PCA part II
& Least Squares Regression

Mathematical Tools for Neuroscience (NEU 314)
Fall, 2021

lecture 12

Jonathan Pillow

Quiz

Suppose A is a 3 x 3 matrix with SVD: $A = USV^T$

(1) If the A matrix above is rank 2, what does that imply about its singular values?

(2) What is U^TU equal to?

(3) What is the SVD of A^T? (Write it using U, S, and V matrices given above)

(4) What is the SVD of A^TA?

(5) Let $\vec{v}_2$ denote the second right singular vector (i.e., the second row of V^T). What is $V^T\vec{v}_2$?

(hint: this is a 3-component vector).

quick review of PCA

PCA summary

the data

$$X = \overbrace{\begin{bmatrix} - & \vec{x}_1 & - \\ - & \vec{x}_2 & - \\ & \vdots & \\ - & \vec{x}_N & - \end{bmatrix}}^{d} \left.\vphantom{\begin{bmatrix} a \\ a \\ a \\ a \end{bmatrix}}\right\} N$$

2nd moment matrix

$$C = X^\top X$$

SVD

$$C = USU^\top$$

first k PCs: $\{u_1, \ldots u_k\}$

sum of squares of data within subspace: $s_1 + \cdots + s_k$

PCA summary

the data

$$X = \overbrace{\begin{bmatrix} \text{—}\ \vec{x}_1\ \text{—} \\ \text{—}\ \vec{x}_2\ \text{—} \\ \vdots \\ \text{—}\ \vec{x}_N\ \text{—} \end{bmatrix}}^{d} \left.\vphantom{\begin{bmatrix}a\\a\\a\\a\end{bmatrix}}\right\} N$$

2nd moment matrix

$$C = X^\top X$$

SVD

$$C = USU^\top$$

first k PCs: $\{u_1, \ldots u_k\}$

fraction of sum of squares: $\dfrac{s_1 + \cdots + s_k}{s_1 + \cdots + s_N}$

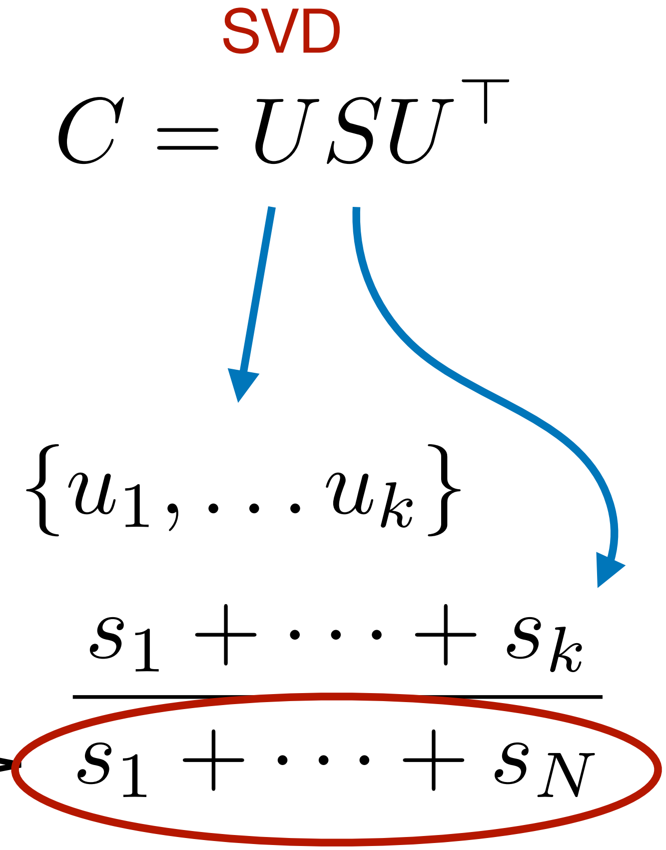
PCA summary

the data

$$X = \left.\overbrace{\begin{bmatrix} \text{---}\ \vec{x}_1\ \text{---} \\ \text{---}\ \vec{x}_2\ \text{---} \\ \vdots \\ \text{---}\ \vec{x}_N\ \text{---} \end{bmatrix}}^{d}\right\} N$$

