投影矩阵

unknown

unknown

在讨论正交投影时，我们简要地讨论投影矩阵。特别地，我们讨论下面的定理。

Theorem 1 令 $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ 是 $\mathbb{R}^n$ 的子空间 W 的正交基。形成 $n \times k$ 矩阵

$$\mathbf{U} = \left[\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k \end{array}\right]$$

那么 $\operatorname{proj}_W \mathbf{v} = \mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{v}$。

矩阵 $\mathbf{U}\mathbf{U}^{\mathrm{T}}$ 称为子空间 W 的投影矩阵。它不依赖于正交基的选择。

如果我们不从 W 的正交基开始呢?

Theorem 2 令 $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ 是 $\mathbb{R}^n$ 的子空间 W 的任意基。形成 $n \times k$ 矩阵

$$\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_k \end{array}\right]$$

那么 W 的投影矩阵是 $\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$。

要知道为什么这个公式是真的，我们需要一个引理。

Lemma 3 假设 $\mathbf{A}$ 是 $n \times k$ 矩阵，其列是线性独立的。那么 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 是可逆的。

为了解释这个引理为什么成立，考虑由 $\mathbf{A}$ 决定的变换 $\mathbf{A}: \mathbb{R}^k \to \mathbb{R}^n$。由于 $\mathbf{A}$ 的列是线性独立的，所以这个变换是一对一的。此外，$\mathbf{A}^{\mathrm{T}}$ 的零空间与 $\mathbf{A}$ 的列空间正交，因此，$\mathbf{A}^{\mathrm{T}}$ 在 $\mathbf{A}$ 的列空间上是一对一的，因此，$\mathbf{A}^{\mathrm{T}}\mathbf{A}: \mathbb{R}^k \to \mathbb{R}^k$ 是一对一的。根据可逆矩阵定理，$\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 是可逆的。

现在我们可以计算 $\mathbf{A}$ 列空间的投影矩阵 (注意，$W = \operatorname{Col}\mathbf{A}$)。矩阵 $\mathbf{A}$ 列空间的任意元素都是 $\mathbf{A}$ 列的线性组合，即，

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_k\mathbf{a}_k$$

如果我们令

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

则

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_k\mathbf{a}_k = \mathbf{A}\mathbf{x}$$

给定 $\mathbb{R}^n$ 中的 $\mathbf{v}$，我们用 $\mathbf{x}_p$ 表示对应于 $\mathbf{v}$ 到 W 的投影的 $\mathbf{x}$。换句话说，令

$$\mathrm{proj}_W \mathbf{v} = \mathbf{A}\mathbf{x}_p$$

我们通过计算 $\mathbf{x}_p$ 求投影矩阵。

$\mathbf{v}$ 向 W 的投影的特点是

$$\mathbf{v} - \mathrm{proj}_W \mathbf{v}$$

与 W 中的每个向量 $\mathbf{w}$ 正交，也就是说，

$$\mathbf{w} \cdot (\mathbf{v} - \mathrm{proj}_W \mathbf{v}) = 0$$

对于 W 中的所有 $\mathbf{w}$ 成立。因为 $\mathbf{w} = \mathbf{A}\mathbf{x}$ 对于某些 $\mathbf{x}$，我们有

$$\mathbf{A}\mathbf{x} \cdot (\mathbf{v} - \mathbf{A}\mathbf{x}_p) = 0$$

对于在 $\mathbb{R}^k$ 中所有的 $\mathbf{x}$ 成立。用矩阵的形式写这个点积可以得到

$$(\mathbf{A}\mathbf{x})^{\mathrm{T}} (\mathbf{v} - \mathbf{A}\mathbf{x}_p) = 0$$

相当于

$$\left(\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\right) (\mathbf{v} - \mathbf{A}\mathbf{x}_p) = 0$$

转换回点积，我们有

$$\mathbf{x} \cdot \mathbf{A}^{\mathrm{T}} (\mathbf{v} - \mathbf{A}\mathbf{x}_p) = 0$$

