 $\boldsymbol{T}_{\boldsymbol{Q}}(\boldsymbol{A}) = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}$ 的逆映射为 $\boldsymbol{T}_{\boldsymbol{Q}^{-1}}$

定理 2

考虑向量空间 $\mathbb{V},\mathbb{W},\mathbb{X}$，则有

- 对于线性映射 $\varphi:\mathbb{V}\to\mathbb{W}$ 和 $\phi:\mathbb{W}\to\mathbb{X}$，则 $\phi(\varphi)$ 也是一个线性映射；
- 如果 $\varphi:\mathbb{V}\to\mathbb{W}$ 是同构映射，则 $\varphi^{-1}:\mathbb{W}\to\mathbb{V}$ 也是一个同构映射；
- 如果 $\varphi:\mathbb{V}\to\mathbb{W},\phi:\mathbb{V}\to\mathbb{W}$ 是线性映射，且 $\lambda\in\mathbb{R}$，则 $\varphi+\phi$ 和 $\lambda\varphi$ 也是线性映射。

用线性映射分类手写数字

- 在 MNIST 数字识别的例子中，我们把图像数据集看作 28×28 维向量空间 $\mathbb{W}$；所有标签向量在 10 维向量空间 $\mathbb{V}$ 中
- 想要找到一个线性映射，将 28×28 维向量空间映射到 10 维向量空间中去。假设图片向量为 $\boldsymbol{x}$，其标签向量为 $\boldsymbol{y}$，我们选择线性映射

$$f : \mathbb{W} \to \mathbb{V}$$

$$\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{x}$$

其中 $\boldsymbol{y}'$ 也是 $\mathbb{V}$ 中的向量，$\boldsymbol{A}$ 是参数矩阵，希望 $\boldsymbol{y}' \approx \boldsymbol{y}$

- 选择合适的损失函数，通过优化得到 $\boldsymbol{A}$

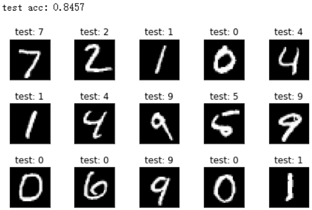

图 2: 线性映射分类准确率

线性映射复合分类手写数字

深度学习中，常常利用映射的复合构建更为强大、准确率更高的分类器，比如在 MNIST 数字识别的例子中，先用 f_1 将图像映射成 50 维的向量，再用 f_2 将 50 维的向量映射为 10 维的向量，但是如果我们利用线性映射的复合构造分类器，即

$$\boldsymbol{h}_1 = f_1(\boldsymbol{x}) = \boldsymbol{A}_1\boldsymbol{x}$$

$$\boldsymbol{y}' = f_2(\boldsymbol{h}_1) = \boldsymbol{A}_2\boldsymbol{h}_1$$

设 $\boldsymbol{A} = \boldsymbol{A}_2\boldsymbol{A}_1$，

$$\boldsymbol{y}' = f_2(f_1(\boldsymbol{x})) = \boldsymbol{A}_2\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}$$

得到的仍然是一个线性映射，并不能提高分类的准确性。

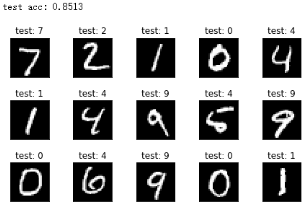

图 3: 双层线性网络准确率

1 4.1 线性映射

2 4.2 线性映射的矩阵表示

3 4.3 线性变换

4 4.4 仿射映射

4.2.1 变换矩阵

考虑一个 n 维向量空间 $\mathbb{V}$ 的基底 $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$，并为基向量规定一个顺序。

定义 6

若一个 n 维向量空间 $\mathbb{V}$ 的基底 $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ 是有序的，我们称 n 元数组 $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ 为向量空间 $\mathbb{V}$ 的一组**有序基**。

回顾：坐标

给定一个向量空间 $\mathbb{V}$ 和 $\mathbb{V}$ 的一个有序基底 $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, 任何的 $\boldsymbol{x} \in \mathbb{V}$，我们得到 $\boldsymbol{x}$ 关于 $\boldsymbol{B}$ 的唯一表示：

$$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \ldots + \alpha_n \boldsymbol{b}_n$$

那么系数组成的向量 $(\alpha_1, \ldots, \alpha_n)$ 就是 $\boldsymbol{x}$ 在 $\boldsymbol{B}$ 下的坐标。