2nd moment matrix

$$C = X^\top X$$



SVD

$$C = USU^\top$$

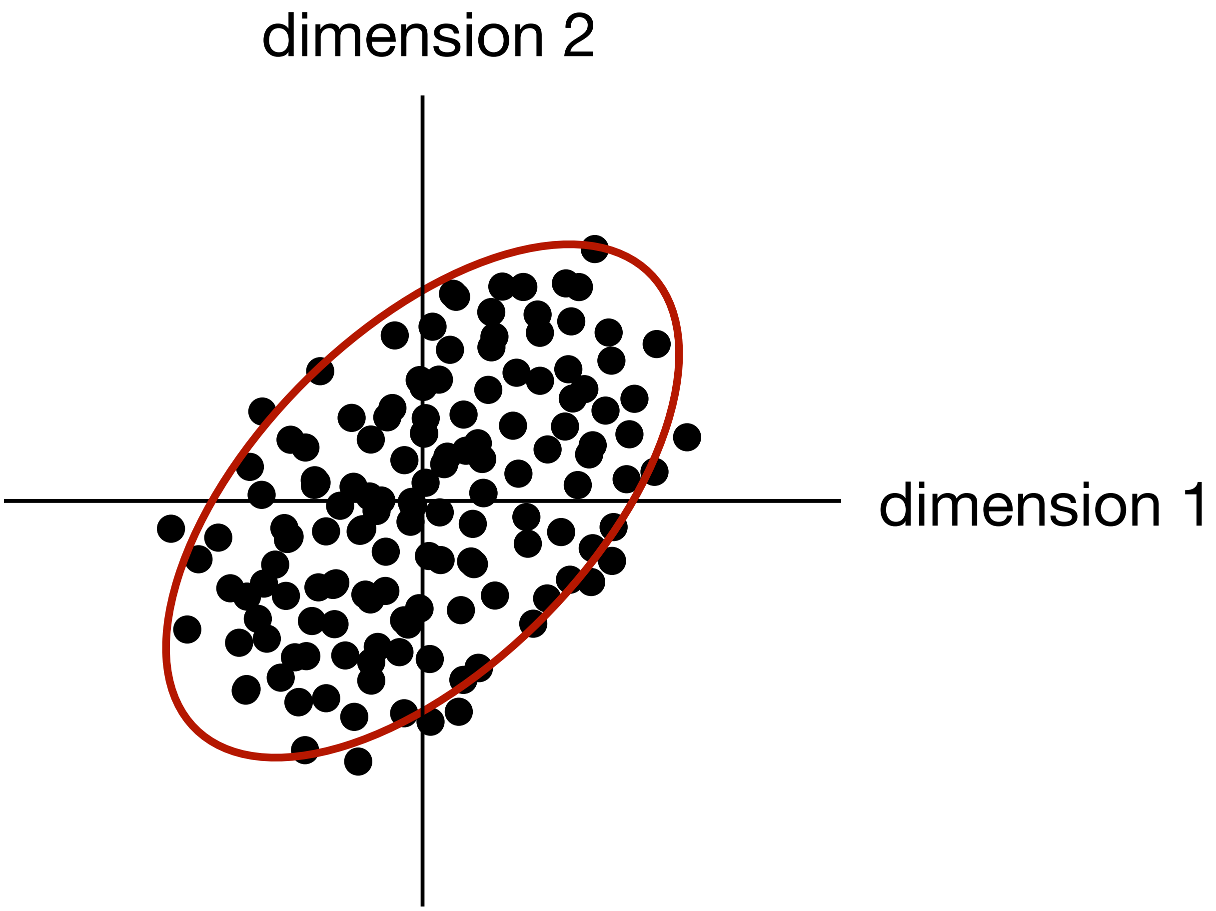
first k PCs: $\{u_1, \ldots u_k\}$

$$\frac{s_1 + \cdots + s_k}{s_1 + \cdots + s_N}$$

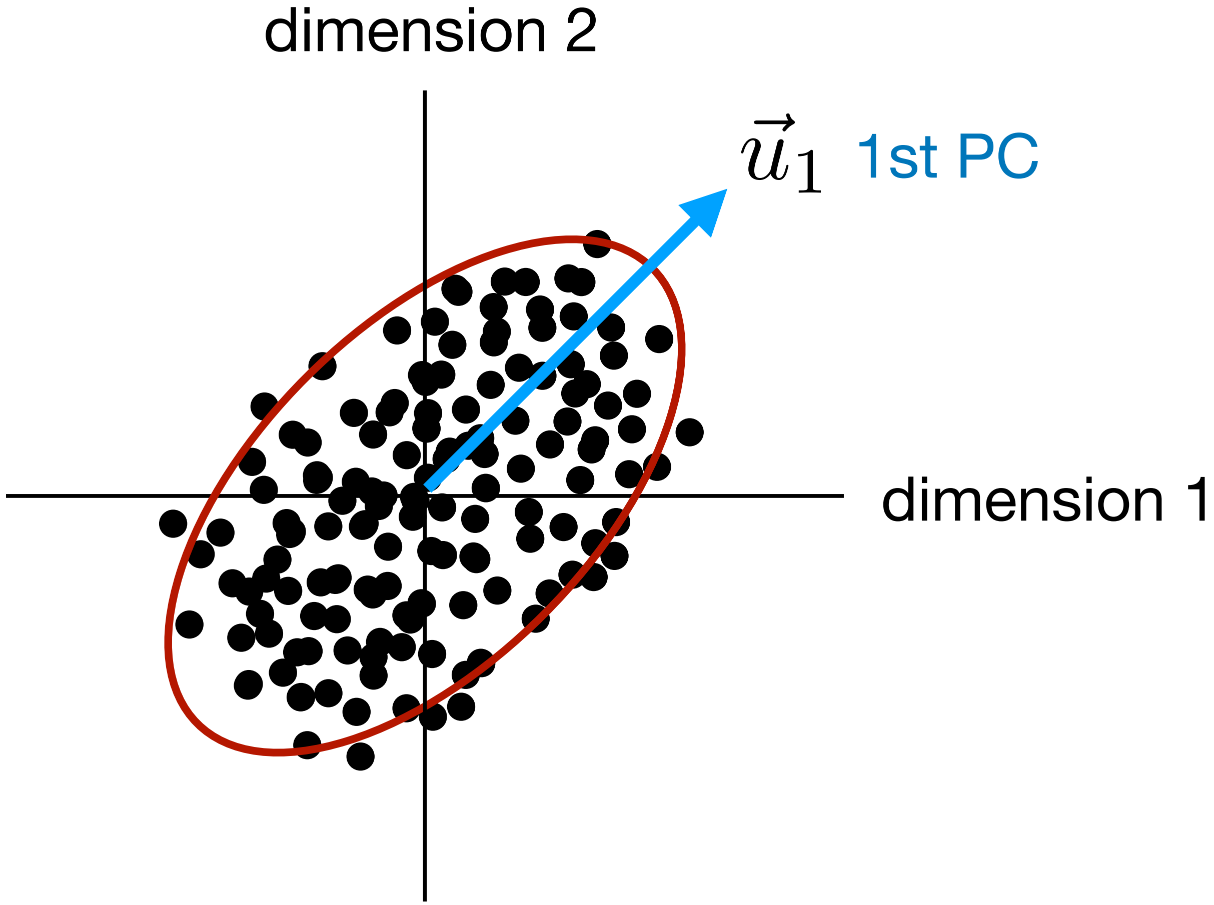
$$||C||_F^2 = \sum_{i=1}^{N} ||\mathbf{x}_i||^2 = \sum_{i,j} x_{ij}^2$$

