Lecture 21 — 2nd Midterm Review

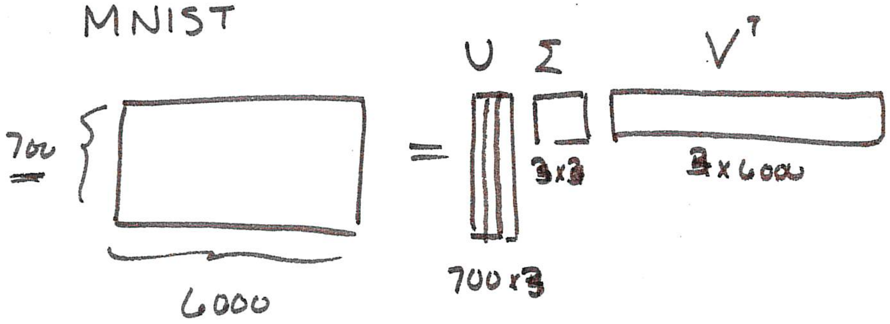
7:15 @ EH 1800

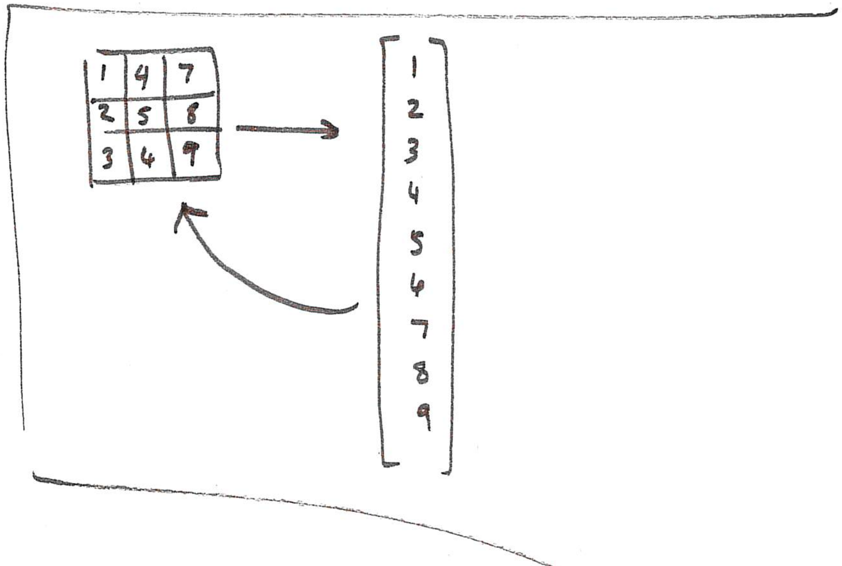
1 8½ x 11" notes
handwritten, one
size.

SVD + PCA

$$X = U\Sigma V^T$$

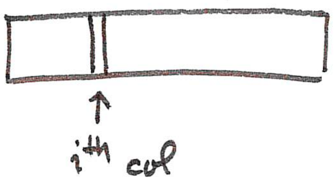
MNIST





$$U^T X = \underline{U^T U}\Sigma V^T = \underbrace{\Sigma V^T}_{3 \times 6000} =$$