我们称向量

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$$

是 $\boldsymbol{x}$ 关于 $\boldsymbol{B}$ 的坐标表示，即 $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\alpha}$。

例 7

考虑一个几何向量 $x \in \mathbb{R}^2$，坐标 $[2,3]^T$，

可以用标准基 $e_1, e_2 \in \mathbb{R}^2$ 来表示。

这意味着，我们可以把 x 表示为 $x = 2e_1 + 3e_2$。

然而，我们不必选择标准基来表示这个向量，

如果我们使用基向量 $b_1 = [1, -1]^T, b_2 = [1\ 1]^T$，

我们将获得坐标 $[-\frac{1}{2}, \frac{5}{2}]^T$ 来表示同一向量。

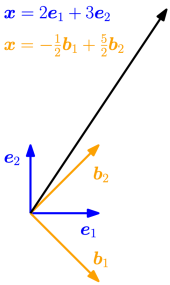

图 4: 不同基下的坐标

定义 7

考虑向量空间 $\mathbb{V},\mathbb{W}$ 的有序基 $\boldsymbol{B}=(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_n)$ 和 $\boldsymbol{C}=(\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m)$. 然后考虑一个线性映射 $\Phi:\mathbb{V}\to\mathbb{W}$, 对于 $j\in\{1,\ldots,n\}$

$$\Phi\left(\boldsymbol{b}_j\right)=a_{1j}\boldsymbol{c}_1+\cdots+a_{mj}\boldsymbol{c}_m=\sum_{i=1}^{m}a_{ij}\boldsymbol{c}_i$$

$\Phi(b_j)$ 是关于 $\boldsymbol{C}$ 的唯一表示，那么我们称这个 $m\times n$ 的矩阵为 $\boldsymbol{A}_\Phi$，它的元素为：

$$\boldsymbol{A}_\Phi(i,j)=a_{ij}$$

是 Φ 的**变换矩阵**，$\Phi(\boldsymbol{b}_j)$ 在 $\mathbb{W}$ 的有序基 C 下的坐标是 $\boldsymbol{A}_\Phi$ 的第 j 列。

记 $\Phi(\boldsymbol{B})=(\Phi(\boldsymbol{b}_1),\Phi(\boldsymbol{b}_2),\cdots,\Phi(\boldsymbol{b}_n))$，则

$$\Phi(\boldsymbol{B})=\boldsymbol{C}\boldsymbol{A}_\Phi.$$

设向量空间 $\mathbb{V}$，$\mathbb{W}$ 的有序基分别为 $\boldsymbol{B}, \boldsymbol{C}$，线性映射 $\Phi : \mathbb{V} \to \mathbb{W}$ 的变换矩阵 $\boldsymbol{A}_\Phi$，如果 $\boldsymbol{x} \in \mathbb{V}$ 关于 $\boldsymbol{B}$ 的坐标是 $\hat{\boldsymbol{x}}$，$\boldsymbol{y} = \Phi(\boldsymbol{x}) \in \mathbb{W}$ 关于 $\boldsymbol{C}$ 的坐标是 $\hat{\boldsymbol{y}}$:

$$\Phi(\boldsymbol{x}) = \Phi(\boldsymbol{B}\hat{\boldsymbol{x}}) = \Phi(\boldsymbol{B})\hat{\boldsymbol{x}} = \boldsymbol{C}\boldsymbol{A}_\Phi\hat{\boldsymbol{x}} = \boldsymbol{C}(\boldsymbol{A}_\Phi\hat{\boldsymbol{x}})$$

$\boldsymbol{A}_\Phi\hat{\boldsymbol{x}}$ 就是 $\Phi(\boldsymbol{x})$ 关于 $\boldsymbol{C}$ 的坐标，由此得到坐标的映射关系:

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_\Phi\hat{\boldsymbol{x}}$$

这意味着这个变换矩阵可以用来计算在两个空间各自基下坐标的映射关系。

例 8

考虑同态 $\Phi : \mathbb{V} \to \mathbb{W}$，$\mathbb{V}$ 有序基 $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$，$\mathbb{W}$ 的有序基 $\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_4)$，有：

$$\begin{aligned}
\Phi\left(\boldsymbol{b}_1\right) &= \boldsymbol{c}_1 - \boldsymbol{c}_2 + 3\boldsymbol{c}_3 - \boldsymbol{c}_4 \\
\Phi\left(\boldsymbol{b}_2\right) &= 2\boldsymbol{c}_1 + \boldsymbol{c}_2 + 7\boldsymbol{c}_3 + 2\boldsymbol{c}_4 \\
\Phi\left(\boldsymbol{b}_3\right) &= 3\boldsymbol{c}_2 + \boldsymbol{c}_3 + 4\boldsymbol{c}_4
\end{aligned}$$