sum of squares of all data

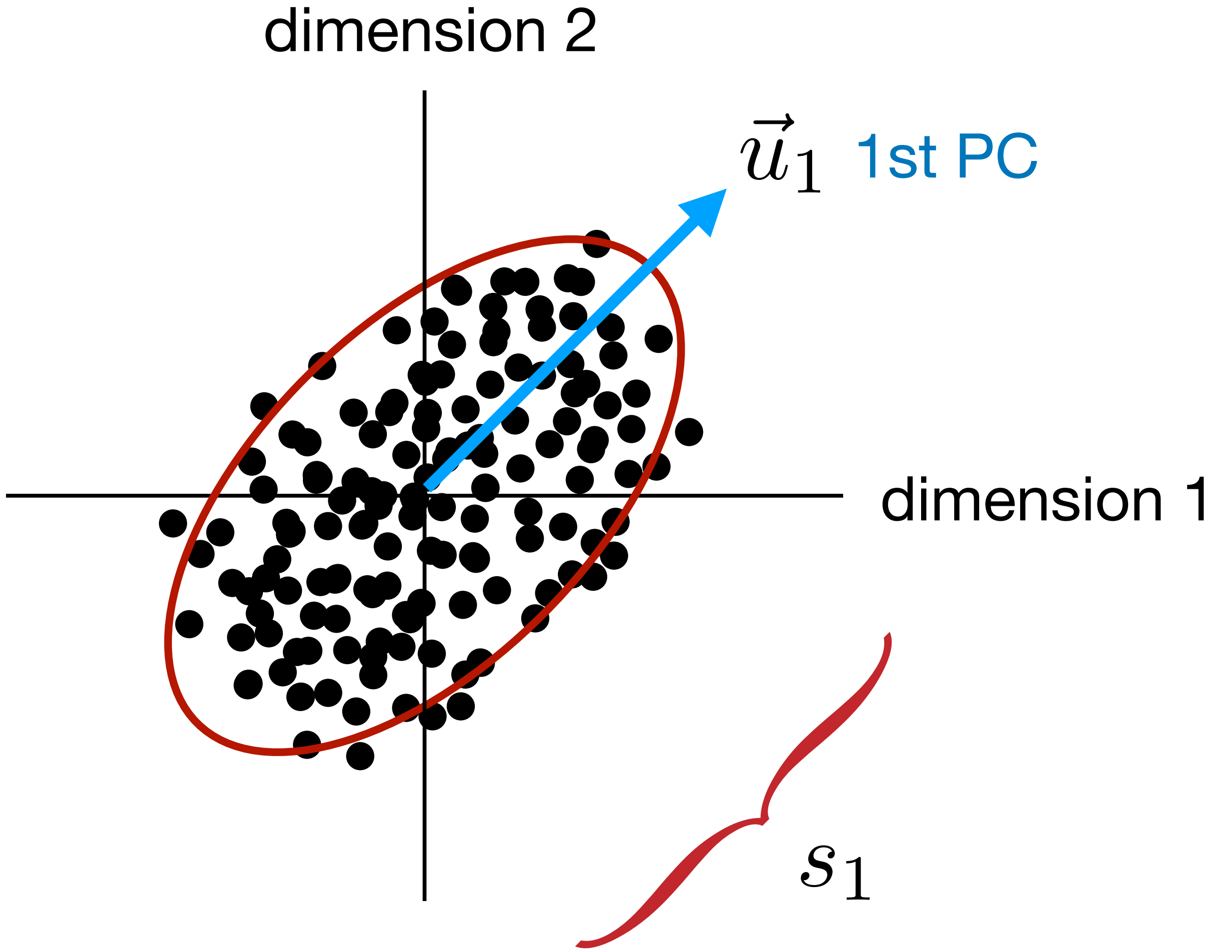
PCA is equivalent to fitting an ellipse to your data



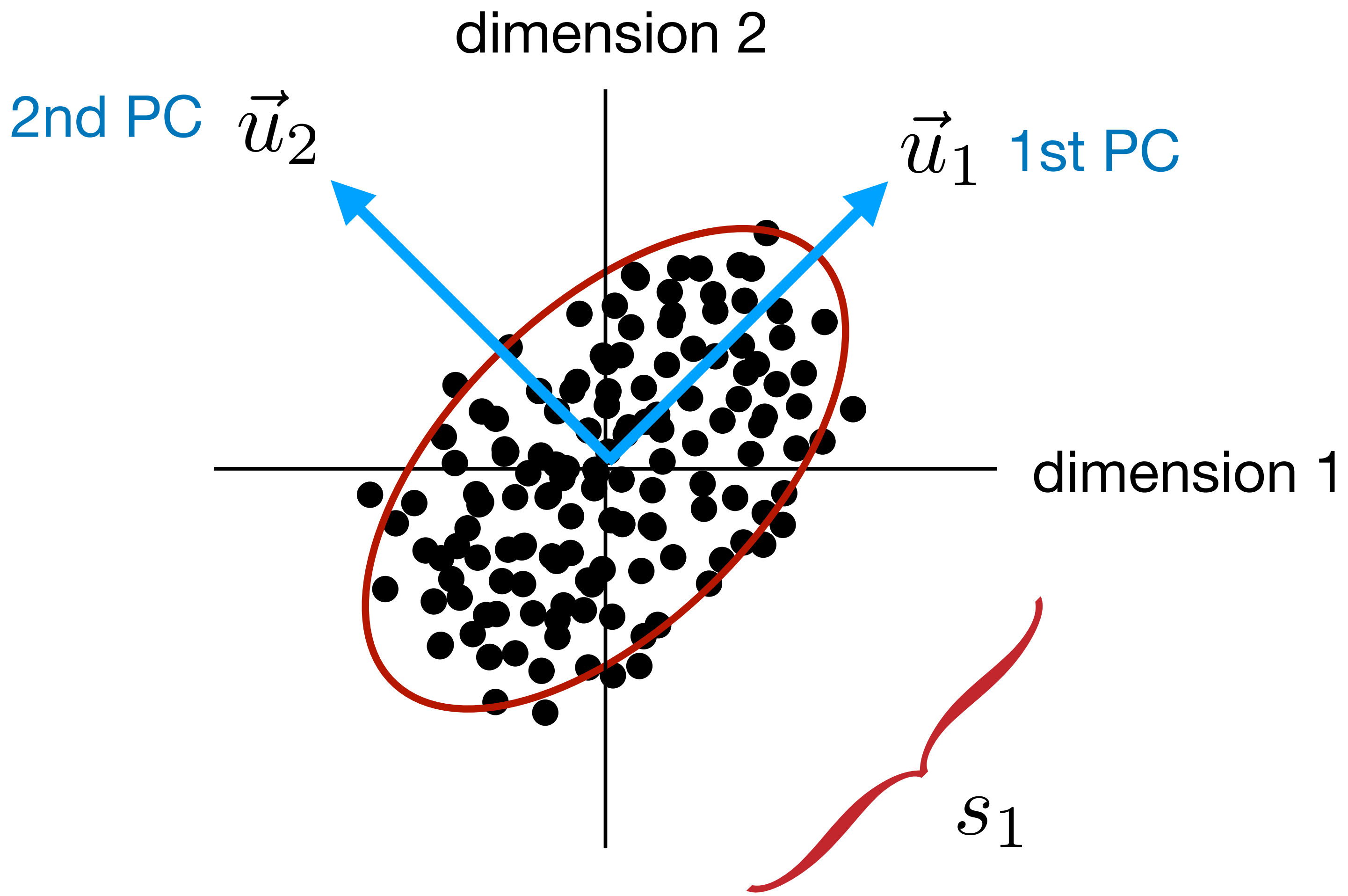
PCA is equivalent to fitting an ellipse to your data



