SWCON253 Machine Learning

Lecture 6. Subspaces, Bases, and Projections in Machine Learning

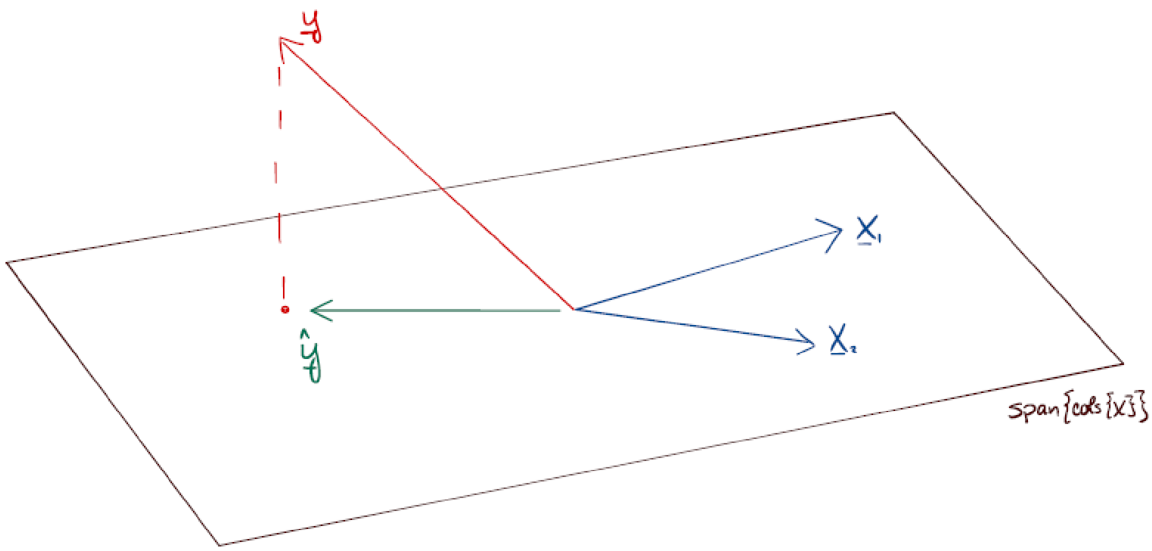
Recall geometric view of least squares

Given $(\underline{x}_i, y_i)$ for i=1, ...,n
Labels $\underline{y} \in \mathbb{R}^p$ for n training samples
Features $X \in \mathbb{R}^{nxp}$ (p features)

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} \cdots \underline{x}_1^T \cdots \\ \cdots \underline{x}_2^T \cdots \\ \\ \cdots \underline{x}_n^T \cdots \end{bmatrix} \in \mathbb{R}^{nxp}$$

We want to find $\underline{\hat{y}} = X\underline{w}$ such that $\left\|\underline{\hat{y}} - \underline{y}\right\|_2^2$ is as small as possible

Let $X_1, X_2, \ldots, X_p$ = p columns of X.
Then, $\hat{y} = w_1 X_1 + w_2 X_2 + \cdots + w_p X_p$



The hyperplane span{cols(X)} is called a **subspace**
If the columns of X are linearly independent, then they form a basis for $\mathcal{X}$.
$\hat{y}$ is the **orthogonal projection** of y onto the subspace.
The 2 columns of X in the image above **span** the subspace.

We will use this notion of least squares with a motivating example.

Subspaces

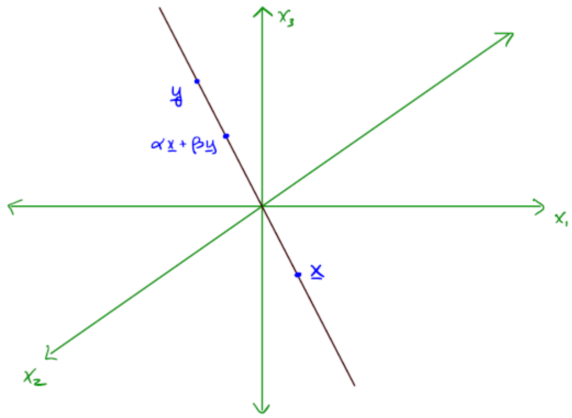
Consider all points $X \in \mathbb{R}^n$. A subspace is a subset of these points that satisfies a few key properties: Specifically, let $\mathcal{S}$ be a subspace and let $\underline{x}$ and $\underline{y}$ be any two points in the subspace. Then for any scalars α and β, the weighted sum $\alpha\underline{x} + \beta\underline{y}$ must also be in the subspace.