其变换矩阵为：

$$\boldsymbol{A}_\Phi = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}$$

其中 $\boldsymbol{a}_j, j = 1, 2, 3$，是 $\Phi(\boldsymbol{b}_j)$ 关于 $\mathbb{W}$ 的有序基 $\boldsymbol{C}$ 的坐标。

4.2.2 基变换

考虑 $\mathbb{V}$ 的两个有序基底：

$$B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n), \tilde{B} = (\tilde{\boldsymbol{b}}_1, ..., \tilde{\boldsymbol{b}}_n)$$

和 $\mathbb{W}$ 的两个有序基底：

$$C = (\boldsymbol{c}_1, ..., \boldsymbol{c}_n), \tilde{C} = (\tilde{\boldsymbol{c}}_1, ..., \tilde{\boldsymbol{c}}_n).$$

$\boldsymbol{A}_\Phi \in \mathbb{R}^{m\times n}$ 是线性映射 $\Phi : \mathbb{V} \to \mathbb{W}$ 关于基底 B, C 的变换矩阵。
$\tilde{\boldsymbol{A}}_\Phi \in \mathbb{R}^{m\times n}$ 是线性映射 $\Phi : \mathbb{V} \to \mathbb{W}$ 关于基底 $\tilde{B}, \tilde{C}$ 的变换矩阵。

我们接下来考察 $\boldsymbol{A}_\Phi$ 和 $\tilde{\boldsymbol{A}}_\Phi$ 的关系是什么，也即当我们把基从 B, C 变换到 $\tilde{B}, \tilde{C}$ 时，$\boldsymbol{A}_\Phi$ 能否变换到 $\tilde{\boldsymbol{A}}_\Phi$、怎么变换到 $\tilde{\boldsymbol{A}}_\Phi$?

例 9

考虑基为 $\boldsymbol{e}_1, \boldsymbol{e}_2$ 的 $\mathbb{R}^2$ 上的变换矩阵 $\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$，如果我们将新的基底定义为

$$B = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$$

我们可以得到关于 B 的一个对角的变换矩阵

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

这个矩阵比 $\boldsymbol{A}$ 更容易处理。

定理 3

对于一个线性映射 $\Phi : \mathbb{V} \to \mathbb{W}$，$\mathbb{V}$ 的两个有序基底:

$$B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n), \tilde{B} = (\tilde{\boldsymbol{b}}_1, ..., \tilde{\boldsymbol{b}}_n)$$

和 $\mathbb{W}$ 的两个有序基底：

$$C = (\boldsymbol{c}_1, ..., \boldsymbol{c}_n), \tilde{C} = (\tilde{\boldsymbol{c}}_1, ..., \tilde{\boldsymbol{c}}_n)$$

以及在基 B, C 下关于 Φ 的变换矩阵 $\boldsymbol{A}_\Phi$，则在基 $\tilde{B}, \tilde{C}$ 下关于 Φ 的变换矩阵 $\tilde{\boldsymbol{A}}_\Phi$ 可以表示为

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S},$$

这里 $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ 是 $\mathbb{V}$ 中自同态映射的坐标变换矩阵（从 B 到 $\tilde{B}$），$\boldsymbol{T} \in \mathbb{R}^{m\times m}$ 是 $\mathbb{W}$ 中自同态映射的坐标变换矩阵（从 C 到 $\tilde{C}$）。

$$
\begin{tikzcd}
\text{向量空间} & V \arrow[r, "\Phi"] & W \\
& B \arrow[r, "\Phi_{CB}", "\boldsymbol{A}_\Phi"'] & C \\
\text{有序基底} & \tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\boldsymbol{S}"'] \arrow[r, "\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C} \arrow[u, "\boldsymbol{T}", "\Xi_{C\tilde{C}}"']
\end{tikzcd}
$$

$$
\begin{tikzcd}
\text{向量空间} & V \arrow[r, "\Phi"] & W \\
& B \arrow[r, "\Phi_{CB}", "\boldsymbol{A}_\Phi"'] & C \arrow[d, "\boldsymbol{T}^{-1}"', "\Xi_{\tilde{C}C} = \Xi^{-1}_{C\tilde{C}}"] \\
\text{有序基底} & \tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\boldsymbol{S}"'] \arrow[r, "\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C}
\end{tikzcd}
$$