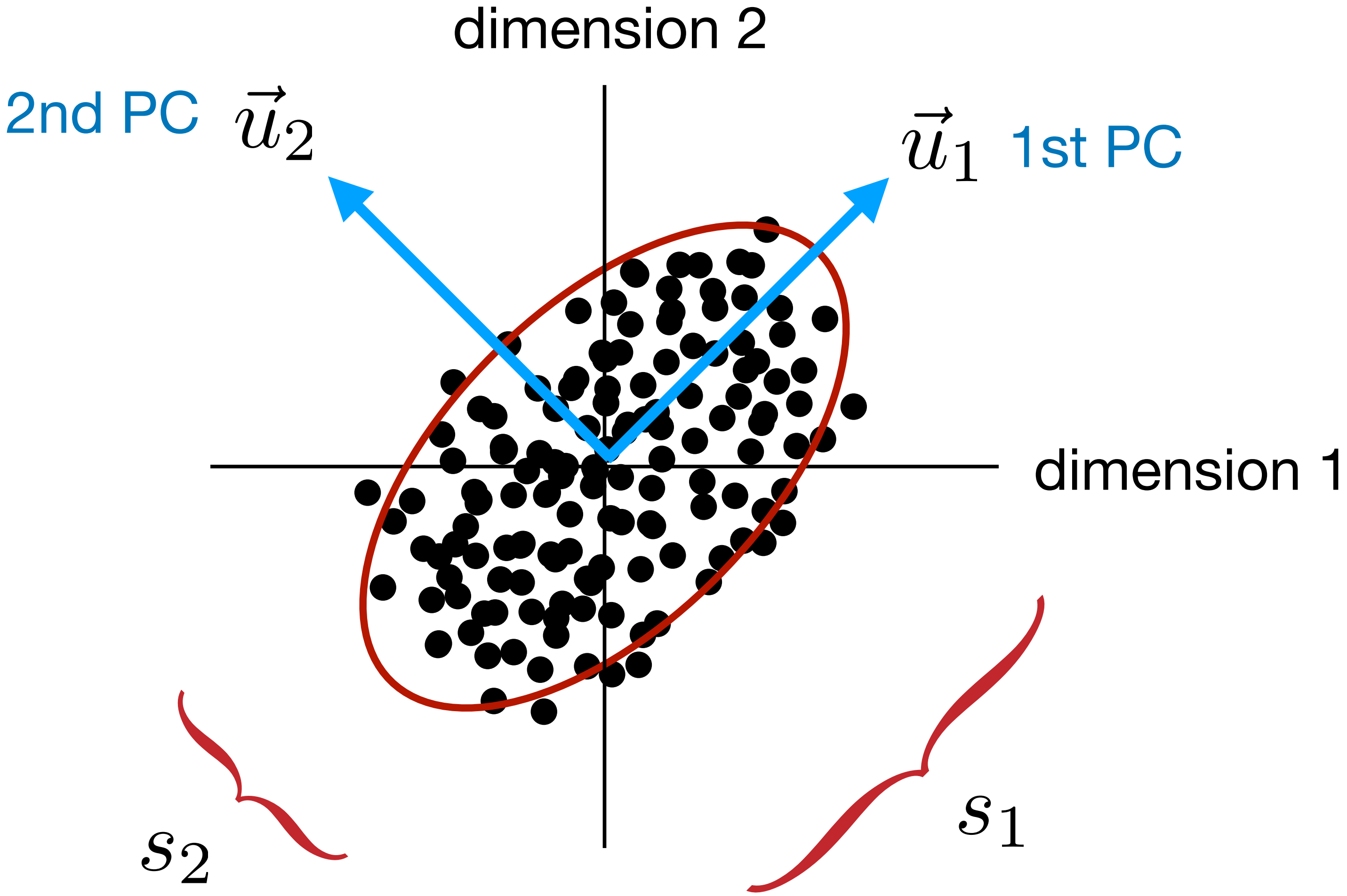
PCA is equivalent to fitting an ellipse to your data