Ex 1. n=3, S={$\underline{x} \in \mathbb{R}^3$: $x_1 = x_2 = -x_3$}

$\underline{x} \in S$

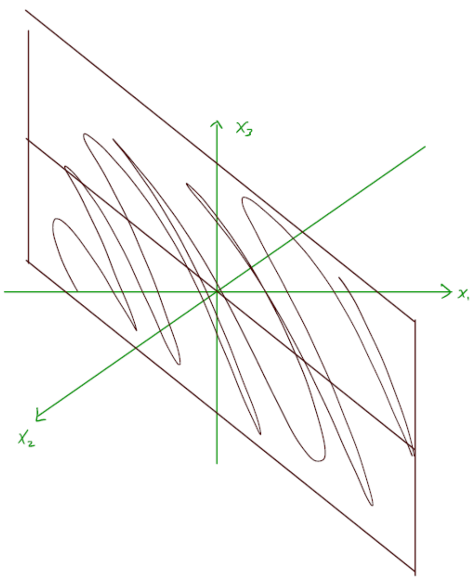
$\underline{x} = a\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ for some a

x_1=a, x_2=a, x_3=-a

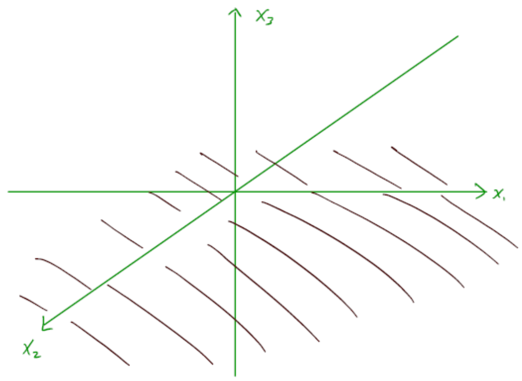


Ex 2. n=3, S={$\underline{x} \in \mathbb{R}^3$: $x_1 = x_2$}

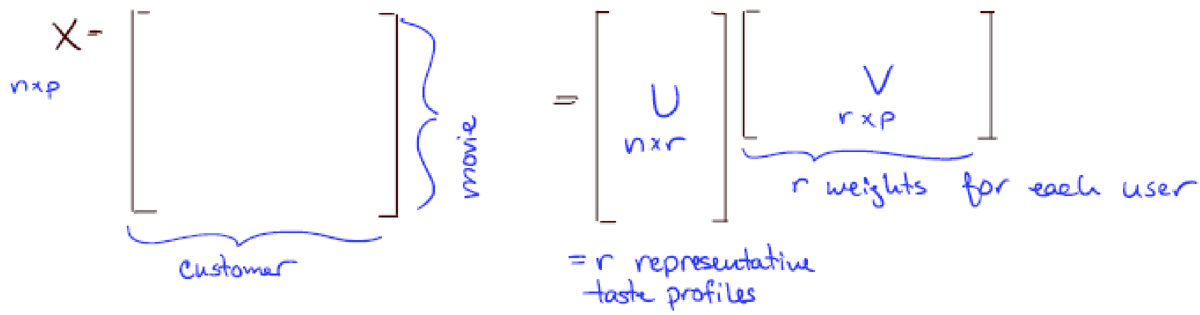
Vertical plane along diagonal



Ex 3. n=3, S={$\underline{x} \in \mathbb{R}^3$: $x_3 = 0$}
Horizontal plane



Ex 4, recommender system



X_{ij} =rating of i^{th} movie by j^{th} customer (user)

The span of columns of U is a subspace.
All columns of X lie in that subspace.

