Clase 15. Proyecciones en Subespacios Vectoriales.

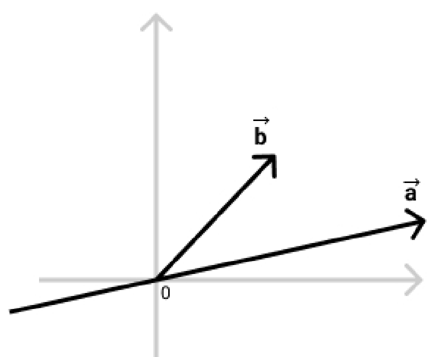
Curso 'Linear Algebra' del MIT.

Resumen

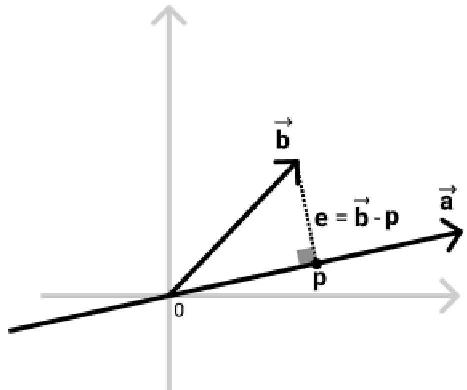
Una forma de aplicar el concepto de **ortogonalidad** visto en la clase anterior, es en la **proyección** de un vector sobre otro o entre dos subespacios. No solo estudiaremos cómo hacerlo, sino también lo usaremos para **resolver sistemas** $Ax = b$ **sin solución**, proyectando el mejor vector posible de soluciones. Un caso de aquello lo veremos en el cálculo de **mínimos cuadrados**, usado mucho en estadística (regresión lineal).

15.1 Proyección sobre un vector.

A continuación tenemos dos vectores $\vec{a}$ y $\vec{b}$ cuyas dimensiones están en $\mathbb{R}^2$.



Busquemos un punto p en $\vec{a}$ que sea lo más cercano a $\vec{b}$ o, en otras palabras, **proyectemos** este último vector sobre el primero. Aquí es donde entra el concepto de **ortogonalidad**, porque el lugar más próximo es aquel donde la recta que podemos trazar entre p y $\vec{b}$ es perpendicular a $\vec{a}$.



Entonces, como vemos en la figura de arriba, p es una aproximación en $\vec{a}$ hacia $\vec{b}$ y el error en aquella **proyección** lo calculamos como $e = \vec{b} - p$.

Si nos damos cuenta, tanto p como e podemos interpretarlos como vectores. En cuanto al primero, podemos entenderlo como el vector $\vec{a}$ escalado ya que está sobre éste. Es decir, sea c un escalar tal que $c \in \mathbb{R}$, entonces:

$$\vec{p} = \vec{a}c$$

En otras palabras, la **proyección** de $\vec{b}$ sobre $\vec{a}$ o el lugar en $\vec{a}$ que es más cercano a $\vec{b}$, **está determinado por un escalar** c. Por lo tanto, tener una forma de conocerlo es clave para nuestro propósito.

Al error e también podemos interpretarlo como $\vec{e}$ y, debido a que $\vec{p} = \vec{a}c$, entonces:

$$\vec{e} = \vec{b} - \vec{a}c$$

Recordemos que estamos asumiendo que $\vec{e}$ es ortogonal a $\vec{a}$. Es decir:

$$\vec{a}^T \cdot \vec{e} = 0$$
$$\vec{a}^T \cdot (\vec{b} - \vec{a}c) = 0$$

Por consiguiente, usemos esta igualdad para tener una fórmula que nos permita conocer el escalar c.

$$\vec{a}^T \cdot (\vec{b} - \vec{a}c) = 0$$
$$\vec{a}^T\vec{b} - \vec{a}^T\vec{a}c = 0$$
$$c = \frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$$

Implicando que:

$$\vec{p} = \vec{a} \cdot \frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$$

En la fórmula de arriba es posible notar que, en particular, $\vec{p}$ es determinado por el vector que estamos proyectando a partir de éste. Es decir, por $\vec{b}$. Por ejemplo, si $\vec{b} = 2\vec{b}$, entonces $\vec{p} = 2\vec{p}$. En cambio, si $\vec{a} = 2\vec{a}$, $\vec{p} = \vec{p}$ (i.e, la proyección se mantiene en el mismo lugar).