PCA is equivalent to fitting an ellipse to your data

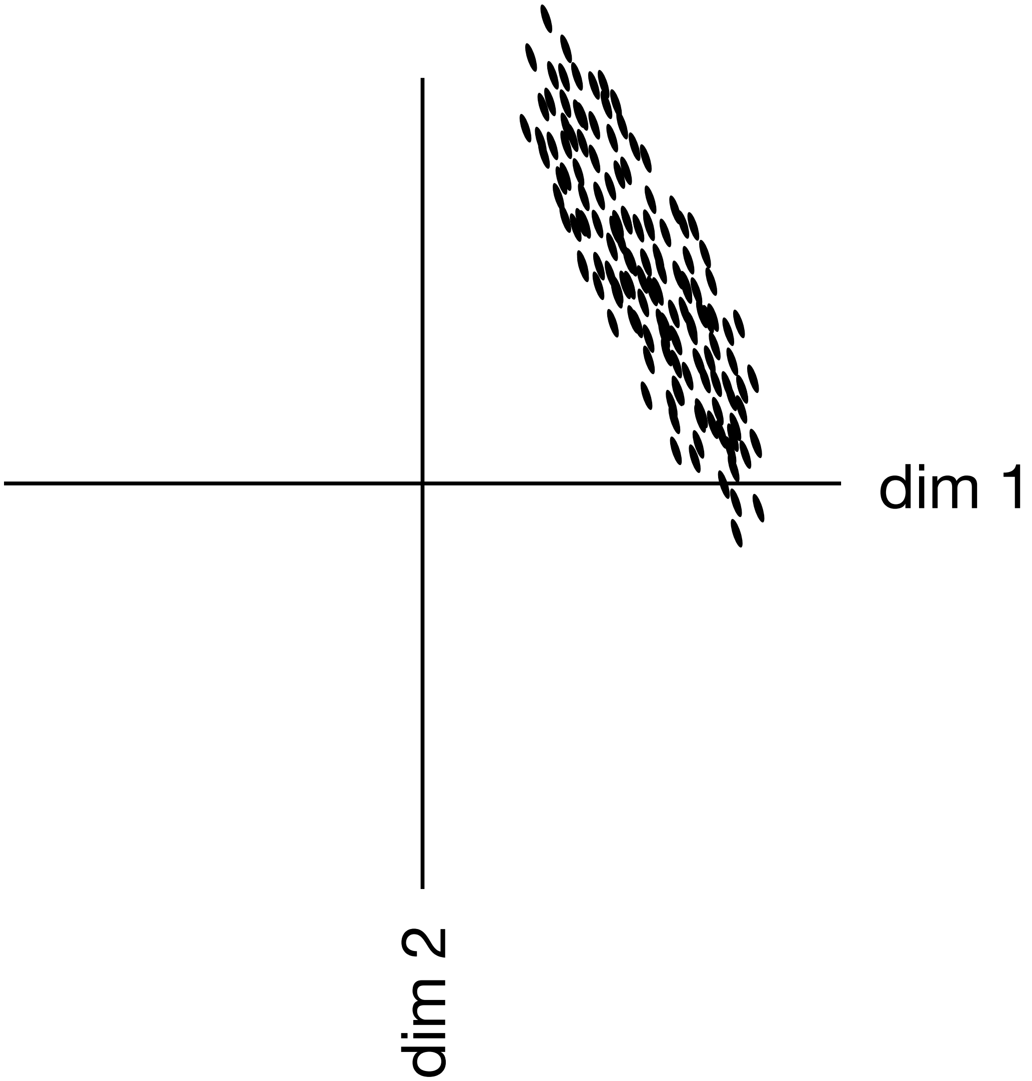


PCA is equivalent to fitting an ellipse to your data

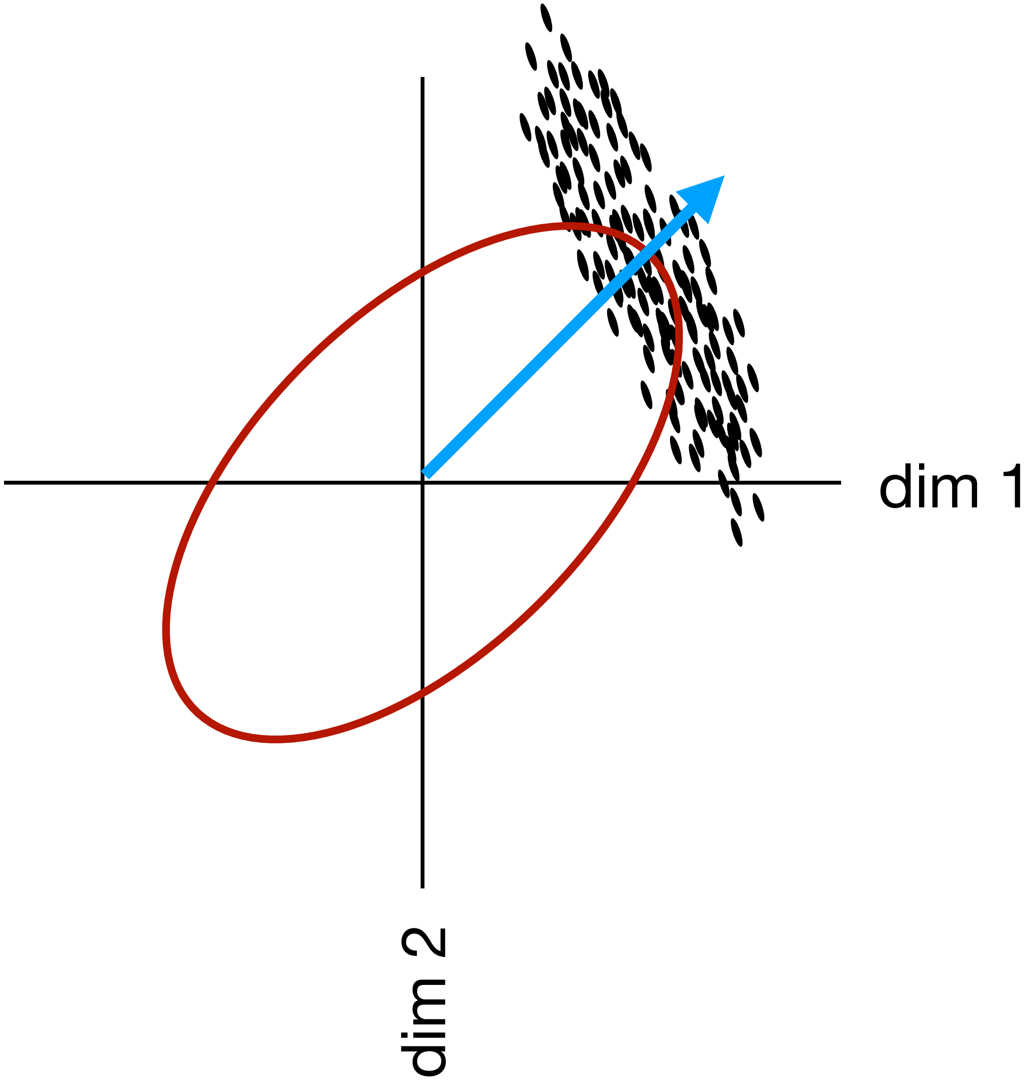


