3.7 ~

3.7 Inner Product of Functions

우리가 이전까지 다뤘던 inner product는 유한한 entires의 벡터들에 대한 것이었다.

이제 inner product를 무한한 entries의 벡터들(countably infinite)와 continuous-valued functions (uncountably infinite)로 좀더 generalize 해보자

따라서 아래의 식과 같이 벡터의 각 구성 요소들을 합친 것은 3.37의 식 처럼 integral로 변한다.

$$\boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \,.$$

An inner product of two functions $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ can be defined as the definite integral

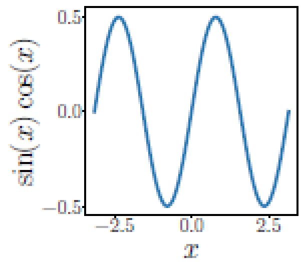
$$\langle u, v \rangle := \int_a^b u(x)v(x)dx \qquad (3.37)$$

- 만약 (3.37)의 결과가 0이 된다면, 함수 u와 v는 수직(orthogonal)하다

예 3.9)

Example 3.9 (Inner Product of Functions)
If we choose $u = \sin(x)$ and $v = \cos(x)$, the integrand $f(x) = u(x)v(x)$ of (3.37), is shown in Figure 3.8. We see that this function is odd, i.e., $f(-x) = -f(x)$. Therefore, the integral with limits $a = -\pi, b = \pi$ of this product evaluates to 0. Therefore, sin and cos are orthogonal functions.

Figure 3.8 $f(x) = \sin(x)\cos(x)$.



이외에도 [-π, π) 범위 안에서 periodic한 함수들의 집합은 모두 서로 orthogonal 하다.

3.8 Orthogonal Projection

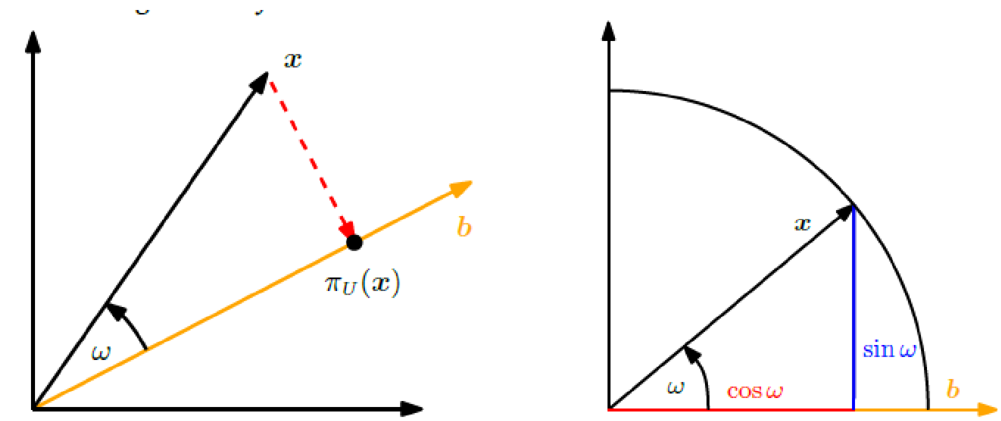
- for data compresion
- 머신러닝을 하다보면 고차원의 데이터를 다룰때가 있다.
- 그러나 고차원의 데이터들은 그중 일부 차원만이 대부분의 정보를 갖고있는 특성이 있다.
- 따라서 우리는 original high-dimensional data를 lower dimensional feature space에 projection함으로써 data compression을 하고자 한다.

Definition 3.10 (Projection). Let V be a vector space and $U \subseteq V$ a subspace of V. A linear mapping $\pi : V \to U$ is called a *projection* if $\pi^2 = \pi \circ \pi = \pi$.

- linear mapping이 tranfermation matrice로 표현될 수 있듯이, projection matrice P π 또한 projection을 표현한다.

$$\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi .$$

3.8.1 Projection onto One-Dimensional Subspaces (lines)



(a) Projection of $x \in \mathbb{R}^2$ onto a subspace U with basis vector b.

(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b.

Figure 3.10 Examples of projections onto one-dimensional subspaces.