换句话说，向量 $\mathbf{A}^{\mathrm{T}} (\mathbf{v} - \mathbf{A}\mathbf{x}_p)$ 与 $\mathbb{R}^k$ 中的每个向量 $\mathbf{x}$ 正交。$\mathbb{R}^k$ 中唯一具有这个性质的向量是零向量，所以我们可以得出这样的结论

$$\mathbf{A}^{\mathrm{T}} (\mathbf{v} - \mathbf{A}\mathbf{x}_p) = \mathbf{0}$$

我们得到

$$\mathbf{A}^{\mathrm{T}}\mathbf{v} = \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}_p$$

从这个引理，我们知道 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 是可逆的，并且我们有

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{v} = \mathbf{x}_p$$

由于 $\mathbf{A}\mathbf{x}_p$ 是所需的投影，我们有

$$\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{v} = \mathrm{proj}_W \mathbf{v}$$

我们得出结论，W 的投影矩阵是

$$\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}$$

注意，任意投影矩阵 $\mathbf{P}$ 都满足这两个性质

1. $\mathbf{P}^2 = \mathbf{P}$，并且
2. $\mathbf{P}$ 是对称的。

任意满足这两个性质的矩阵都是 $\mathbb{R}^n$ 中的某个子空间的投影矩阵 (see Exercise 36 in Section 7.1 of Lay)。

5.3.27 如果 A 是可对角化的，则对于某些可逆矩阵 P 和对角矩阵 D，$A = PDP^{-1}$。因为 A 是可逆的，则 0 不是 A 的一个特征值。所以 D 的对角项 (就是 A 的特征值) 不为 0，并且 D 是可逆的。根据乘积的逆定理，$A^{-1} = (PDP^{-1})^{-1} = (P^{-1})^{-1} D^{-1}P^{-1}$。因为 D^{-1} 显然是对角的，所以 A^{-1} 是可对角化的。

5.3.28 如果 A 有 n 个线性独立的特征向量，那么根据对角化定理，可逆矩阵 P 和对角矩阵 D，$A = PDP^{-1}$。利用转置的性质，$A^{\mathrm{T}} = (PDP^{-1})^{\mathrm{T}} = \left(P^{\mathrm{T}}\right)^{-1} D^{\mathrm{T}}P^{\mathrm{T}} = QDQ^{-1}$，其中 $Q = \left(P^{\mathrm{T}}\right)^{-1}$。因此 A^{T} 是可对角化的。根据对角化定理，Q 矩阵的列是 A^{T} 的 n 个线性无关的特征向量。

6.3.23 根据正交分解定理，在 $\mathbb{R}^n$ 中的每个 $\mathbf{x}$ 都可以唯一地写成 $\mathbf{x} = \mathbf{p}+\mathbf{u}$，其中 $\mathbf{p}$ 在 Row A 中，$\mathbf{u}$ 在 $(\text{Row } A)^{\perp}$ 中。根据第 6.1 节中的定理 3，$(\text{Row } A)^{\perp} = \text{Nul } A$，因此 $\mathbf{u}$ 在 Nul A 中。接下来，假设 $A\mathbf{x} = \mathbf{b}$ 是一致的。设 $\mathbf{x}$ 为一个解，并如上所述写出 $\mathbf{x} = \mathbf{p}+\mathbf{u}$。那么 $A\mathbf{p} = A(\mathbf{x}-\mathbf{u}) = A\mathbf{x}-A\mathbf{u} = \mathbf{b}-\mathbf{0} = \mathbf{b}$，那么方程 $A\mathbf{x} = \mathbf{b}$ 在 Row A 中至少有一个解 $\mathbf{p}$。最后，假设 $\mathbf{p}$ 和 $\mathbf{p}_1$ 都在 Row A 中，并且都满足 $A\mathbf{x} = \mathbf{b}$。那么 $\mathbf{p}-\mathbf{p}_1$ 在 $\text{Nul } A = (\text{Row } A)^{\perp}$ 中，因为 $A(\mathbf{p}-\mathbf{p}_1) = A\mathbf{p}-A\mathbf{p}_1 = \mathbf{b}-\mathbf{b} = \mathbf{0}$。方程 $\mathbf{p} = \mathbf{p}_1+\mathbf{p}-\mathbf{p}_1$ 和 $\mathbf{p} = \mathbf{p}+\mathbf{0}$ 都将 $\mathbf{p}$ 分解为 Row A 中的向量和 $(\text{Row } A)^{\perp}$ 中的向量之和。通过正交分解的唯一性 (定理 8)，$\mathbf{p} = \mathbf{p}_1$，$\mathbf{p}$ 是唯一的。