- PCs are major axes of ellipse (or “ellipsoid”)
- singular values specify lengths of axes

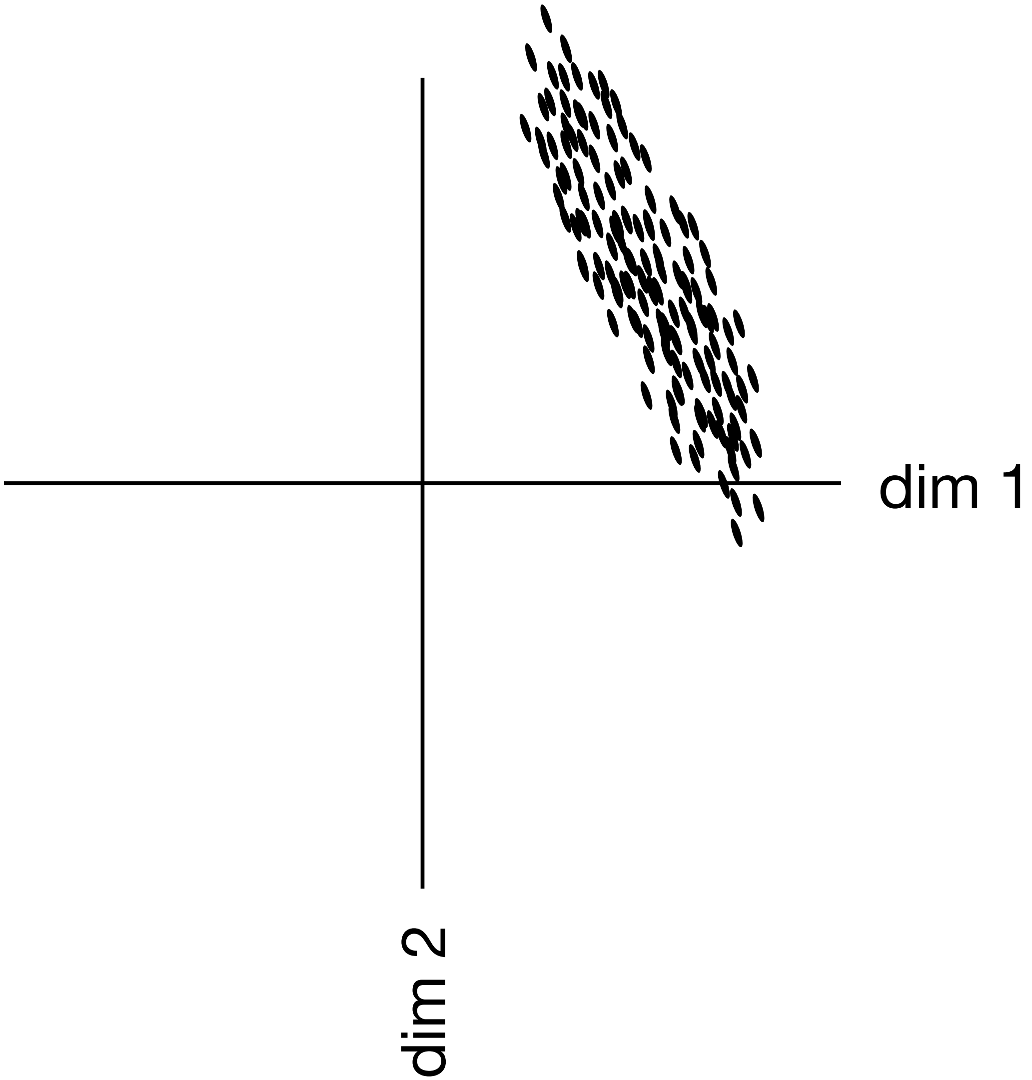
what is the top singular vector of $X^\top X$?



what is the top singular vector of $X^\top X$?