Assume we are given a line (one-dimensional subspace) through the origin with basis vector $b \in \mathbb{R}^n$. The line is a one-dimensional subspace $U \subseteq \mathbb{R}^n$ spanned by b. When we project $x \in \mathbb{R}^n$ onto U, we seek the vector $\pi_U(x) \in U$ that is closest to x. Using geometric arguments, let

us characterize some properties of the projection $\pi_U(x)$ (Figure 3.10(a) serves as an illustration):

<projection의 특성>

- The projection $\pi_U(x)$ is closest to x, where "closest" implies that the distance $\|x - \pi_U(x)\|$ is minimal. It follows that the segment $\pi_U(x) - x$ from $\pi_U(x)$ to x is orthogonal to U, and therefore the basis vector b of U. The orthogonality condition yields $\langle \pi_U(x) - x, b \rangle = 0$ since angles between vectors are defined via the inner product.
- The projection $\pi_U(x)$ of x onto U must be an element of U and, therefore, a multiple of the basis vector b that spans U. Hence, $\pi_U(x) = \lambda b$, for some $\lambda \in \mathbb{R}$.

λ is then the coordinate of $\pi_U(x)$ with respect to b.

다음 3단계를 통해 coodinate 람다, projection π(x), projection matrix Pπ를 구해보자

1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x})=\lambda \boldsymbol{b}}{\iff} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \,. \tag{3.39}$$

We can now exploit the bilinearity of the inner product and arrive at

With a general inner product, we get $\lambda = \langle \boldsymbol{x}, \boldsymbol{b} \rangle$ if $\|\boldsymbol{b}\| = 1$.

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \iff \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle} = \frac{\langle \boldsymbol{b}, \boldsymbol{x} \rangle}{\|\boldsymbol{b}\|^2} \,. \tag{3.40}$$

In the last step, we exploited the fact that inner products are symmetric. If we choose $\langle \cdot, \cdot \rangle$ to be the dot product, we obtain

$$\lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \,. \tag{3.41}$$

If $\|\boldsymbol{b}\| = 1$, then the coordinate λ of the projection is given by $\boldsymbol{b}^\top \boldsymbol{x}$.

2. Finding the projection point $\pi_U(\boldsymbol{x}) \in U$. Since $\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}$, we immediately obtain with (3.40) that

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\|\boldsymbol{b}\|^2} \boldsymbol{b} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \boldsymbol{b} \,, \tag{3.42}$$

where the last equality holds for the dot product only. We can also compute the length of $\pi_U(\boldsymbol{x})$ by means of Definition 3.1 as

$$\|\pi_U(\boldsymbol{x})\| = \|\lambda \boldsymbol{b}\| = |\lambda| \, \|\boldsymbol{b}\| \,. \tag{3.43}$$

Hence, our projection is of length $|\lambda|$ times the length of $\boldsymbol{b}$. This also adds the intuition that λ is the coordinate of $\pi_U(\boldsymbol{x})$ with respect to the basis vector $\boldsymbol{b}$ that spans our one-dimensional subspace U.

If we use the dot product as an inner product, we get

$$\|\pi_U(\boldsymbol{x})\| \overset{(3.42)}{=} \frac{|\boldsymbol{b}^\top \boldsymbol{x}|}{\|\boldsymbol{b}\|^2} \|\boldsymbol{b}\| \overset{(3.25)}{=} |\cos\omega| \, \|\boldsymbol{x}\| \, \|\boldsymbol{b}\| \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}\|^2} = |\cos\omega| \, \|\boldsymbol{x}\| \,. \tag{3.44}$$

Here, ω is the angle between $\boldsymbol{x}$ and $\boldsymbol{b}$. This equation should be familiar from trigonometry: If $\|\boldsymbol{x}\| = 1$, then $\boldsymbol{x}$ lies on the unit circle. It follows that the projection onto the horizontal axis spanned by $\boldsymbol{b}$ is exactly $\cos\omega$, and the length of the corresponding vector $\pi_U(\boldsymbol{x}) = |\cos\omega|$. An illustration is given in Figure 3.10(b).

$$\cos\omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}\,. \tag{3.25}$$

3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping (see Definition 3.10). Therefore, there exists a projection matrix $\boldsymbol{P}_\pi$, such that $\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$. With the dot product as inner product and

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\lambda = \boldsymbol{b}\frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\boldsymbol{x}\,, \tag{3.45}$$

we immediately see that

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\,. \tag{3.46}$$

Note that $\boldsymbol{b}\boldsymbol{b}^\top$ (and, consequently, $\boldsymbol{P}_\pi$) is a symmetric matrix (of rank 1), and $\|\boldsymbol{b}\|^2 = \langle \boldsymbol{b}, \boldsymbol{b}\rangle$ is a scalar.