For example, for one column of X, we can think about this column as **a weighted sum** of the columns of U and the j^{th} column of V that tells us what those **weights** are. So every column of X is a weighted sum of the columns of U for some sort of weights (V) here and this again coincides with our notion of subspaces because the subspace corresponding to the span of the columns of U.

How to represent a subspace?

(a) Represent $\mathcal{S}$ as the span of a set of vectors

(b) Represent $\mathcal{S}$ as the span of a set of linearly independent vectors (called subspace **basis**)

(c) Represent $\mathcal{S}$ as the span of a set of orthonormal vectors (called subspace **orthonormal basis**)

Recall

n=3, $\mathcal{S}=\{\underline{\mathrm{x}} \in \mathbb{R}^3 : \; x_3 = 0\} \rightarrow$ horizontal plane

(a) $\mathcal{S}$ = span$\left\{\begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\}$

(b) $\mathcal{S}$ = span$\left\{\begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}\right\}$

(c) $\mathcal{S}$ = span$\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\}$

Two vectors $\underline{u}_1$ and $\underline{u}_2$ are orthogonal if $\langle \underline{u}_1, \underline{u}_2 \rangle = \underline{u}_1^T \underline{u}_2 = \underline{u}_2^T \underline{u}_1 = 0$

A vector u is normal if $\|u\|_2 = \|u\|_2^2 = \langle \underline{u}, \underline{u} \rangle = \underline{u}^T \underline{u} = 1$

A set of vectors $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_p$ is orthonormal if

$\langle \underline{u}_i, \underline{u}_j \rangle = 1$ if i=j

$\langle \underline{u}_i, \underline{u}_j \rangle = 0$ if i≠j

Properties of the orthonormal basis matrix

If $\mathcal{S}$ = span$\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p\}$ where the vectors are orthonormal, then

U = $\begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \dots & u_p \\ \vdots & \vdots & & \vdots \end{bmatrix}$ is a (orthogonal) basis matrix
U is an orthogonal matrix

C =U^TU $\rightarrow$ C_{ij} = $\langle \underline{u}_i, \underline{u}_j \rangle$
$\langle \underline{u}_i, \underline{u}_j \rangle$=1 if i=j
$\langle \underline{u}_i, \underline{u}_j \rangle$=0 if i≠j

C = $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$=I

U is (squared) length preserving
Let $\underline{\mathrm{v}} \in \mathbb{R}^p$ Consider Uv

$$||Uv||_2^2 = (Uv)^T(Uv) = v^T U^T Uv = v^T v = ||v||_2^2$$

Dimension of subspace
dim($\mathcal{S}$) = number of vectors in subspace basis
If $\mathcal{S}$=span(cols(X)), then dim($\mathcal{S}$)=rank(X)

$X \in \mathbb{R}^{nxp}$

Projection

The projection of a point $\underline{y}$ onto a set is the point in the set closest to $\underline{y}$.

$\underline{\hat{y}}$ = projection of $\underline{y}$ onto set $\mathcal{X}$ = $P_{\mathcal{X}}\underline{y}$ = $\underset{\underline{x}\in\mathcal{X}}{\operatorname{argmin}} \left\|\underline{y}-\underline{x}\right\|_2^2$

If $\mathcal{X}$ is a subspace spanned by columns of $X \in \mathbb{R}^{nxp}$ with LI columns, any point in $\mathcal{X}$ has form $\underline{\hat{y}} = w_1\underline{x}_1 + w_2\underline{x}_2 + \cdots + w_p\underline{x}_p$.

