# Lecture 8: Least Squares, Subspaces, & Classification

Last time model $\hat{y} = X\underline{w} \longrightarrow X \in \mathbb{R}^{n \times p}$

$\hat{y} = w_1 \underline{x}_1 + w_2 \underline{x}_2 + \cdots + w_p \underline{x}_p$

$\underline{x}_2$: 2nd col of $X$

$\hat{y} \in \text{Span}\{x_1, \ldots, x_p\}$ ← this a subspace.

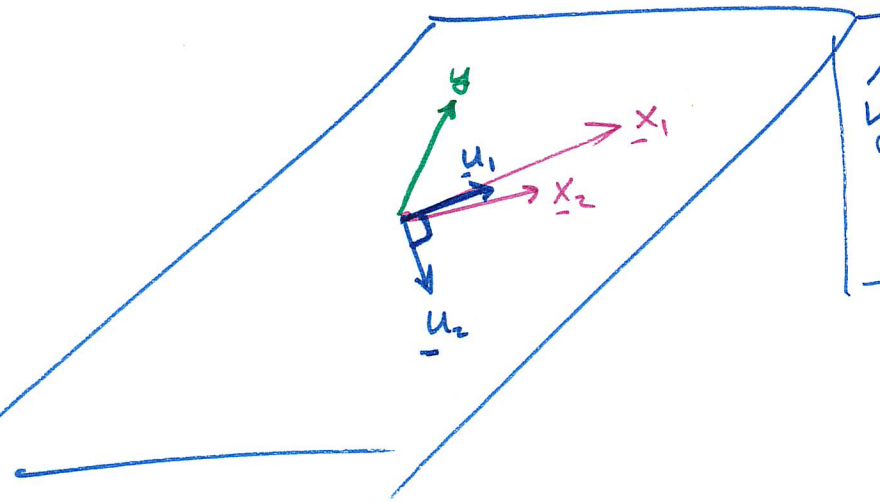
Find an orthonormal basis for subspace (via Gram-Schmidt orthogonalization); let $U$ be matrix w/ basis vectors as columns: $U = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r]$. Then $U^T U = I$.

Each col of $U$, $u_i$ is $\in$ Span $\{x_1, \ldots, x_p\}$

Let $u_i = \alpha_{i1} x_1 + \alpha_{i2} x_2 + \alpha_{i3} x_3 + \cdots + \alpha_{ip} x_p$

or $U = XA$ what $i^{th}$ col of $A$ is $a_i = \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ip} \end{bmatrix} \iff u_i = X a_i$

$n \times r \quad n \times p \quad p \times r$




$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_p x_p$

$= v_1 u_1 + v_2 u_2 + \cdots + v_r u_r$

$= v_1 X a_1 + v_2 X a_2 + \cdots + v_r X a_r$

$= v_1 (\alpha_{11} x_1 + \alpha_{12} x_2 + \cdots + \alpha_{1p} x_p)$

$+ v_2 (\alpha_{21} x_1 + \alpha_{22} x_2 + \alpha_{2p} x_p)$

$\vdots$

$+ v_r (\alpha_{r1} x_1 + \alpha_{r2} x_2 \cdots + \alpha_{rp} x_p)$

$= x_1 (v_1 \alpha_{11} + v_2 \alpha_{21} \cdots + v_r \alpha_{r1})$ $w_1$

$+ x_2 (v_2 \alpha_{21} \cdots)$ $w_2$

$+ \cdots + x_p (\quad)$

$\hat{y} = XAv$

$w$

Given new sample $x_{new} \in \mathbb{R}^p$, how to predict label $y_{new}$?