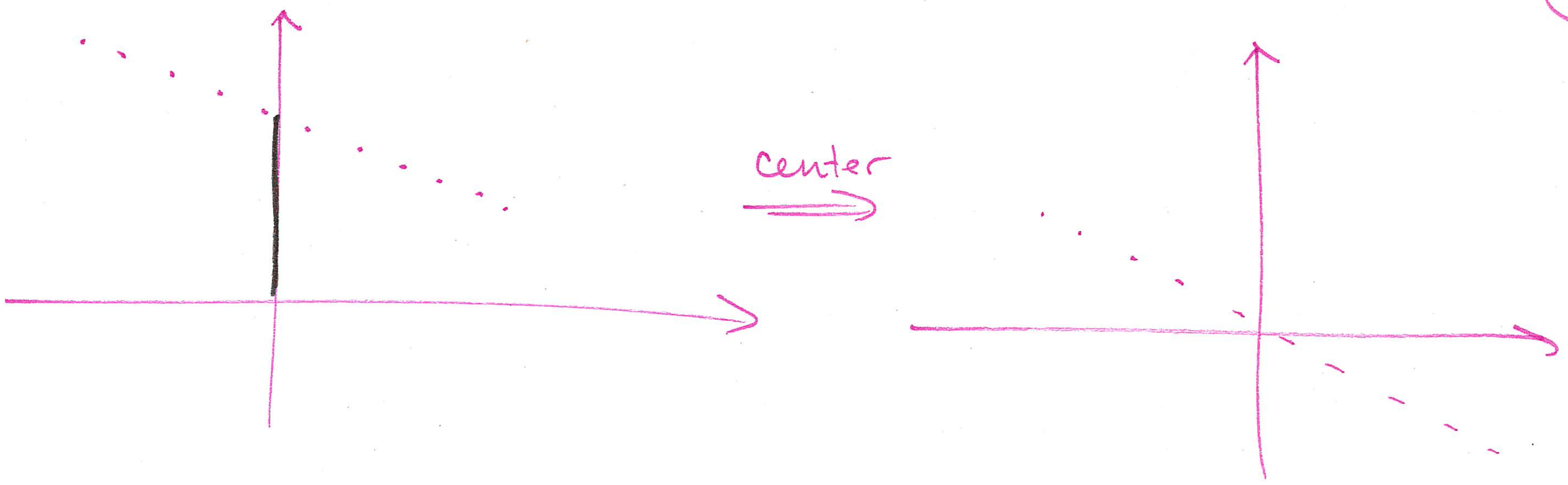
"W"



= Pric. comp of ith sample

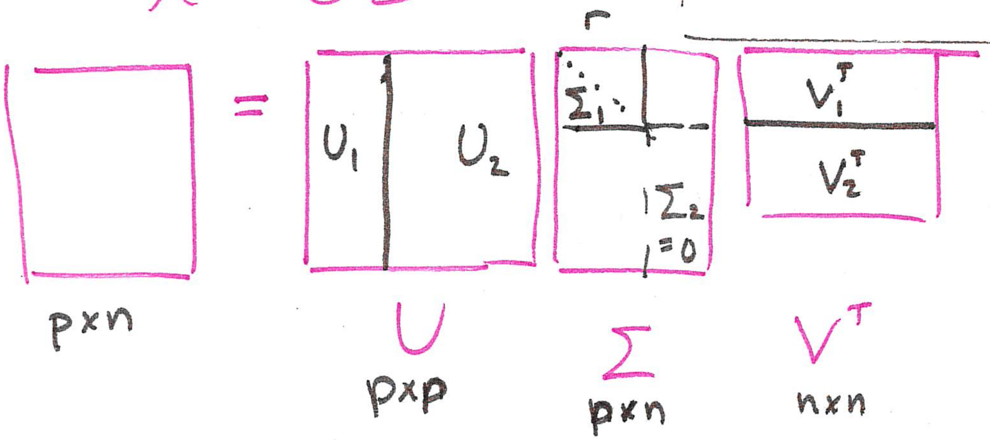
alt: $Y = X^T \in \mathbb{R}^{6000 \times 700}$

$$\mathrm{SVD}(Y) = V\Sigma U^T$$



Classification using SVD

$$X = U\Sigma V^T = \boxed{U_1\Sigma_1V_1^T + U_2\Sigma_2V_2^T}$$



① $\tilde{X} = U_1\Sigma_1V_1^T$

② $\underbrace{\overset{Z}{\cancel{W}}}_{r\times n} = U_1^T\tilde{X} = \Sigma_1V_1^T$

$$Z\ \cancel{W} = \left[\overset{z_1}{\cancel{w_1}}\ \ \overset{z_2}{\cancel{w_2}}\ \cdots\ \overset{z_n}{\cancel{w_n}}\right]$$

③ use z_i's as training sample feature vectors

$V^TV = I$

$V_1^TV_2 = 0$

④ $y_i \approx w^T z_i$

$Xw = 0$?

