Analytic Geometry

April 6, 2023

1 Orthogonal Projections

Projections are an important class of linear transformations and play an important role in graphics, coding theory, statistics and machine learning.

In machine learning, dealing with high dimensional data can be very problematic, because it is hard to interpret and visualize. Sometimes only a small fraction of the features contains the most information. We can use projections to transform the high dimensional data into lower dimensional data while preserving as much information as possible (Principal Component Analysis).

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

Since linear mappings can be expressed by transformation matrices, the preceding definition applies equally to a special kind of transformation matrices, the *projection matrices* $\boldsymbol{P}_\pi$, which exhibit the property that $\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi$.

2 Projection onto One-Dimensional Subspaces (Lines)

Assume we have a basis vector $\boldsymbol{b}$ that spans a line (one dimension) and defines the subspace $U \subseteq \mathbb{R}^n$. We are interested in the projection $\pi_U(\boldsymbol{x}) \in U$, that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto the subspace U. The projection $\pi_U(\boldsymbol{x})$ can be characterized as follows:

- The projection $\pi_U(\boldsymbol{x})$ is closest to $\boldsymbol{x}$, where "closest" implies that the distance $\|\boldsymbol{x} - \pi_U(\boldsymbol{x})\|$ is minimal. It follows that the segment $\pi_U(\boldsymbol{x}) - \boldsymbol{x}$ from $\pi_U(\boldsymbol{x})$ to $\boldsymbol{x}$ is orthogonal to U, and therefore the basis vector $\boldsymbol{b}$ of U. The orthogonality condition yields $\langle \pi_U(\boldsymbol{x}) - \boldsymbol{x}, \boldsymbol{b} \rangle = 0$ since angles between vectors are defined via the inner product.

- The projection $\pi_U(\boldsymbol{x})$ of $\boldsymbol{x}$ onto U must be an element of U and, therefore, a multiple of the basis vector $\boldsymbol{b}$ that spans U. Hence, $\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}$, for some $\lambda \in \mathbb{R}$.

In the following three steps, we determine the coordinate λ, the projection $\pi_U(\boldsymbol{x}) \in U$, and the projection matrix $\boldsymbol{P}_\pi$ that maps any $\boldsymbol{x} \in \mathbb{R}^n$ onto U:

1. Finding the coordinate λ. The orthogonality condition yields

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \overset{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}}{\iff} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \tag{1}$$

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

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

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

$$\lambda = \frac{\boldsymbol{b}^T \boldsymbol{x}}{\boldsymbol{b}^T \boldsymbol{b}} = \frac{\boldsymbol{b}^T \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \tag{3}$$

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

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

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\|\boldsymbol{b}\|^2} \boldsymbol{b} = \frac{\boldsymbol{b}^T \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \boldsymbol{b} \tag{4}$$

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

$$\|\pi_U(\boldsymbol{x})\| = \|\lambda \boldsymbol{b}\| = |\lambda| \|\boldsymbol{b}\| \tag{5}$$

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. I we use the dot product as an inner product, we get

$$\|\pi_U(\boldsymbol{x})\| = \frac{|\boldsymbol{b}^T \boldsymbol{x}|}{\|\boldsymbol{b}\|^2} \|\boldsymbol{b}\| = |\cos \omega| \|\boldsymbol{x}\| \|\boldsymbol{b}\| \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}\|^2} = |\cos \omega| \|\boldsymbol{x}\| \tag{6}$$

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 $\pi_U(\boldsymbol{x}) = |\cos \omega|$.

3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping. 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

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b} \lambda = \boldsymbol{b} \frac{\boldsymbol{b}^T \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b} \boldsymbol{b}^T}{\|\boldsymbol{b}\|^2} \boldsymbol{x} \tag{7}$$

we immediatley see that

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b} \boldsymbol{b}^T}{\|\boldsymbol{b}\|^2} \tag{8}$$

Note that $\boldsymbol{b} \boldsymbol{b}^T$ (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.

3 Projection onto General Subspaces

After covering the case for projections onto one dimensional subspaces, now we will look at projections on general lower dimensional subspaces $U \subseteq \mathbb{R}^n$ with $dim(U) = m \geqslant 1$.