Por otra parte, si proyectamos $\vec{a}$ sobre este mismo, entonces $\vec{p} = \vec{a}$, puesto que $\vec{a}^T\vec{a}$ corresponde al escalar[1] $||\vec{a}||^2$:

$$\text{Si } \vec{b} = \vec{a} \text{ entonces } \vec{p} = \vec{a} \cdot \frac{\vec{a}^T\vec{a}}{\vec{a}^T\vec{a}} = \vec{a} \cdot 1 = \vec{a}$$

Y si el $\vec{b}$ es ortogonal a $\vec{a}$, entonces el vector proyección $\vec{p} = \vec{0}$, ya que $\vec{a}^T\vec{b} = 0$.

$$\vec{p} = \vec{a} \cdot \frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}} = \vec{a} \cdot 0 = \vec{0}$$

Es decir, cuando dos vectores son perpendiculares entre sí, la proyección de uno sobre el otro es el vector cero.

15.1.1 Matriz de Proyección.

El vector $\vec{p}$ también podemos escribirlo como:

$$\vec{p} = \frac{\vec{a}\vec{a}^T}{\vec{a}^T\vec{a}} \cdot \vec{b}$$

El producto $\vec{a}\vec{a}^T$ corresponde a una **matriz cuadrada**[2]. Al multiplicar el escalar $1/(\vec{a}^T\vec{a})$ a esta matriz, obtenemos la **Matriz de Proyección** P.

$$P = \left(\frac{1}{\vec{a}^T\vec{a}}\right) \cdot \vec{a}\vec{a}^T$$

De modo que la proyección $\vec{p}$ también podemos escribirla como:

$$\vec{p} = P \cdot \vec{b}$$

Esto nos lleva más al mundo del Álgebra Lineal, porque ahora la proyección de un vector sobre otro no solo lo podemos saber a partir de una constante/escalar, sino que también por medio de una matriz.

En la Clase 11 señalamos que las **matrices de rango 1** se construyen como el producto entre un vector columna y su transpuesta. En ese sentido, $\text{rango}(\vec{a}\vec{a}^T) = 1$, implicando que:

[1] Y es un producto punto, también.

[2] Si $\vec{a}$ es de $n \times 1$, entonces $\vec{a}^T$ es de $1 \times n$. En consecuencia, $\vec{a}\vec{a}^T$ es una matriz de $n \times n$.

$$\text{rango}(P) = 1$$

Por otra parte, el espacio columna[3] $C(P)$ lo podemos obtener calculando $P\vec{b}$, donde los vectores que lo generan **siempre estarán adentro de** $\vec{a}$.

Y como P es de rango 1, también signfica que $\dim(C(P)) = 1$. Es decir, su base (*basis*) corresponde a solo un vector: el vector columna $\vec{a}$.

Otra característica de la matriz de proyección, es que es **simétrica**. Es decir, $P^T = P$:

$$P^T = \left(\frac{1}{\vec{a}^T\vec{a}}\vec{a}\vec{a}^T\right)^T = \frac{1}{\vec{a}^T\vec{a}}(\vec{a}\vec{a}^T)^T = \frac{1}{\vec{a}^T\vec{a}}(\vec{a}^T)^T\vec{a}^T = \frac{1}{\vec{a}^T\vec{a}}\vec{a}\vec{a}^T = P$$

Además, si, por decir, proyectamos $\vec{b}$ dos veces sobre $\vec{a}$ que es lo mismo que $\vec{p} = P\cdot P\cdot\vec{b} = P^2\cdot\vec{b}$, este último vector se mantendrá en el mismo lugar, implicando que $P^2 = P$ y, a su vez, que $P^2\cdot\vec{b} = \vec{p}$. Lo mismo ocurrirá si lo hacemos más veces.

Entonces, resumiendo lo que hemos aprendido sobre la matriz de proyección P:

- $P = (1/(\vec{a}^T\vec{a}))\cdot\vec{a}\vec{a}^T$.
- $C(P) =$ vectores a través de $\vec{a}$.
- $\text{rango}(P) = 1$ y, por consiguiente, $\dim(C(P)) = 1$, donde $\text{base}(C(P)) = \{\vec{a}\}$.
- P es simétrica, ya que $P^T = P$
- $P^2 = P$, ya que $P\cdot P\cdot\vec{b} = P^2\cdot\vec{b} = \vec{p} = P\cdot\vec{b}$.

15.2 Resolviendo $Ax = \vec{b}$ sin solución: Proyección en un Subespacio Vectorial.

La clase anterior vimos preliminarmente cómo resolver un sistema $A\vec{x} = \vec{b}$ sin solución, donde señalamos que la idea es buscar **el mejor vector solución posible**.