$$Xw = \underbrace{U_1}_{1^{st}\text{ 2 cols}} \Sigma_1 V_1^T w + \underset{\substack{\uparrow \\ 3^{rd}\text{ col}}}{U_2} \underset{\substack{\uparrow \\ \text{zeros}}}{\Sigma_2} V_2^T w$$

$$Xw = U_1 \Sigma_1 V_1^T w$$

$Xw = 0$ if $U_1 \Sigma_1 V_1^T w = 0$

or $\underline{V_1^T w} = 0$

if $\boxed{w = V_2 c}$, then $V_1^T w = V_1^T V_2 c$

$= 0$

V_2 tells us null space of X when

$\Sigma_2 = 0$

2 regimes

~~Ⓐ~~ $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$

Ⓐ $\Sigma_2 = 0$

$\Rightarrow$ rank(X) = sidelength in Σ_1

Ⓑ $\Sigma_2 \neq 0$, small

Σ_2 small $\Longrightarrow$ subspace approximation

assume $\underline{\underline{\Sigma_1 \sim r \times r}}$

$\hat{X}^{(r)} :=$ best rank-r approximation to X

$$= U_1 \Sigma_1 V_1^T$$

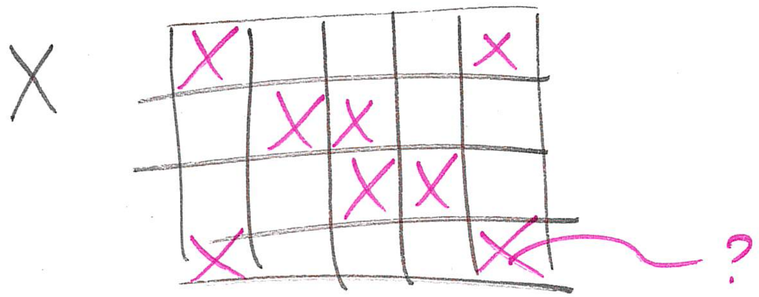
$$X = U_1 \underbrace{\Sigma_1}_{r \times r} V_1^T + \underline{\underline{U_2 \Sigma_2 V_2^T}}$$

$$X - \hat{X}^{(r)} = U_2 \Sigma_2 V_2^T$$

$$\|X - \hat{X}^{(r)}\|_F^2 = \|\underline{\underline{U_2}} \Sigma_2 \underline{\underline{V_2^T}}\|_F^2$$

$$= \|\Sigma_2\|_F^2 = \sum_{i=r+1}^{\min(n,p)} \sigma_i^2$$

Matrix Completion



assume X is rank r

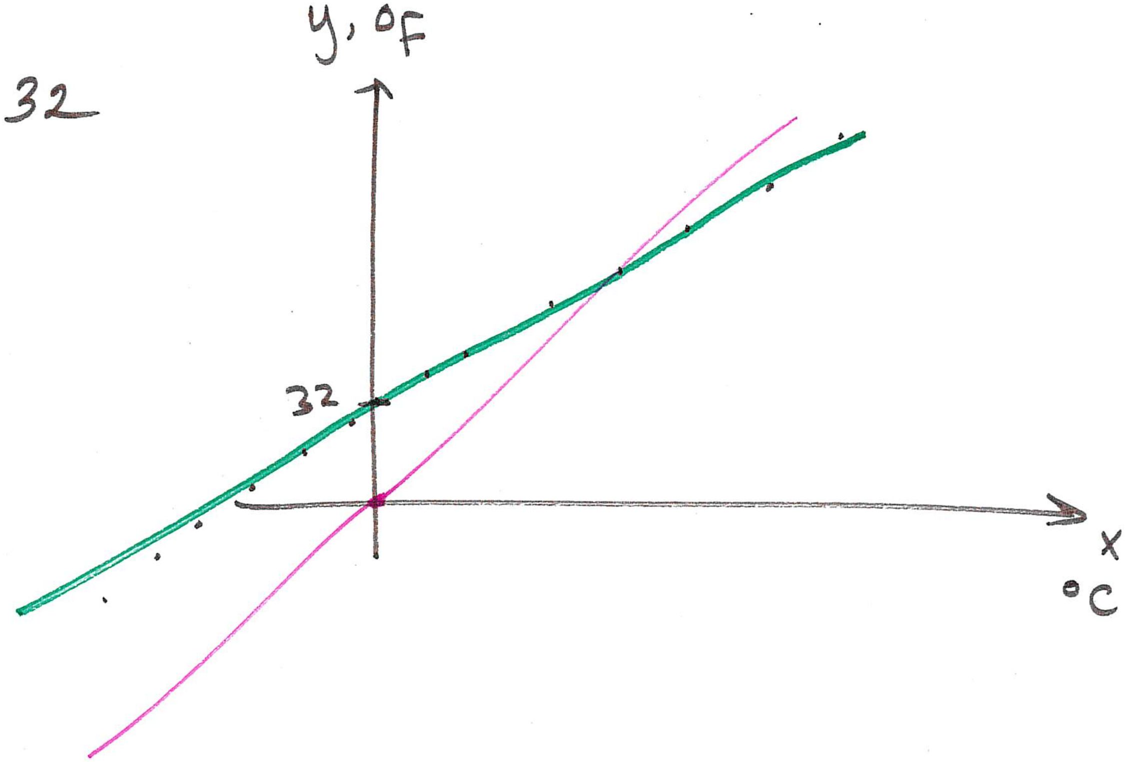
$\hat{X}$ = estimate.