The projection matrix $\boldsymbol{P}_\pi$ projects any vector $\boldsymbol{x} \in \mathbb{R}^n$ onto the line through the origin with direction $\boldsymbol{b}$ (equivalently, the subspace U spanned by $\boldsymbol{b}$).

Remark. The projection $\pi_U(\boldsymbol{x}) \in \mathbb{R}^n$ is still an n-dimensional vector and not a scalar. However, we no longer require n coordinates to represent the projection, but only a single one if we want to express it with respect to the basis vector $\boldsymbol{b}$ that spans the subspace U: λ. ◇

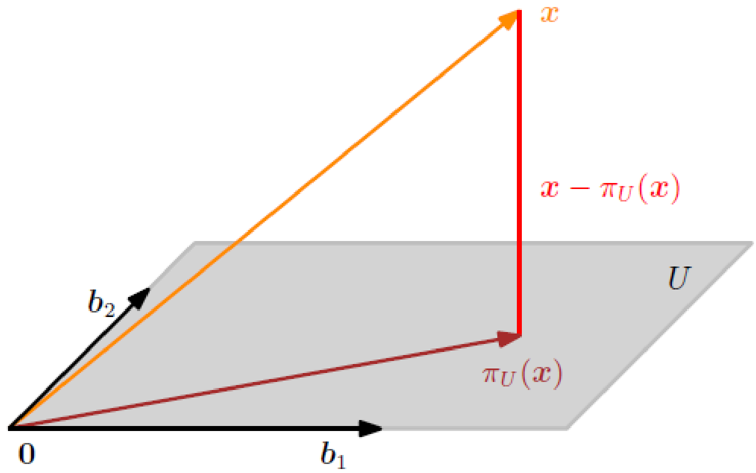


Figure 3.11 Projection onto a two-dimensional subspace U with basis b_1, b_2. The projection $\pi_U(x)$ of $x \in \mathbb{R}^3$ onto U can be expressed as a linear combination of b_1, b_2 and the displacement vector $x - \pi_U(x)$ is orthogonal to both b_1 and b_2.

Example 3.10 (Projection onto a Line)
Find the projection matrix P_π onto the line through the origin spanned by $b = \begin{bmatrix}1 & 2 & 2\end{bmatrix}^\top$. b is a direction and a basis of the one-dimensional subspace (line through origin).

With (3.46), we obtain

$$P_\pi = \frac{bb^\top}{b^\top b} = \frac{1}{9}\begin{bmatrix}1\\2\\2\end{bmatrix}\begin{bmatrix}1 & 2 & 2\end{bmatrix} = \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}. \tag{3.47}$$

Let us now choose a particular x and see whether it lies in the subspace spanned by b. For $x = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\top$, the projection is

$$\pi_U(x) = P_\pi x = \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \frac{1}{9}\begin{bmatrix}5\\10\\10\end{bmatrix} \in \operatorname{span}\left[\begin{bmatrix}1\\2\\2\end{bmatrix}\right]. \tag{3.48}$$

Note that the application of P_π to $\pi_U(x)$ does not change anything, i.e., $P_\pi \pi_U(x) = \pi_U(x)$. This is expected because according to Definition 3.10, we know that a projection matrix P_π satisfies $P_\pi^2 x = P_\pi x$ for all x.

3.8.2 Projection onto General Subspaces

Assume that $(b_1, \ldots, b_m)$ is an ordered basis of U. Any projection $\pi_U(x)$ onto U is necessarily an element of U. Therefore, they can be represented

The basis vectors form the columns of $B \in \mathbb{R}^{n \times m}$, where $B = [b_1, \ldots, b_m]$.

as linear combinations of the basis vectors $b_1, \ldots, b_m$ of U, such that $\pi_U(x) = \sum_{i=1}^{m} \lambda_i b_i$.