图 5: 不同基底下的坐标关系

当我们分别在 $\mathbb{V}$ 中把基底从 B 变换为 $\tilde{B}$ 以及在 $\mathbb{W}$ 中把基底从 C 变换为 $\tilde{C}$，我们可以通过一些步骤来得到相应的变换矩阵 $\tilde{\boldsymbol{A}}_\Phi$。

- 首先，我们写出联系新基底 $\tilde{B}$ 下坐标和旧基底 B 下坐标的线性映射 $\Psi_{B\tilde{B}} : \mathbb{V} \to \mathbb{V}$ 所对应的矩阵表示。
- 然后我们再使用 Φ_{CB} 的变换矩阵 $\boldsymbol{A}_\Phi$ 将坐标映射到以 C 为基底的 $\mathbb{W}$ 中。
- 最后我们再使用线性映射 $\Xi_{\tilde{C}C} : \mathbb{W} \to \mathbb{W}$ 把坐标从用基底 C 表示到用基底 $\tilde{C}$ 表示。

因此我们可以将线性映射 $\Phi_{\tilde{C}\tilde{B}}$ 表示为:

$$\Phi_{\tilde{C}\tilde{B}} = \Xi_{\tilde{C}C} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}} = \Xi^{-1}_{C\tilde{C}} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}}$$

定义 8

如果对于两个矩阵 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$，存在可逆矩阵 $\boldsymbol{S} \in \mathbb{R}^{n\times n}$，$\boldsymbol{T} \in \mathbb{R}^{m\times m}$ 使得，$\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{S}$ 成立，则称 $\boldsymbol{A},\boldsymbol{B}$ **等价**。

定义 9

如果对于两个矩阵 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$，存在可逆矩阵 $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ 使得，$\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{S}$ 成立，则称 $\boldsymbol{A},\boldsymbol{B}$ **相似**。

所以两个相似的矩阵必定等价，反之则不然。

例 10

考虑一个线性映射 $\Phi:\mathbb{R}^3\to\mathbb{R}^4$ 其在标准基下的变换矩阵为

$$\begin{pmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{pmatrix}$$

我们寻找一个新的基下的 Φ 的变换矩阵。令新的基分别为

$$\tilde{B}=(\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix},\begin{pmatrix}1\\0\\1\end{pmatrix}),\tilde{C}=(\begin{pmatrix}1\\1\\0\\0\end{pmatrix},\begin{pmatrix}1\\0\\1\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\\1\end{pmatrix})$$

所以

$$\boldsymbol{S}=\begin{pmatrix}1&0&1\\1&1&0\\0&1&1\end{pmatrix},\ \boldsymbol{T}=\begin{pmatrix}1&1&0&1\\1&0&1&0\\0&1&1&0\\0&0&0&1\end{pmatrix}$$

因此我们可以得到

$$\tilde{\boldsymbol{A}}_\Phi=\boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S}$$

$$=\frac{1}{2}\begin{pmatrix}1&1&-1&-1\\1&-1&1&-1\\-1&1&1&1\\0&0&0&2\end{pmatrix}\begin{pmatrix}1&2&0\\-1&1&3\\3&7&1\\-1&2&4\end{pmatrix}\begin{pmatrix}1&0&1\\1&1&0\\0&1&1\end{pmatrix}=\begin{pmatrix}-4&-4&-2\\6&0&0\\4&8&4\\1&6&3\end{pmatrix}$$

复合线性映射的变换矩阵

考虑向量空间 $\mathbb{V}, \mathbb{W}, \mathbb{X}$，我们知道线性映射的复合仍是线性映射

$$\Phi : \mathbb{V} \to \mathbb{W}$$

$$\Psi : \mathbb{W} \to \mathbb{X}$$

$$\Psi \circ \Phi : \mathbb{V} \to \mathbb{X}$$

记 $\boldsymbol{A}_\Phi, \boldsymbol{A}_\Psi$ 是对应的变换矩阵，则 $\boldsymbol{A}_{\Psi \circ \Phi} = \boldsymbol{A}_\Psi \boldsymbol{A}_\Phi$