- find rank-r approx to $\hat{X}$
- fill in the missing entries

Offset in SVM

$$°F = °C \cdot 1.8 + 32$$

$$y_i \qquad x_i$$



w/o offset

$$y_i = x_i \cdot w$$

w/o offset

$$y_i = x_i w_1 + w_0$$

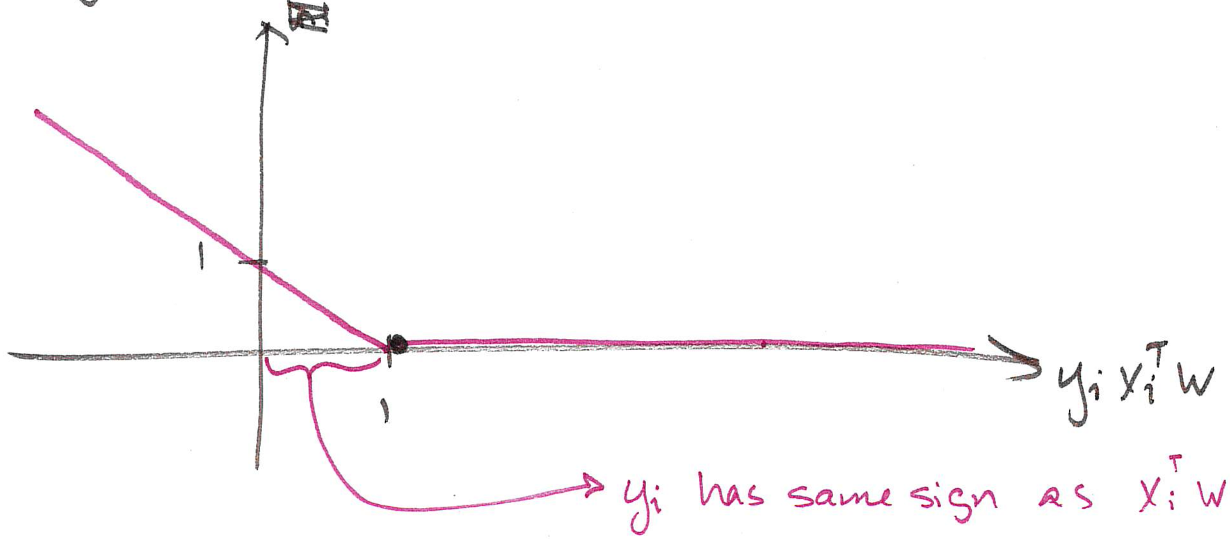
Convergence of GD and operator norm

$$\min_{w} \underbrace{\|y - Xw\|_2^2}_{f(w)}$$

GD: $\hat{w}^{(k+1)} = \hat{w}^{(k)} - \tau X^T(X\hat{w}^{(k)} - y)$

claim: if $\tau < \frac{2}{\|X\|_{op}^2}$, then GD converges

hinge loss



y_i has same sign as $x_i^T w$ (x_i is correcly classified)
but loss $\neq 0$

$$X = \begin{bmatrix} 1 & & & 0 \\ & 2 & & \\ & & 3 & \\ 0 & & & 4 \end{bmatrix}$$