Como recordaremos de la Clase 8, un sistema $A\vec{x} = \vec{b}$ tiene al menos una solución sí y solo sí $\vec{b} \in C(A)$. Es decir, $\vec{b}$ debe ser uno de los vectores que se generan en las combinaciones lineales entre los vectores columna de A y $\vec{x}$.

Un caso donde suele ser que $\vec{b} \notin C(A)$, es cuando en A de $n \times m$, $n > m$. Es decir, no solo cuando A es rectangular, sino que también cuando la cantidad de ecuaciones es mayor a la

[3] El conjunto de todas las combinaciones lineales posibles entre sus vectores columnas

de incógnitas y, por consiguiente, $\vec{b}$ tendrá más componentes de lo necesario.

Para resolver un sistema de este tipo, vamos a **proyectar a $\vec{b}$ a partir de un vector** $\vec{p} \in C(A)$, lo que también significa que existe $\vec{e} = \vec{b} - \vec{p}$ ortogonal al subespacio[4] $C(A)$.

La idea es usar a $\vec{p}$ en reemplazo de $\vec{b}$ para resolver el sistema:

$$A\hat{x} = \vec{p}$$

donde $\hat{x}$ es el **mejor vector solución posible del sistema** $A\vec{x} = \vec{b}$.

Algo a tener en cuenta, es que $\hat{x} \neq \vec{x}$, pero como $\vec{p} \in C(A)$, tenemos garantía que $\hat{x}$ existe.

Lo primero que debemos hacer, es tener una forma de encontrar a $\vec{p}$. Anteriormente señalamos que $\vec{p} = \vec{a}c$, con $c =$ constante. En ese momento trabajamos con un solo $\vec{a}$, ahora lo haremos con m de ellos.

Establezcamos que base$(C(A)) = \{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_m\}$ y están en $\mathbb{R}^n$. Es decir, son m vectores columna de n dimensiones que pertenecen al $C(A)$ y son **linealmente independientes** los que, en conjunto, conforman la matriz A.

Sabemos que existe un vector $\vec{e}$, donde $\vec{e} \perp C(A)$. Esto significa que el producto punto entre cada vector transpuesto del $C(A)$ y $\vec{e}$, es igual a cero. Como indicamos que los vectores base de dicho subespacio están contenidos en la matriz A, podemos establecer que:

$$A^T \cdot \vec{e} = \vec{0}$$

La igualdad de arriba nos confirma que $\vec{e} \perp C(A)$, porque acá el $\vec{e}$ es el $N(A^T)$ y vimos en la clase pasada en la clase pasada que el $C(A)$ es un complemento ortogonal[5] del $N(A^T)$.

Recordemos que $\vec{e} = \vec{b} - \vec{p}$:

$$A^T \cdot (\vec{b} - \vec{p}) = \vec{0}$$

Y establecimos que $A\hat{x} = \vec{p}$:

$$A^T \cdot (\vec{b} - A\hat{x}) = \vec{0}$$

Aplicando cuidadosamente un poco de álgebra:

$$A^T\vec{b} - A^T A\hat{x} = \vec{0}$$
$$A^T A\hat{x} = A^T\vec{b}$$

[4]Es decir, $\vec{p}$ es perpendicular a todos los vectores que componen el $C(A)$.

[5]Todos los vectores que componen el $N(A^T)$ son perpendiculares a todos los del $C(A)$ y viceversa.

Nuestro propósito es encontrar a $\hat{x}$, pero no podemos hacerlo de la misma manera a cómo lo hicimos con c en la sección anterior, porque A^TA es una matriz, no un escalar. No obstante, si bien A es rectangular, nos encontramos ante dos hechos útiles:

1. A^TA **siempre** es simétrica, lo que significa que también es cuadrada.
2. Los vectores columna de A son linealmente independientes. Por lo tanto, también lo son en A^TA.

En consecuencia, A^TA es **simétrica** e **invertible**.

Con este hecho, multipliquemos a la ecuación de arriba por $(A^TA)^{-1}$ (la inversa de A^TA).

$$\begin{aligned}(A^TA)^{-1}A^TA\hat{x} &= (A^TA)^{-1}A^T\vec{b}\\ I\hat{x} &= (A^TA)^{-1}A^T\vec{b}\\ \hat{x} &= (A^TA)^{-1}A^T\vec{b}\end{aligned}$$

Entonces, con la fórmula de arriba podemos encontrar a $\hat{x}$, el mejor vector solución posible.