例 11

令

$$\Phi:\mathbb{R}^n\to\mathbb{R}^n,\Phi(\boldsymbol{x})=3\boldsymbol{x},\ \Psi:\mathbb{R}^n\to\mathbb{R},\Psi(\boldsymbol{x})=\sum_{i=1}^{n}x_i$$

则

$$\boldsymbol{A}_\Phi=3\boldsymbol{I}$$

$$\boldsymbol{A}_\Psi=\left(1,1,...,1\right)$$

而 $\Psi\circ\Phi(\boldsymbol{x})=\sum_{i=1}^{n}3x_i$

$$\boldsymbol{A}_{\Psi\circ\Phi}=\left(3,3,...,3\right)=\left(1,1,...,1\right)3\boldsymbol{I}=\boldsymbol{A}_\Psi\boldsymbol{A}_\Phi$$

4.2.3 核与像

线性性映射的像与核两个重要的线性子空间。

定义 10

对于 $\Phi : \mathbb{V} \to \mathbb{W}$ 我们定义**核空间** *(零空间)*:

$$ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\boldsymbol{v} \in \mathbb{V} : \Phi(\boldsymbol{v}) = \mathbf{0}_W\}$$

像空间 *(值域)*:

$$Im(\Phi) := \Phi(\mathbb{V}) = \{\boldsymbol{w} \in \mathbb{W} | \exists \boldsymbol{v} \in \mathbb{V} : \Phi(\boldsymbol{v}) = \boldsymbol{w}\}$$

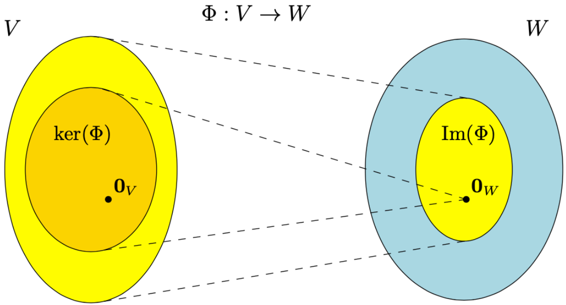

图 6: 核空间与像空间

接下来给出一些关于像空间与核空间的结论。

考虑线性映射 $\Phi : \mathbb{V} \to \mathbb{W}$，其中 $\mathbb{V}, \mathbb{W}$ 是线性空间。

- 总有 $\Phi(\mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{W}}$ ，因此 $\mathbf{0}_{\mathbb{V}} \in ker(\Phi)$
  也就是说零空间永远非空
- $Im(\Phi) \subseteq \mathbb{W}$ 是 $\mathbb{W}$ 的子空间
- $ker(\Phi) \subseteq \mathbb{V}$ 是 $\mathbb{V}$ 的子空间
- Φ 是单射当且仅当 $ker(\Phi) = \mathbf{0}$
- $rank(\boldsymbol{A}) = \dim(Im(\Phi))$
- Φ 的核空间是方程 $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ 的解空间。

定义 11

$\boldsymbol{A}$ 的列向量张成空间叫做**列空间**

- 考虑 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 和线性映射 $\Phi : \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{x} \to \boldsymbol{A}\boldsymbol{x}$
  对于 $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_n]$，我们可以得到

$$Im(\Phi) = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} = \{\sum_{i=1}^{n} x_i \boldsymbol{a}_i\} = L(\boldsymbol{a}_1, ..., \boldsymbol{a}_n) \subseteq \mathbb{R}^m$$

  所以 Φ 的像空间是可以由 $\boldsymbol{A}$ 的列向量张成的。

例 12

考虑映射

$$\Phi : \mathbb{R}^4 \to \mathbb{R}^2, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{pmatrix}$$

是线性映射。

Φ 的像空间就是变换矩阵的列空间。

$$Im(\Phi) = L(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix})$$

为了得到零空间我们需要解 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1/2 \end{pmatrix}$$

最终我们可以给出

$$ker(\Phi) = L(\begin{pmatrix} 0 \\ 1/2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1/2 \\ 0 \\ 1 \end{pmatrix})$$