6.3.24

a. 根据假设，向量 $\mathbf{w}_1, \ldots, \mathbf{w}_p$ 是成对正交的，并且向量 $\mathbf{v}_1, \ldots, \mathbf{v}_q$ 是成对正交的。因为对于任意 i，$\mathbf{w}_i$ 在 W 中，对于任意 j，$\mathbf{v}_j$ 在 $W^{\perp}$ 中，对于任意 i 和 j，$\mathbf{w}_i \cdot \mathbf{v}_j = 0$ 表示。因此 $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ 形成一个正交集。

b. 对于 $\mathbb{R}^n$ 中的任意 $\mathbf{y}$，如正交分解定理所示，写为 $\mathbf{y} = \hat{\mathbf{y}}+\mathbf{z}$，其中 $\hat{\mathbf{y}}$ 在 W 中，$\mathbf{z}$ 在 $W^{\perp}$ 中. 然后存在标量 $c_1, \ldots, c_p$ 和 $d_1, \ldots, c_q$，使得 $\mathbf{y} = \hat{\mathbf{y}}+\mathbf{z} = c_1\mathbf{w}_1 + \ldots + c_p\mathbf{w}_p + d_1\mathbf{v}_1 + \ldots + d_q\mathbf{v}_q$。因此，集合 $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ 涵盖 $\mathbb{R}^n$。

c. 集合 $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ 与 (a) 线性无关，并且 (b) 涵盖 $\mathbb{R}^n$，因此是 $\mathbb{R}^n$ 的基向量。因此 $\text{dim}W + \text{dim}W^{\perp} = p+q = \text{dim}\mathbb{R}^n$。

7.1.27 令 A 为 $n \times n$ 对称矩阵。则 $(A\mathbf{x})\cdot\mathbf{y} = (A\mathbf{x})^{\mathrm{T}}\mathbf{y} = \mathbf{x}^{\mathrm{T}}A^{\mathrm{T}}\mathbf{y} = \mathbf{x}^{\mathrm{T}}A\mathbf{y} = \mathbf{x}\cdot(A\mathbf{y})$，因为 $A^{\mathrm{T}} = A$。

7.1.28 因为 A 是对称的，$\left(B^{\mathrm{T}}AB\right) = B^{\mathrm{T}}A^{\mathrm{T}}\left(B^{\mathrm{T}}\right)^{\mathrm{T}} = B^{\mathrm{T}}AB$，并且 $B^{\mathrm{T}}AB$ 是对称的。将此结果应用于 $A = I$，则 $B^{\mathrm{T}}B$ 是对称的。最后，$\left(BB^{\mathrm{T}}\right)^{\mathrm{T}} = \left(B^{\mathrm{T}}\right)^{\mathrm{T}} B^{\mathrm{T}}$ ，所以 BB^{T} 是对称的。

7.1.29 由于 A 是正交可对角化的，$A = PDP^{-1}$，其中 P 是正交的，D 是对角的。因为 A 是可逆的，所以 $A^{-1} = (PDP^{-1})^{-1} = (P^{-1})^{-1} D^{-1}P^{-1}$。注意，$D^{-1}$ 是对角矩阵，所以 A^{-1} 是正交可对角化的。

7.1.30 如果 A 和 B 是正交可对角化的，则 A 和 B 根据定理 2 说明是对称的。如果 $AB=BA$，则 $(AB)^{\mathrm{T}}=(BA)^{\mathrm{T}}=A^{\mathrm{T}}B^{\mathrm{T}}=AB$。所以 AB 是对称的，并因此可以通过定理 2 说明是正交可对角化的。

7.1.31 第 5.3 节的对角化定理指出，P 的列向量是线性独立的特征向量，对应于 D 的对角线上列出的 A 的特征值。因此 P 有 k 列与 λ 相对应的特征向量。这些 k 列向量构成特征空间的基向量。