Assume that $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ is an ordered basis of U. Any projection $\pi_U(\boldsymbol{x})$ onto U is necessarily an element of U. Therefore, they can be represented as linear combinations of the basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ of U, such that $\pi_U(\boldsymbol{x}) = \sum_{i=1}^m \lambda_i \boldsymbol{b}_i$.
As in the 1D case, we follow a three-step procedure to find the projection $\pi_U(\boldsymbol{x})$ and the projection matrix $\boldsymbol{P}_\pi$:

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

$$\pi_U(\boldsymbol{x}) = \sum_{i=1}^m \lambda_i \boldsymbol{b}_i = \boldsymbol{B} \boldsymbol{\lambda} \tag{9}$$

$$\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m] \in \mathbb{R}^{n \times m}, \quad \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^T \in \mathbb{R}^m \tag{10}$$

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)

$$\langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_U(\boldsymbol{x})\rangle = \boldsymbol{b}_1^T(\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \tag{11}$$

$$\vdots$$

$$\langle \boldsymbol{b}_m, \boldsymbol{x} - \pi_U(\boldsymbol{x})\rangle = \boldsymbol{b}_m^T(\boldsymbol{x} - \pi_U(\boldsymbol{x})) = 0 \tag{12}$$

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

$$\boldsymbol{b}_1^T(\boldsymbol{x} - \boldsymbol{B\lambda}) = 0 \tag{13}$$

$$\vdots$$

$$\boldsymbol{b}_m^T(\boldsymbol{x} - \boldsymbol{B\lambda}) = 0 \tag{14}$$

such that we obtain a homogeneous linear equation system

$$\begin{aligned} \begin{bmatrix} \boldsymbol{b}_1^T \\ \vdots \\ \boldsymbol{b}_m^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{B\lambda} \end{bmatrix} = \boldsymbol{0} &\iff \boldsymbol{B}^T(\boldsymbol{x} - \boldsymbol{B\lambda}) = \boldsymbol{0} \\ &\iff \boldsymbol{B}^T\boldsymbol{B\lambda} = \boldsymbol{B}^T\boldsymbol{x} \end{aligned} \tag{15}$$

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

$$\boldsymbol{\lambda} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{x} \tag{16}$$

The matrix $(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T$ is also called the *pseudo-inverse* of $\boldsymbol{B}$, which can be computed for non-square matrices $\boldsymbol{B}$. It only requires that $\boldsymbol{B}^T\boldsymbol{B}$ is positive definite, which is the case if $\boldsymbol{B}$ is full rank.

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

$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{x} \tag{17}$$

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

$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T \tag{18}$$

Projections allow us to look at situations where we have a linear system $\boldsymbol{Ax} = \boldsymbol{b}$ without a solution. In this case, we can find an *approximate solution.* The idea is to find the vector in the subspace spanned by the columns of $\boldsymbol{A}$ that is closest to $\boldsymbol{b}$, i.e., we compute the orthogonal projection of $\boldsymbol{b}$ onto the subspace spanned by the columns of $\boldsymbol{A}$. This problems occurs often in practice and the solution is called *least-squares solution* of an overdetermined system.

4 Gram-Schmidt Orthogonalization

This methods of orthogonalization allows us to transform any basis of an n-dimensional vector space V into an orthogonal/orthonormal basis of V. It iteratively constructs an orthogonal basis $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$ from any basis $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ of V as follows:

$$\begin{aligned} &\boldsymbol{u}_1 := \boldsymbol{b}_1 \\ &\boldsymbol{u}_n := \boldsymbol{b}_k - \pi_{span[\boldsymbol{u}_1,\dots,\boldsymbol{u}_{k-1}]}(\boldsymbol{b}_k), \quad k = 2, \dots, n \end{aligned} \tag{19}$$

The k-th basis vector $\boldsymbol{b}_k$ is projected onto the subspace spanned by the first $k-1$ constructed orthogonal vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}$. 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 $k = 1, \ldots, n$.

5 Projection onto Affine Subspaces

This is similar to examples we have seen before. 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 and obtain the projection $\pi_U(\boldsymbol{x} - \boldsymbol{x}_0)$. 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

$$\pi_L(\boldsymbol{x}) = \boldsymbol{x}_0 + \pi_U(\boldsymbol{x} - \boldsymbol{x}_0) \tag{20}$$

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

$$\begin{aligned} d(\boldsymbol{x}, L) &= \|\boldsymbol{x} - \pi_L(\boldsymbol{x})\| = \|\boldsymbol{x} - (\boldsymbol{x}_0 + \pi_U(\boldsymbol{x} - \boldsymbol{x}_0))\| \\ &= d(\boldsymbol{x} - \boldsymbol{x}_0, \pi_U(\boldsymbol{x} - \boldsymbol{x}_0)) = d(\boldsymbol{x} - \boldsymbol{x}_0, U) \end{aligned} \tag{21}$$

Projections onto affine subspaces can be used to derive the concept of a seperating hyperplane.