1 4.1 线性映射

2 4.2 线性映射的矩阵表示

3 4.3 线性变换

4 4.4 仿射映射

定义 12

设 $\mathbb{V}$ 是数域 $\mathbb{K}$ 上的向量空间，如果对任何向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$ 及任意的 $\alpha, \beta \in \mathbb{K}$，有

$$\mathcal{A}(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha \mathcal{A}(\boldsymbol{x}) + \beta \mathcal{A}(\boldsymbol{y}), \tag{2}$$

则称 $\mathcal{A}$ 为 $\mathbb{V}$ 上的**线性变换**，$\mathcal{A}(\boldsymbol{x})$ 和 $\mathcal{A}(\boldsymbol{y})$ 代表元素 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 在变换 $\mathcal{A}$ 下的像。

例 13

下列线性映射是线性变换

- **恒等映射**$\epsilon : \mathbb{V} \to \mathbb{V}, \epsilon(\boldsymbol{x}) = \boldsymbol{x}$。
- 例4中的线性映射 $\boldsymbol{T}_{\boldsymbol{Q}}(\boldsymbol{A}) = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}$，其中 $\boldsymbol{Q}$ 是可逆矩阵，我们称其为矩阵的**相似变换**。

设 $\mathbb{V}$ 是数域 $\mathbb{K}$ 上的 n 维向量空间，$\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n$ 是 $\mathbb{V}$ 的一组基，现在我们来建立线性变换与矩阵之间的关系。

空间 $\mathbb{V}$ 中任一向量 $\boldsymbol{a}$ 可以被基 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n$ 线性表出，即有

$$\boldsymbol{a} = a_1\boldsymbol{\varepsilon}_1 + a_2\boldsymbol{\varepsilon}_2 + \cdots + a_n\boldsymbol{\varepsilon}_n$$

其中系数是唯一确定的，它们就是 $\boldsymbol{a}$ 在这组基下的坐标。由于线性变换保持线性关系不变，因而在 $\boldsymbol{a}$ 的像 $\mathcal{A}\boldsymbol{a}$ 与基的像 $\mathcal{A}\boldsymbol{\varepsilon}_1, \mathcal{A}\boldsymbol{\varepsilon}_2, \cdots, \mathcal{A}\boldsymbol{\varepsilon}_n$ 满足：

$$\begin{aligned}\mathcal{A}\boldsymbol{a} &= \mathcal{A}(a_1\boldsymbol{\varepsilon}_1 + a_2\boldsymbol{\varepsilon}_2 + \cdots + a_n\boldsymbol{\varepsilon}_n) \\ &= a_1\mathcal{A}(\boldsymbol{\varepsilon}_1) + a_2\mathcal{A}(\boldsymbol{\varepsilon}_2) + \cdots + a_n\mathcal{A}(\boldsymbol{\varepsilon}_n).\end{aligned}$$

定理 4

设 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 是向量空间 $\mathbb{V}$ 的一组基，如果线性变换 $\mathcal{A}$ 与 $\mathcal{B}$ 在这组基上的作用相同，即

$$\mathcal{A}\varepsilon_i = \mathcal{B}\varepsilon_i, \quad i = 1, 2, \cdots, n,$$

那么 $\mathcal{A} = \mathcal{B}$.

定理4指出，一个线性变换完全被它在一组基上的作用所决定，然而，基向量的像却完全可以是任意的，也就是说:

定理 5

设 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 是向量空间 $\mathbb{V}$ 的一组基，对于任意一组向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$，一定存在一个线性变换 $\mathcal{A}$ 使得

$$\mathcal{A}\varepsilon_i = \boldsymbol{a}_i, \quad i = 1, 2, \cdots, n.$$

结合以上两点，则有

定理 6

设 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 是向量空间 $\mathbb{V}$ 的一组基，$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 是 $\mathbb{V}$ 中任意 n 个向量，存在唯一的线性变换 $\mathcal{A}$ 使得

$$\mathcal{A}\varepsilon_i = \boldsymbol{a}_i, \quad i = 1, 2, \cdots, n.$$

有了以上的讨论，就可以建立线性变换与矩阵之间的关系。

定义 13

设 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是数域 $\mathbb{K}$ 上 n 维向量空间 $\mathbb{V}$ 的一组基，$\mathcal{A}$ 是 $\mathbb{V}$ 的一个线性变换，基向量的像可以被基线性表出：

$$\begin{cases} \mathcal{A}\varepsilon_1 = a_{11}\varepsilon_1 + a_{21}\varepsilon_2 + \cdots + a_{n1}\varepsilon_n, \\ \qquad\qquad\vdots \\ \mathcal{A}\varepsilon_n = a_{1n}\varepsilon_1 + a_{2n}\varepsilon_2 + \cdots + a_{nn}\varepsilon_n. \end{cases}$$