7.1.32 如果 $A=PRP^{-1}$,则 $P^{-1}AP=R$。因为 P 是正交的,$R=P^{\mathrm{T}}AP$。因此 $R^{\mathrm{T}}=\left(P^{\mathrm{T}}AP\right)^{\mathrm{T}}=P^{\mathrm{T}}A^{\mathrm{T}}\left(P^{\mathrm{T}}\right)^{\mathrm{T}}=P^{\mathrm{T}}AP=R$，其中显示 R 是对称的。因为 R 也是上三角形矩阵，所以它在对角线上方的条目必须是零，以匹配对角线下方的零。因此 R 是对角矩阵。

7.1.35

a. 给定 $\mathbb{R}^n$ 中的 $\mathbf{x}$，$b\mathbf{x}=\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)\mathbf{x}=\mathbf{u}\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}\right)=\left(\mathbf{u}^{\mathrm{T}}\mathbf{x}\right)\mathbf{u}$，因为 $\mathbf{u^T x}$ 是一个标量。所以 $B\mathbf{x}=(\mathbf{x}\cdot\mathbf{u})\,\mathbf{u}$。因为 $\mathbf{u}$ 是单位向量，所以 $B\mathbf{x}$ 是 $\mathbf{x}$ 在 $\mathbf{u}$ 上的正交投影。

b. 因为 $B^{\mathrm{T}}=\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)^{\mathrm{T}}=\left(\mathbf{u}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{u}^{\mathrm{T}}=\mathbf{u}\mathbf{u}^{\mathrm{T}}=B$，$B$ 是对称矩阵。所以 $B^2=\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)=\mathbf{u}\left(\mathbf{u}^{\mathrm{T}}\mathbf{u}\right)\mathbf{u}^{\mathrm{T}}=\mathbf{u}\mathbf{u}^{\mathrm{T}}=B$，因为 $\mathbf{u}^{\mathrm{T}}\mathbf{u}=1$。

c. 因为 $\mathbf{u}^{\mathrm{T}}\mathbf{u}=1$，$B\mathbf{x}=\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right)\mathbf{u}=\mathbf{u}\left(\mathbf{u}^{\mathrm{T}}\mathbf{u}\right)=\mathbf{u}\,(1)=\mathbf{u}$，所以 $\mathbf{u}$ 是 B 的特征向量，对应的特征值为 1。

7.1.36 给定 $\mathbb{R}^n$ 中的任意 $\mathbf{y}$，设 $\hat{\mathbf{y}}=B\mathbf{y}$，并且 $\mathbf{z}=\mathbf{y}$-$\hat{\mathbf{y}}$。假设 $B^{\mathrm{T}}=B$ 和 $B^2=B$，则 $B^{\mathrm{T}}B=BB=B$。

a. 因为 $\mathbf{z}\cdot\hat{\mathbf{y}}=(\mathbf{y}-\hat{\mathbf{y}})\cdot(B\mathbf{y})=\mathbf{y}\cdot(B\mathbf{y})-\hat{\mathbf{y}}\cdot(B\mathbf{y})=\mathbf{y}^{\mathrm{T}}B\mathbf{y}-(B\mathbf{y})^{\mathrm{T}}B\mathbf{y}=\mathbf{y}^{\mathrm{T}}B\mathbf{y}-\mathbf{y}^{\mathrm{T}}B^{\mathrm{T}}B\mathbf{y}=\mathbf{0}$，所以 $\mathbf{z}$ 正交于 $\hat{\mathbf{y}}$。

b. 在 $W=\mathrm{Col}B$ 中的任意向量，对于某些 $\mathbf{u}$，有形式为 $B\mathbf{u}$。注意到 B 是对称的，练习 28 给出 $(\mathbf{y}-\hat{\mathbf{y}})\cdot(B\mathbf{u})=[B\,(\mathbf{y}-\hat{\mathbf{y}})]\cdot\mathbf{u}=[B\mathbf{y}-BB\mathbf{y}]\cdot\mathbf{u}=\mathbf{0}$，因为 $B^2=B$，所以 $\mathbf{y}-\hat{\mathbf{y}}$ 在 $W^{\perp}$ 中，并分解为 $\mathbf{y}=\hat{\mathbf{y}}+(\mathbf{y}-\hat{\mathbf{y}})$，将 $\mathbf{y}$ 表示为 W 中的向量和 $W^{\perp}$ 中的向量之和。根据第 6.3 节中的正交分解定理，该分解是唯一的，因此 $\hat{\mathbf{y}}$ 必须是 $\mathrm{proj}_W\mathbf{y}$。