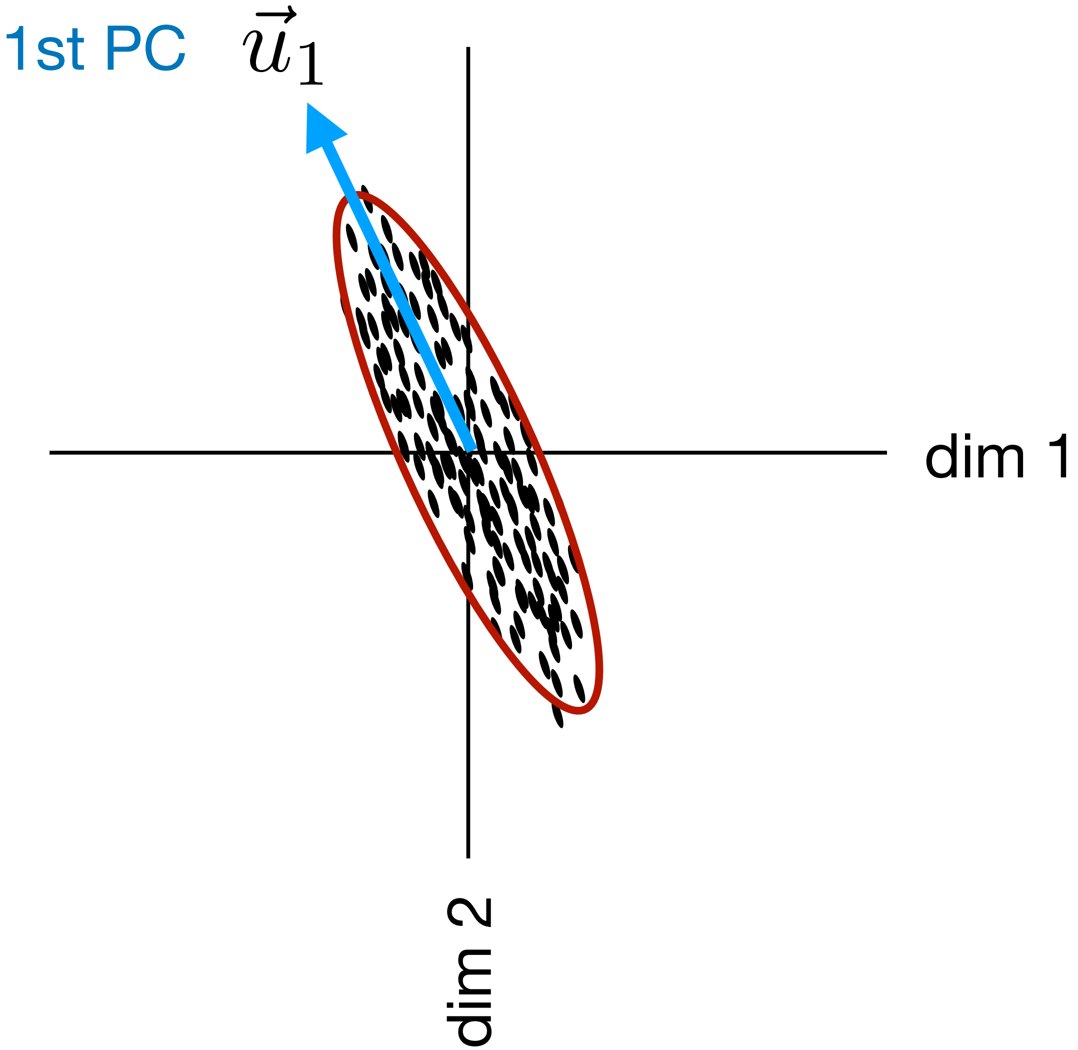


Centering the data $\vec{x}_i - \bar{x}$



Centering the data $\vec{x}_i - \bar{x}$

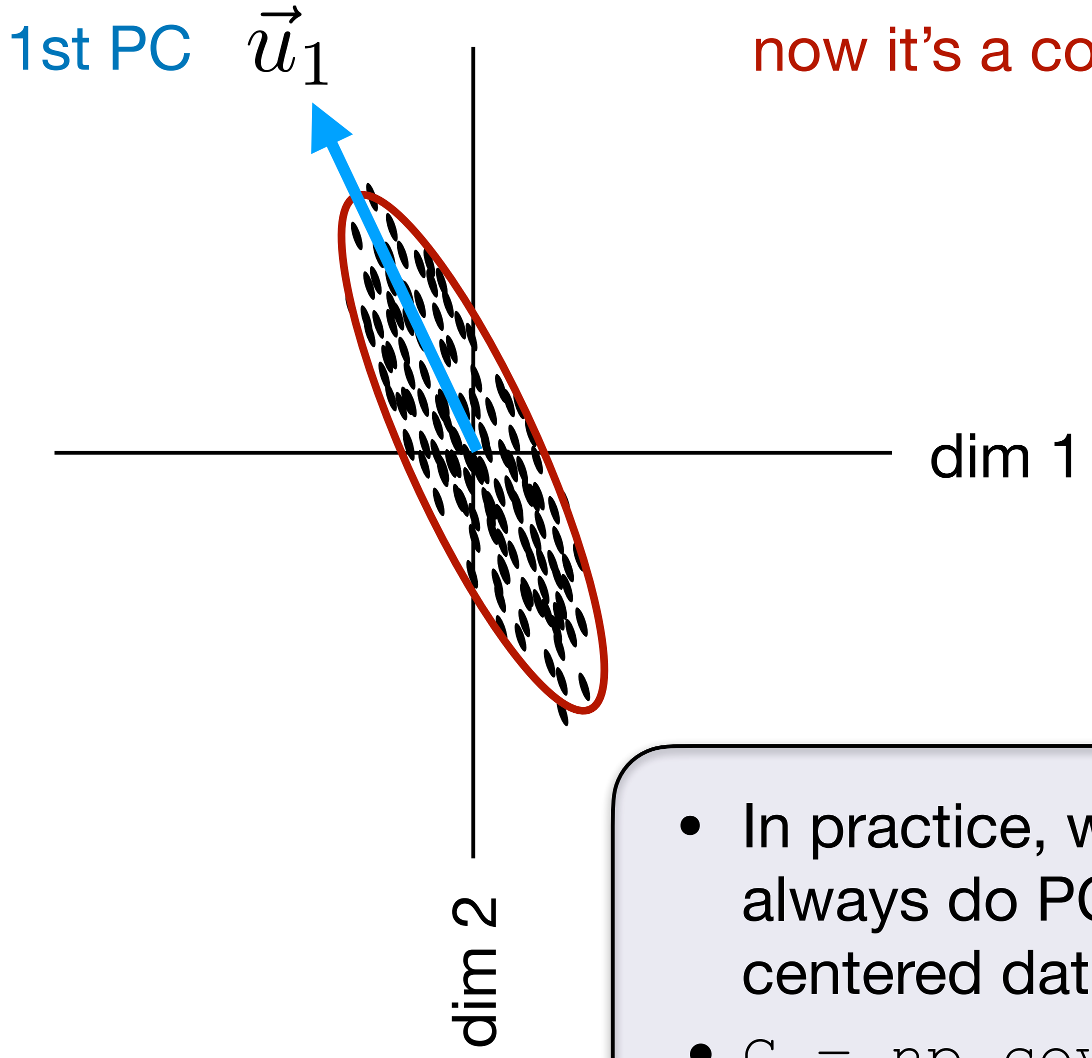
$$C = (X - \bar{x})^\top (X - \bar{x})$$