用矩阵来表示就是

$$\mathcal{A}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n) = (\mathcal{A}\varepsilon_1,\mathcal{A}\varepsilon_2,\cdots,\mathcal{A}\varepsilon_n) \quad = (\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)\boldsymbol{A}$$

矩阵 $\boldsymbol{A}$ 称为 $\mathcal{A}$ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的矩阵。

定理 7

设线性变换 $\mathcal{A}$ 在基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的矩阵是 $\boldsymbol{A}$，且向量 $\boldsymbol{x}$ 在基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的坐标为 $(x_1, x_2, \cdots, x_n)$，则 $\mathcal{A}\boldsymbol{x}$ 在基 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 下的坐标 $(y_1, y_2, \cdots, y_n)$ 可以按照公式

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

计算。

4.3.2 线性变换的几何意义

例 14

考虑线性变换 $\mathcal{A}$ 在数域 $\mathbb{R}^2$ 上的三组矩阵

$$\boldsymbol{A}_1=\begin{pmatrix}\cos(\pi/4) & -\sin(\pi/4)\\ \sin(\pi/4) & \cos(\pi/4)\end{pmatrix}$$

$$\boldsymbol{A}_2=\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}\ \boldsymbol{A}_3=\begin{pmatrix}3/2 & -1/2\\ 1/2 & -1/2\end{pmatrix}$$

它们对原数据的改变如图所示。

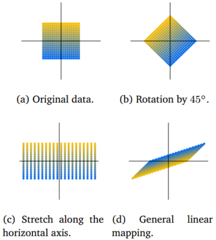

图 7: 线性变换对原数据的影响

1 4.1 线性映射

2 4.2 线性映射的矩阵表示

3 4.3 线性变换

4 4.4 仿射映射

4.4.1 仿射映射

与线性空间之间的映射相似，我们可以定义两个仿射空间的映射。

定义 14

设两个线性空间 $\mathbb{V},\mathbb{W}$ 与一个线性映射 $\Phi:\mathbb{V}\to\mathbb{W}$，则映射 $\phi:\mathbb{V}\to\mathbb{W}$ 且 $\boldsymbol{a}\in\mathbb{W}$

$$\phi(\boldsymbol{x})=\boldsymbol{a}+\Phi(\boldsymbol{x})$$

是一个从 $\mathbb{V}$ 到 $\mathbb{W}$ 的**仿射映射**。向量 $\boldsymbol{a}$ 称为 ϕ 的平移向量。

- 每个仿射映射 $\phi:\mathbb{V}\to\mathbb{W}$ 是一个线性映射 $\Phi:\mathbb{V}\to\mathbb{W}$ 和一个平移变换 $\tau:\mathbb{V}\to\mathbb{W}$ 复合而成的，即 $\phi=\tau\circ\Phi$，其中映射 Φ 和 τ 是唯一确定的。
- 对于仿射映射 $\phi:\mathbb{V}\to\mathbb{W},\phi':\mathbb{W}\to\mathbb{X}$ 的复合 $\phi\circ\phi'$ 也是仿射映射。
- 仿射映射保持几何结构不变。它还保持维度和平行性。

4.4.1 仿射映射

例 15

函数 $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ 就是一个仿射函数。设仿射空间

$$\mathbb{V} = \boldsymbol{c} + L(\boldsymbol{d})$$

则 f 将这个仿射空间的函数映射到了仿射空间

$$\mathbb{W} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{b} + L(\boldsymbol{A}\boldsymbol{d})$$

例 16

在计算机视觉领域的任务中，常利用已有的数据进行翻转、平移或旋转等操作，增加训练集中的数据量，以提高模型的泛化能力。我们以 *MNIST* 数据集中的数字图像为例：
设以原图像的点 $\boldsymbol{s} = (s_1, s_2)$ 为旋转中心。假设经过旋转后，该旋转中心在旋转后的目标图像位置为 $\boldsymbol{d} = (d_1, d_2)$。如何获得逆时针旋转角度 δ 后的图像呢？