$$\text{SVD} \Rightarrow U = I,\ V = I,\ \Sigma = \begin{bmatrix} 1 & & & 0 \\ & 2 & & \\ & & 3 & \\ 0 & & & 4 \end{bmatrix} \quad ✗$$

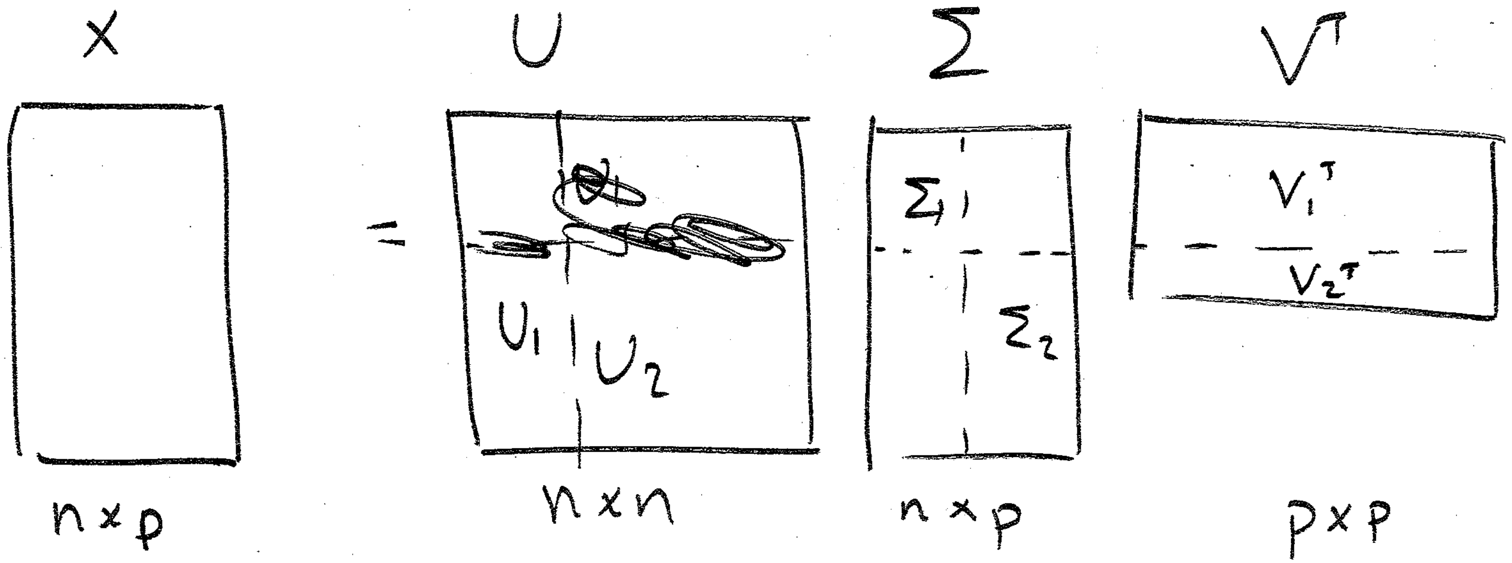
$$\text{instead } U = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & 0 \end{bmatrix},\ V = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & 0 \end{bmatrix},\ \Sigma = \begin{bmatrix} 4 & & & 0 \\ & 3 & & \\ & & 2 & \\ 0 & & & 1 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow U = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Big/ \sqrt{3}$$

$$UU^T = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} / 3$$

$$= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} / 3$$

$$\Rightarrow U \cdot 3 \cdot U^T = X$$



$U^T U = U U^T = I$

$V^T V = V V^T = I$

$U_1^T U_1 = I \ , \quad U_1 U_1^T \neq I$

$U_1^T U_2 = 0$