Questions about HW 2?

$$\vec{x}^T \vec{y} = \vec{y}^T \vec{x} \qquad \underbrace{\vec{v}^T B \vec{u}}_{\text{scalar}} = \vec{u}^T B^T \vec{v}$$

Taking derivatives
or gradients $\longrightarrow$ see video

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$(\nabla_{\vec{x}} C)_i = \underline{\underline{\frac{\partial C}{\partial x_i}}} = \sum_j (\cdots) \underbrace{\frac{\partial x_j}{\partial x_i}}_{\substack{1 \text{ if } i=j \\ 0 \text{ if } i \neq j}}$$

Reference to eqn.

Linear algebra review

- Tuesday 10/6 at noon
- will be recorded

$$\underline{X} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} .5 \\ 0 \end{bmatrix}$$

$$\underline{X}^T \left(\underline{X}\vec{\beta} \approx \vec{y} \right)$$

normal eqns. : $\underline{X}^T \underline{X} \vec{\beta} = \underline{X}^T \vec{y}$

least-squares $\vec{\beta}$

$$C = \| \underline{X}\vec{\beta} - \vec{y} \|^2$$

$$\nabla_{\vec{\beta}} (\text{cost}) = 0$$
$$= 2\underline{X}^T \underline{X}\vec{\beta} - 2\underline{X}^T\vec{y}$$

1) cost function → minimize ... zero gradient
2) compute gradient, set $= 0$
3) that gives normal equs.

$$\vec{\beta}^* = \arg\min_{\vec{\beta}} \|\underline{X}\vec{\beta} - \vec{y}\|^2$$

"vector argument that minimizes"

Last time

SVD of $\underline{X} = USV^T$

$$\vec{\beta}^* = \underline{X}^+ \vec{y} = VS^+U^T\vec{y} = \sum_{i=1}^{\text{rank}(\underline{X})} \vec{v}_i \left(\frac{\vec{u}_i^T \vec{y}}{\sigma_i} \right)$$

solves normal equations

Uses of SVD:

- pseudoinverse, solves linear least-squares
- low-rank approximation

$\|\cdot\|_{Fro} = \sqrt{\sum_{i,j} x_{ij}^2}$

$\|\cdot\|_2$

given X find $X_r = \arg\min_{X'} \|X - X'\|$ where $\text{rank}(X') = r$

where

$X = USV^T$

$\text{rank}(X) > r$

$$X_r = \sum_{i=1}^{r} \vec{u}_i \vec{v}_i^T \sigma_i$$

same as setting σ_i, $i > r$, $= 0$

keep r largest singular vectors

- PCA principal components analysis
  unsupervised learning method
  (no labels)

$$\underset{(n \times d)}{X} = \begin{bmatrix} - \vec{x}_1 - \\ \vdots \\ - \vec{x}_n - \end{bmatrix}$$

dimension reduction

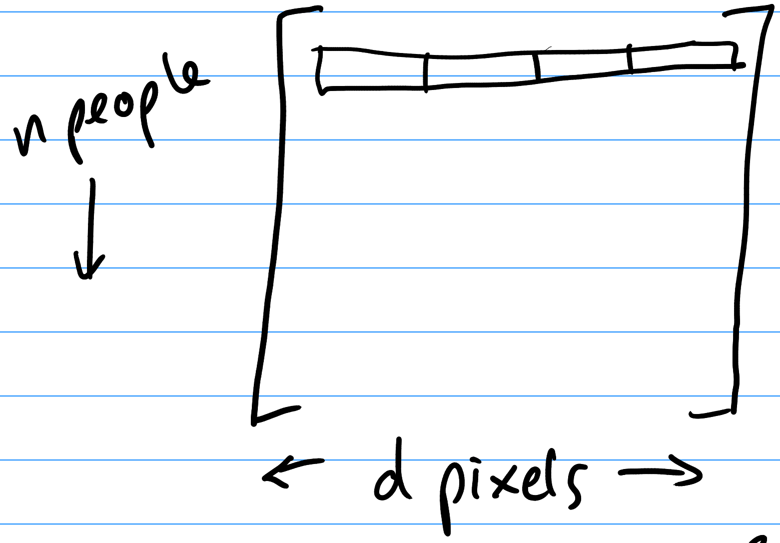
pick some new dimension $r < d$ to reduce to