Por otra parte, también podemos conocer al vector de proyección $\vec{p}$ puesto que, como $\hat{x} = (A^TA)^{-1}A^T\vec{b}$, entonces:

$$\begin{aligned}A\hat{x} &= \vec{p}\\ A(A^TA)^{-1}A^T\vec{b} &= \vec{p}\end{aligned}$$

Esta expresión podemos reducirla un poco más, porque $A(A^TA)^{-1}A^T$ es la **Matriz de Proyección** P:

$$P = A(A^TA)^{-1}A^T$$

Por lo tanto:

$$P\vec{b} = \vec{p}$$

De este modo, también podemos decir que para resolver un sistema $A\vec{x} = \vec{b}$ sin solución, buscamos la matriz de proyección P, la que nos permite obtener el vector $\hat{x}$ de las mejores soluciones posibles.

Esta matriz P también se caracteriza por ser simétrica ($P^T = P$) y por $P^2 = P$. En cuanto a lo primero:

$$
\begin{aligned}
P^T &= (A(A^TA)^{-1}A^T)^T \\
&= (A^T)^T((A^TA)^{-1})^TA^T \\
&= A((A^TA)^T)^{-1}A^T \\
&= A(A^T(A^T)^T)^{-1}A^T \\
&= A(A^TA)^{-1}A^T \\
P^T &= P
\end{aligned}
$$

Y con respecto a $P \cdot P = P^2 = P$:

$$
\begin{aligned}
P^2 &= A(A^TA)^{-1}A^T \cdot A(A^TA)^{-1}A^T \\
&= A(A^TA)^{-1} \cdot (A^TA(A^TA)^{-1})A^T \\
&= A(A^TA)^{-1} \cdot I \cdot A^T \\
&= A(A^TA)^{-1}A^T \\
P^2 &= P
\end{aligned}
$$

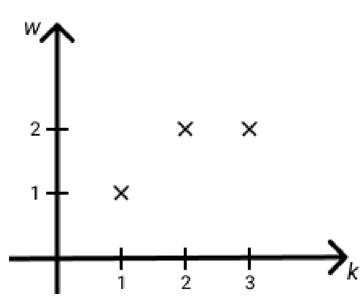
15.3 Método de Mínimos Cuadrados.

Una aplicación de la proyección en un subespacio vectorial, es en la búsqueda de una ecuación lineal que se **ajuste** de la mejor manera posible a un conjunto de datos que no siempre tiene una relación lineal.

Por ejemplo, digamos que tenemos dos variables k y w y queremos predecir los valores que puede tomar la segunda a partir de la primera. Es decir, asumamos que w es dependiente de k. Modelaremos esta relación a partir de una ecuación lineal:

$$w = c + dk \text{ (} c \text{ y } d \text{ constantes)}$$

Supongamos que las variables toman los valores $k = \{1,\ 2,\ 3\}$ y $w = \{1,\ 2,\ 2\}$. El problema es que no siguen una relación estrictamente lineal, como lo vemos a continuación.



Por lo tanto, no podemos predecir los valores que puede tener w con una sola ecuación. En este caso, vamos a necesitar tres. Es decir, una para cada punto:

$$\begin{cases} 1 = c + 1d \\ 2 = c + 2d \\ 2 = c + 3d \end{cases}$$

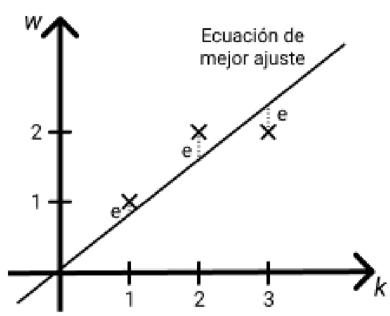
En otras palabras, vamos a resolver este sistema de ecuaciones lineales para conocer los valores que puede ir tomando w. Sin embargo, nos enfretamos a un segundo problema: tenemos más ecuaciones (3) que incógnitas (2).

En efecto, si pasamos estas ecuaciones matrices:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \vec{x} = \begin{bmatrix} c \\ d \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

ocurre que $A\vec{x} = \vec{b}$ no tiene solución.

Como, en realidad, nuestro propósito inicial es usar una sola ecuación para predecir los valores de w, lo que haremos es **proyectar una ecuación lineal que se ajuste mejor a los datos**.



La ecuación lineal que tenga el mejor ajuste, será aquella que pase más cerca de todos los puntos. Es decir, aquella donde el error e de cada punto sea el más pequeño. Todo esto se conoce como el **Método de Mínimos Cuadrados** y lo aplicaremos multiplicando el sistema $A\vec{x} = \vec{b}$ por A^T:

$$A^T A\hat{x} = A^T\vec{b}$$

Ya sabemos que, al resolverlo, conoceremos $\hat{x}$, $\vec{p}$ y P. Todo esto lo veremos la siguiente clase.