(A) $\boxed{\hat{y}_{new} = \langle \underline{x}_{new}, \hat{\underline{w}} \rangle}$

(B) ~~=~~ what if using orthonormal basis $U$?

recall $U = XA$ (n×r = n×p p×r) $\Rightarrow$ $\underline{u}_{new}^T = \underline{x}_{new}^T A$

or $\underline{u}_{new} = A \underline{x}_{new}$

$\Rightarrow \boxed{\hat{y}_{new} = \langle \underline{u}_{new}, \hat{v} \rangle}$

So what is A? if cols of X are linearly independent ($r = p$)

Recall $\underline{u}_i = X \underline{a}_i$. To get $\underline{a}_i$, use LS: $\underline{a}_i = (X^T X)^{-1} X^T \underline{u}_i$

---

Prediction using orthonormal basis: Given $X \in \mathbb{R}^{n \times p}$, $y \in \mathbb{R}^n$

① compute basis $U$ from $X$ (Gram Schmidt)

② $\hat{v} = \underset{v}{\operatorname{argmin}} \| y - U \underline{v} \|_2^2 = U^T y$

③ out of sample prediction: $x_{new} \rightarrow u_{new}$; $\hat{y}_{new} = \langle \underline{u}_{new}, \hat{v} \rangle$

**Theorem:** Let $X \in \mathbb{R}^{n \times p}$, $n \geq p$, $X$ is full rank (p cols are lin. indep)

$y \in \mathbb{R}^n$

and let $u_1, \ldots, u_p$ be orthonormal basis vectors

for $\text{span}\{x_1, \ldots, x_p\} = \text{span}\{u_1, \ldots, u_p\}$

Then $\hat{y} = X\hat{w}$ where $\hat{w} = \underset{w}{\text{argmin}} \|y - Xw\|_2^2$

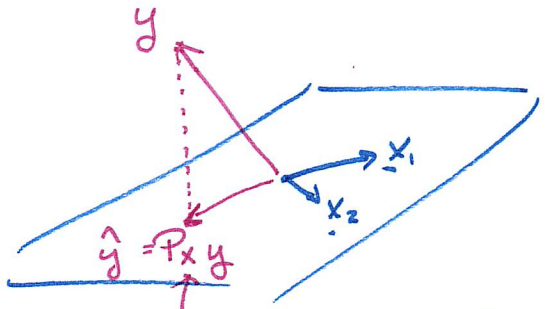
is given by $\hat{y} = UU^T y$ where $U = [u_1, \ldots, u_p]$

note: because $U$ is an orthogonal matrix, $U^TU = I \neq UU^T$

**Proof:**

$$\hat{y} = X\hat{w} = \underbrace{X(X^TX)^{-1}X^T}\, y$$

"projection matrix" $P_X$




recall $\text{span}\{\text{cols of } X\} = \text{span}\{u_1, \ldots, u_p\}$

$$\Rightarrow P_X y = P_U y$$

$$P_X = P_U = U \underbrace{(U^T U)^{-1}}_{=I} U^T$$

$$= UU^T$$

$$\hat{y} = P_X y = P_U y = UU^T y$$

---

## Least Squares & Classification

Setup: $n$ training samples $(x_i, y_i) \in \mathbb{R}^p \times \{-1, 1\}$ for $i = 1, \ldots, n$

$x_i \in \mathbb{R}^p$ $\quad$ $y_i \in \{-1, +1\}$

$$y \approx Xw \qquad X = \begin{bmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \{-1, +1\}^n$$

Assume $n \geq p$, $X$ full rank

$$\hat{w} = \arg\min_w \|y - Xw\|_2^2 = (X^T X)^{-1} X^T y$$

$\hat{y}$ = predicted labels on training sample

$$= X\hat{w} = X(X^T X)^{-1} X^T y = P_X y$$

## Predictions

given new sample $x_{new} \in \mathbb{R}^p$, want to predict $y_{new}$.

(A) $\hat{y}_{new} = x_{new}^T \hat{w} \in \mathbb{R}$

(B) $\hat{y}_{new} = \text{sign}(x_{new}^T \hat{w}) \in \{+1, -1\}$

} "linear classifier"

Classification rule: predict $+1$ if $x_{new}^T w > 0$

$-1$ if $x_{new}^T w < 0$