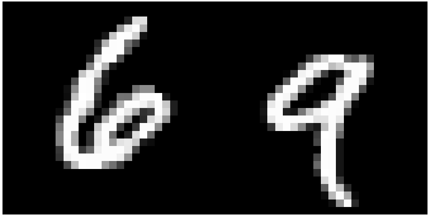

图 8: MNIST 数据集中的数字图像

设一个像素的原位置 $\boldsymbol{x} = (x_1, x_2)$，旋转后的目标位置 $\boldsymbol{y} = (y_1, y_2)$。
它们相对各自旋转中心的相对位置为：

$$(x_1 - s_1, x_2 - s_2), \qquad (y_1 - d_1, y_2 - d_2)$$

相对位置和旋转角度 δ 的关系为：

$$\begin{pmatrix} y_1 - d_1 \\ y_2 - d_2 \end{pmatrix} = \begin{pmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{pmatrix} \begin{pmatrix} x_1 - s_1 \\ x_2 - s_2 \end{pmatrix}$$

则原位置与目标位置关系为：

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{pmatrix} \begin{pmatrix} x_1 - s_1 \\ x_2 - s_2 \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

通过原像素位置计算目标操作位置进行旋转操作的方式称为前向映射。
测试一下我们用前向变换方式得到的逆时针旋转 60° 图像。

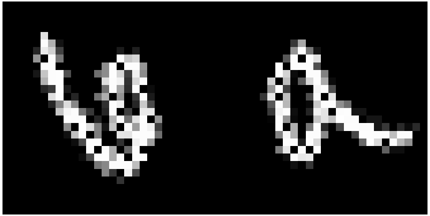

图 9: 利用前向映射进行 MNIST 图像旋转后的结果

观察到旋转后的图像中有一些黑色的噪点。为什么会有噪点？如何去除？

一个像素的原位置 $\boldsymbol{x}=(x_1,x_2)$ 与旋转后的目标位置 $\boldsymbol{y}=(y_1,y_2)$ 的关系为：

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{pmatrix} \begin{pmatrix} x_1 - s_1 \\ x_2 - s_2 \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

利用矩阵的逆，得到变化后的目标像素 $\boldsymbol{y}=(y_1,y_2)$ 的原坐标 $\boldsymbol{x}=(x_1,x_2)$ 为：

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\delta & \sin\delta \\ -\sin\delta & \cos\delta \end{pmatrix} \begin{pmatrix} y_1 - d_1 \\ y_2 - d_2 \end{pmatrix} + \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

因此，只需要对目标图像进行遍历，依次填充原图像的像素，即可得到旋转后的图像。这种方式称为反向映射。

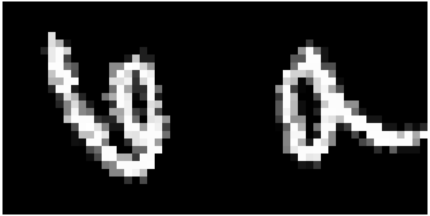

图 10: 利用反向映射进行 MNIST 图像旋转后的结果

在 MNIST 数字分类例子中，即使我们将线性映射改为仿射映射，重新构造模型并使用优化算法求解，最终得到的模型的准确率也几乎不能提升。同样，由于多个仿射函数的复合函数仍然是仿射函数，也不能提高模型的预测能力。

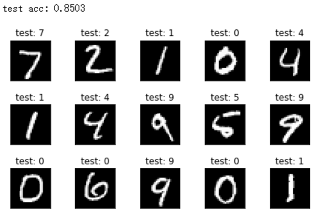

图 11: 仿射映射模型分类准确率

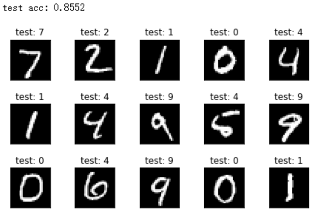

图 12: 仿射映射复合模型分类准确率

线性映射与线性变换小结

本节主要内容

- 线性映射
- 线性变换
- 矩阵表示、基变换
- 仿射映射
- ⋯

线性模型

- 线性映射
- 仿射映射
- 线性或仿射的复合
- ⋯

线性模型的表达能力有限，需要引入非线性映射。除了非线性激活函数，还包括行列式和二次型等以矩阵为变量的矩阵基本特征和各种非线性核函数等。