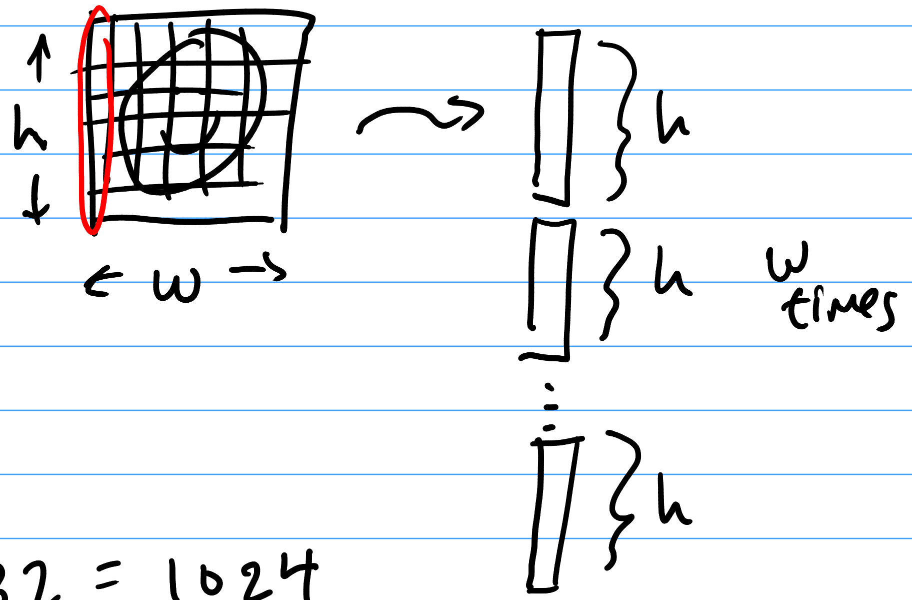
$$\downarrow$$

$$X' = \underbrace{\begin{bmatrix} - \tilde{x}_1 - \\ \vdots \\ - \tilde{x}_n - \end{bmatrix}}_{r < d}$$

$d = w \cdot h$

ex/

$X =$ 



$32 \times 32 = 1024$

PCA

1. subtract off mean $\vec{\mu} = \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i$
2. take SVD of $(\underbrace{X - \overbrace{\vec{1}\,\vec{\mu}^T}^{\text{matrix w/ } \vec{\mu}^T \text{ in every row}}}_{\text{mean-subtracted matrix}})$

call this matrix Y

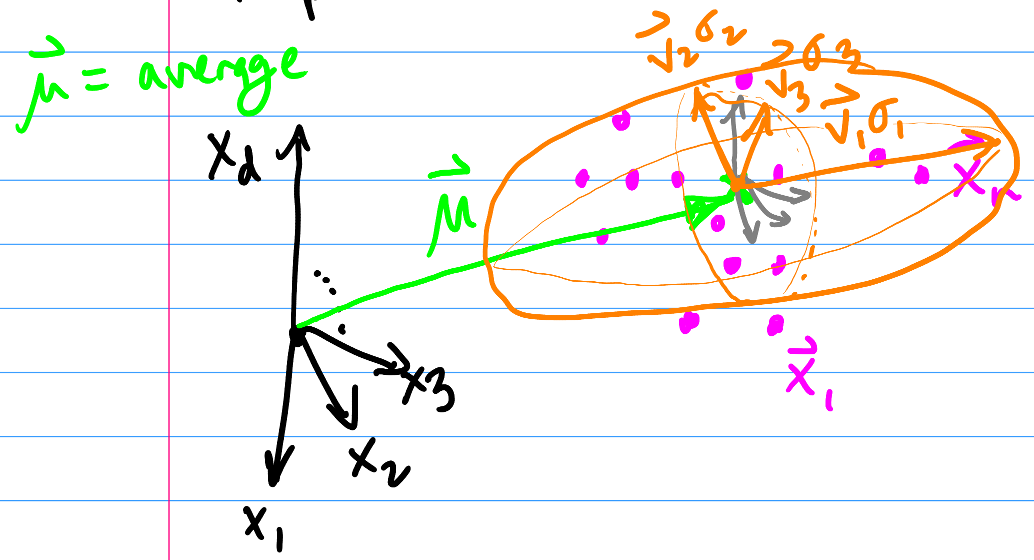
$$\underset{(n\times d)}{Y} = \underset{(n\times n)}{U}\, S\, \underset{(d\times d)}{V^T}$$