As in the 1D case, we follow a three-step procedure to find the projection $\pi_U(x)$ and the projection matrix P_π:

<Projection π(x)와 projection matrix Pπ를 구하기 위한 3가지 절차>

1. Find the coordinates $\lambda_1, \ldots, \lambda_m$ of the projection (with respect to the basis of U), such that the linear combination

$$\pi_U(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i \boldsymbol{b}_i = \boldsymbol{B}\boldsymbol{\lambda}, \tag{3.49}$$

$$\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m] \in \mathbb{R}^{n \times m}, \quad \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m, \tag{3.50}$$

is closest to $\boldsymbol{x} \in \mathbb{R}^n$. As in the 1D case, "closest" means "minimum distance", which implies that the vector connecting $\pi_U(\boldsymbol{x}) \in U$ and $\boldsymbol{x} \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain m simultaneous conditions (assuming the dot product as the inner product)

$$\langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle = \boldsymbol{b}_1^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \tag{3.51}$$

$$\vdots$$

$$\langle \boldsymbol{b}_m, \boldsymbol{x} - \pi_U(\boldsymbol{x}) \rangle = \boldsymbol{b}_m^\top (\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \tag{3.52}$$

which, with $\pi_U(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\lambda}$, can be written as

$$\boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{3.53}$$

$$\vdots$$

$$\boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{3.54}$$

such that we obtain a homogeneous linear equation system

$$\begin{bmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_m^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda} \end{bmatrix} = \boldsymbol{0} \iff \boldsymbol{B}^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = \boldsymbol{0} \tag{3.55}$$

$$\iff \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{B}^\top \boldsymbol{x}. \tag{3.56}$$

The last expression is called *normal equation*. Since $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ are a basis of U and, therefore, linearly independent, $\boldsymbol{B}^\top \boldsymbol{B} \in \mathbb{R}^{m \times m}$ is regular and can be inverted. This allows us to solve for the coefficients/ coordinates

$$\boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}. \tag{3.57}$$

The matrix $(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ is also called the *pseudo-inverse* of $\boldsymbol{B}$, which can be computed for non-square matrices $\boldsymbol{B}$. It only requires that $\boldsymbol{B}^\top \boldsymbol{B}$ is positive definite, which is the case if $\boldsymbol{B}$ is full rank. In practical applications (e.g., linear regression), we often add a "jitter term" $\epsilon \boldsymbol{I}$ to

2. Find the projection $\pi_U(\boldsymbol{x}) \in U$. We already established that $\pi_U(\boldsymbol{x}) = \boldsymbol{B\lambda}$. Therefore, with (3.57)

$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{x}\,. \tag{3.58}$$

3. Find the projection matrix $\boldsymbol{P}_\pi$. From (3.58), we can immediately see that the projection matrix that solves $\boldsymbol{P}_\pi \boldsymbol{x} = \pi_U(\boldsymbol{x})$ must be

$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top\,. \tag{3.59}$$

Remark. The solution for projecting onto general subspaces includes the 1D case as a special case: If $\dim(U) = 1$, then $\boldsymbol{B}^\top \boldsymbol{B} \in \mathbb{R}$ is a scalar and we can rewrite the projection matrix in (3.59) $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top$ as $\boldsymbol{P}_\pi = \frac{\boldsymbol{B}\boldsymbol{B}^\top}{\boldsymbol{B}^\top \boldsymbol{B}}$, which is exactly the projection matrix in (3.46). $\diamondsuit$

3.8.3 Gram-Schmidt Orthogonalization

- Projection은 Gram-Schmidt Orthogonalization의 핵심이 된다.
- Gram-Schmidt Orthogonalization : any basis (b1, ..., bn) of V 로부터 orthogonal basis (u1, ..., un)을 구성하는 것

$$\boldsymbol{u}_1 := \boldsymbol{b}_1 \tag{3.67}$$

$$\boldsymbol{u}_k := \boldsymbol{b}_k - \pi_{\mathrm{span}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{k-1}]}(\boldsymbol{b}_k)\,, \quad k = 2,\ldots,n\,. \tag{3.68}$$