Let $\underline{\hat{w}} = \underset{\underline{w}}{\operatorname{argmin}} \left\|\underline{y} - X\underline{w}\right\|_2^2$ and $\underline{\hat{y}} = X\underline{\hat{w}}$

Least squares
$\underline{\hat{\boldsymbol{w}}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\underline{\boldsymbol{y}}$

Projection matrix
$\underline{\hat{y}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\underline{y}$
$= P_X\underline{y}$

P_X is squares

$P_X = P_X^2\underline{y} = P_X P_X$

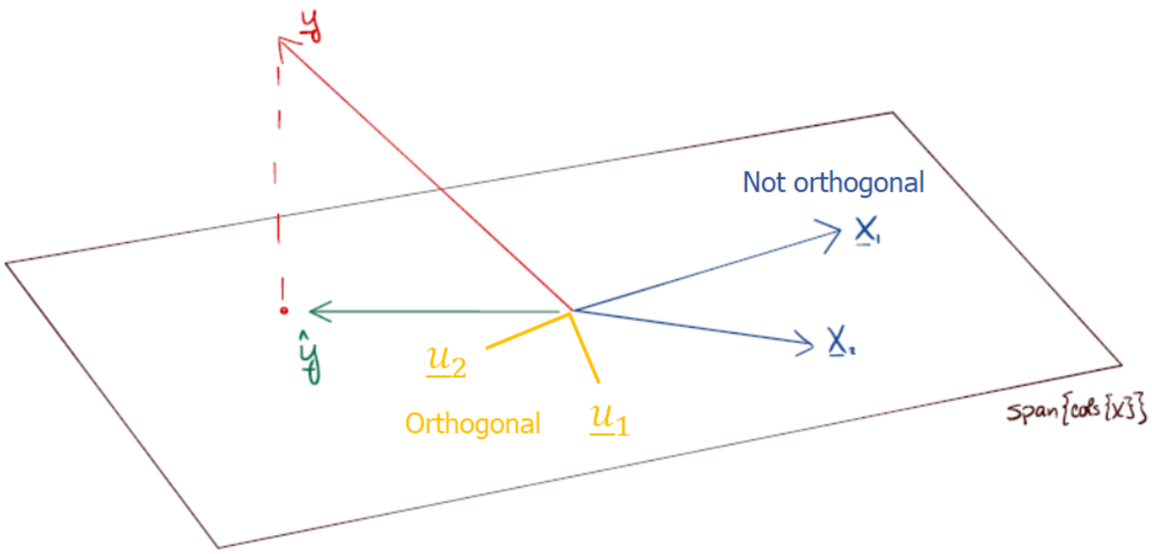
If $\underline{\hat{y}} \in \mathcal{X}$, then $P_{\mathcal{X}}\underline{\hat{y}} = \underline{\hat{y}}$

Orthogonal Subspace Bases and Least Squares

Let $X \in \mathbb{R}^{nxp}$, $\underline{y} \in \mathbb{R}^n$
Let U be orthonormal basis matrix for subspace spanned by columns of X
span(cols(U)) = span(cols(X))

$\underline{\hat{y}} = X\underline{\widehat{w}} \; = \; U\underline{\widetilde{w}}$ for any $\underline{\hat{y}} \in \mathcal{X}$, there are both $\underline{\widehat{w}}, \underline{\widetilde{w}}$ so that $\underline{\hat{y}} = X\underline{\widehat{w}} \; = \; U\underline{\widetilde{w}}$

We can write any point in the space as
$w_1\underline{x}_1 + w_2\underline{x}_2 \; = \; \widetilde{w}_1\underline{u}_1 + \widetilde{w}_2\underline{u}_2$



Use least squares to find $\underline{\widetilde{w}}$

$$\underline{\widetilde{w}} = \underset{\underline{w}}{\operatorname{argmin}} \left\| \underline{y} - U\underline{w} \right\|_2^2$$

$=(U^TU)^{-1}U^T\underline{y}$

$\underline{\hat{y}} = \quad \boldsymbol{U(U^TU)^{-1}U^T\underline{y}} \quad\quad = \quad\quad \boldsymbol{X(X^TX)^{-1}X^T\underline{y}}$

Projection onto span(cols(U)) = Projection onto span(cols(X))

$\mathrm{U(U^TU)^{-1}U^T}$

$\mathrm{UIU^T = UU^T}$