$$Y^T Y = V S^2 V^T$$

columns $\vec{v}_i$ of V

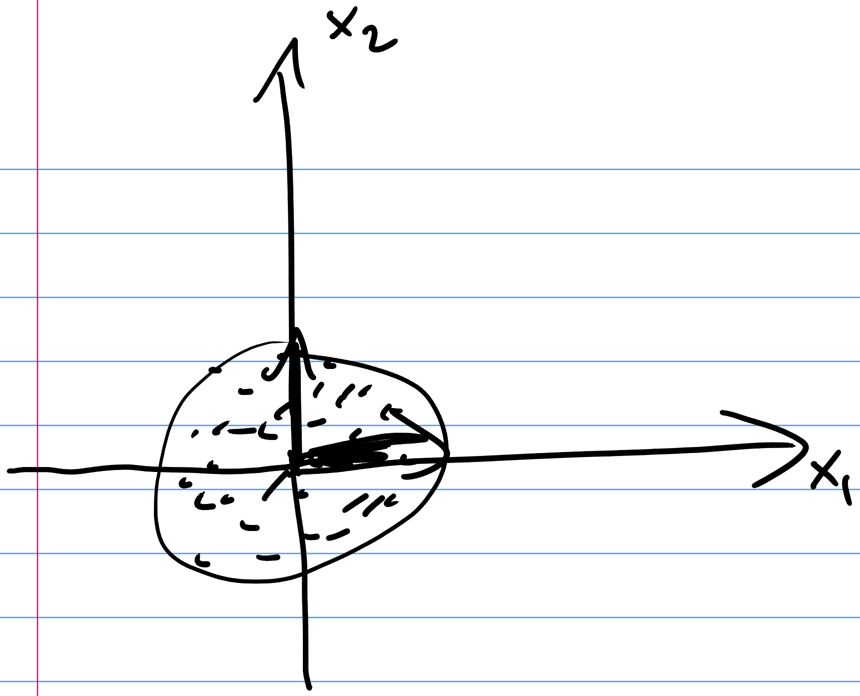
Defn: principal components of X as the eigenvectors of Y^TY (covariance matrix) with "importance weightings" equal to the eigenvalues

Importance here means variance captured by i^{th} component

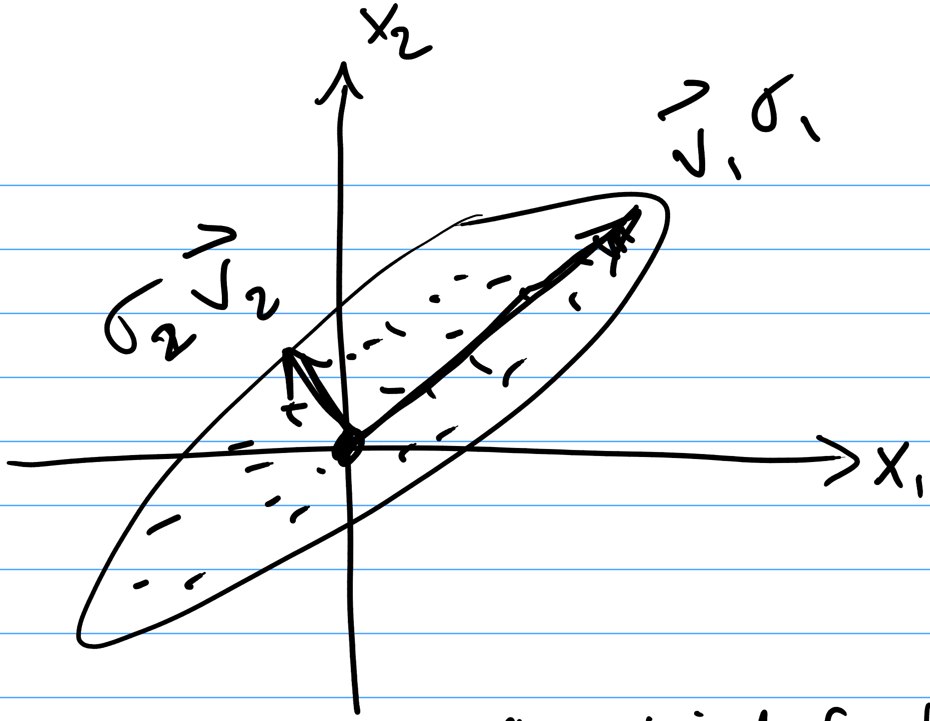


- Subtracting mean = moving origin to center of data
- SVD gives directions of axes of the ellipse $\vec{v}_i$
- Length of axes σ_i

  variance σ_i^2



perfectly spherical data,
any orthonormal
basis V works
(use coordinate axes)



any kind of elliptic
shape to data,
you end up with
unique axes

Fraction of variance explained (σ_i = standard deviation in i^{th} direction)

$$\frac{\sigma_1^2}{\sum_j \sigma_j^2} = \text{fraction of variance explained by } 1^{st} \text{ component}$$

(can replace σ_1 by σ_i for i^{th} component)

$$\frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{j=1}^{\mathrm{rank}(X)} \sigma_j^2} = \text{fraction of variance explained by components } 1, 2, \ldots, r \text{ together}$$