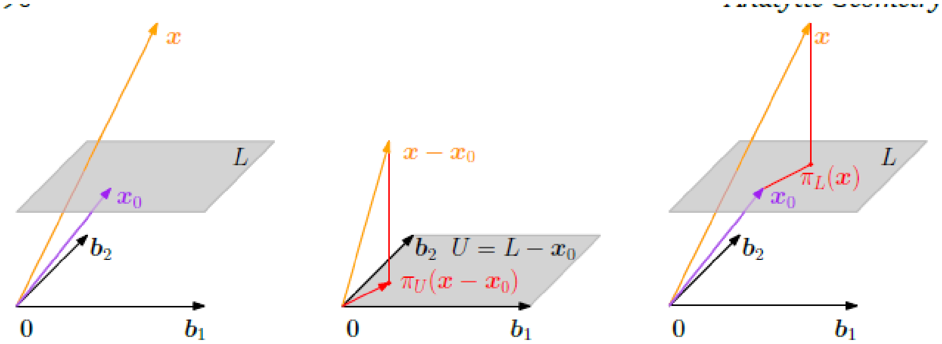
- 이때 bk는 (orthogonal vectors u1,...uk-1에 의해 span 되는 subspace에) project 된다.

tion 3.8.2. This projection is then subtracted from $\boldsymbol{b}_k$ and yields a vector $\boldsymbol{u}_k$ that is orthogonal to the $(k-1)$-dimensional subspace spanned by $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{k-1}$. Repeating this procedure for all n basis vectors $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_n$ yields an orthogonal basis $(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$ of V. If we normalize the $\boldsymbol{u}_k$, we obtain an ONB where $\|\boldsymbol{u}_k\| = 1$ for $k = 1,\ldots,n$.

3.8.4 Projection onto Affine Subspace

- 지금까지 더 낮은 차원 subspace 로 벡터를 projection하는 것에 대해 배웠다.
- 이제 Affine subspace 에 Projection하는 것에 대해 알아보자

Figure 3.13 Projection onto an affine space. (a) original setting; (b) setting shifted by $-\boldsymbol{x}_0$ so that $\boldsymbol{x} - \boldsymbol{x}_0$ can be projected onto the direction space U; (c) projection is translated back to $\boldsymbol{x}_0 + \pi_U(\boldsymbol{x} - \boldsymbol{x}_0)$, which gives the final orthogonal projection $\pi_L(\boldsymbol{x})$.



(a) Setting.

(b) Reduce problem to projection π_U onto vector subspace.

(c) Add support point back in to get affine projection π_L.

Consider the setting in Figure 3.13(a). We are given an affine space $L = \boldsymbol{x}_0 + U$, where $\boldsymbol{b}_1, \boldsymbol{b}_2$ are basis vectors of U. To determine the orthogonal projection $\pi_L(\boldsymbol{x})$ of $\boldsymbol{x}$ onto L, we transform the problem into a problem that we know how to solve: the projection onto a vector subspace. In order to get there, we subtract the support point $\boldsymbol{x}_0$ from $\boldsymbol{x}$ and from L, so that $L - \boldsymbol{x}_0 = U$ is exactly the vector subspace U. We can now use the orthogonal projections onto a subspace we discussed in Section 3.8.2 and obtain the projection $\pi_U(\boldsymbol{x} - \boldsymbol{x}_0)$, which is illustrated in Figure 3.13(b). This projection can now be translated back into L by adding $\boldsymbol{x}_0$, such that we obtain the orthogonal projection onto an affine space L as

$$\pi_L(\boldsymbol{x}) = \boldsymbol{x}_0 + \pi_U(\boldsymbol{x} - \boldsymbol{x}_0)\,, \tag{3.72}$$

where $\pi_U(\cdot)$ is the orthogonal projection onto the subspace U, i.e., the direction space of L; see Figure 3.13(c).

From Figure 3.13, it is also evident that the distance of $\boldsymbol{x}$ from the affine space L is identical to the distance of $\boldsymbol{x} - \boldsymbol{x}_0$ from U, i.e.,

$$d(\boldsymbol{x}, L) = \|\boldsymbol{x} - \pi_L(\boldsymbol{x})\| = \|\boldsymbol{x} - (\boldsymbol{x}_0 + \pi_U(\boldsymbol{x} - \boldsymbol{x}_0))\| \tag{3.73a}$$

$$= d(\boldsymbol{x} - \boldsymbol{x}_0, \pi_U(\boldsymbol{x} - \boldsymbol{x}_0)) = d(\boldsymbol{x} - \boldsymbol{x}_0, U)\,. \tag{3.73b}$$

We will use projections onto an affine subspace to derive the concept of a separating hyperplane in Section 12.1.