Centering the data $\vec{x}_i - \bar{x}$

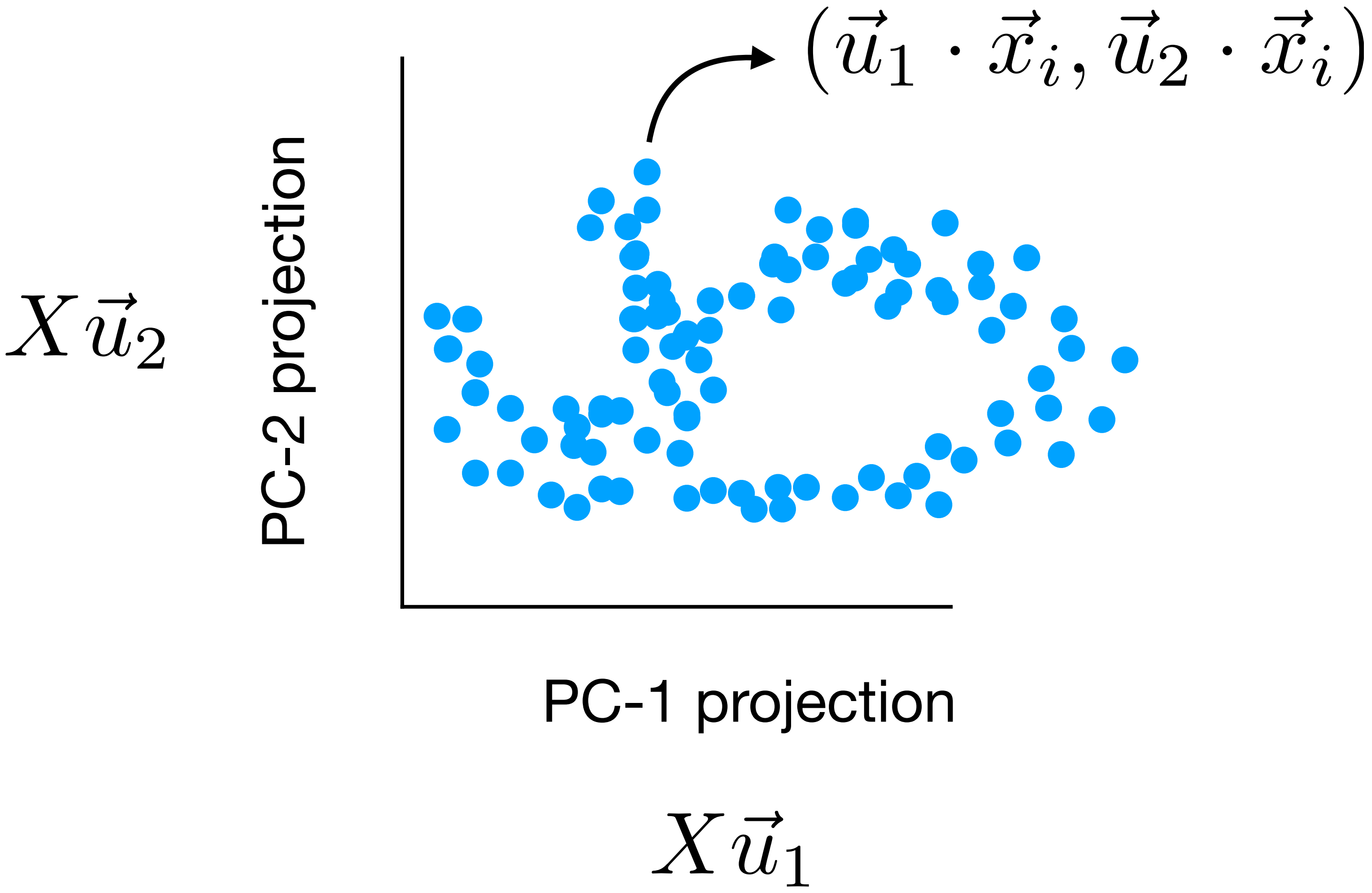
$$C = \frac{1}{N}(X - \bar{x})^\top (X - \bar{x})$$

now it’s a covariance!



- In practice, we almost always do PCA on centered data!
- `C = np.cov(X)`

Projecting onto the PCs



- visualize low-dimensional projection that captures most variance

Full derivation of PCA: see notes

Two equivalent formulations:

1. $$\hat{B}_{pca} = \arg\max_{B} ||XB||_F^2$$

$$\text{such that } B^\top B = I$$

find subspace that preserves *maximal* sum-of-squares

Full derivation of PCA: see notes

Two equivalent formulations:

1. $$\hat{B}_{pca} = \arg\max_{B} ||XB||_F^2$$
$$\text{such that } B^\top B = I$$

find subspace that preserves *maximal* sum-of-squares

2. $$\hat{B}_{pca} = \arg\min_{B} ||X - \underbrace{XBB^\top}||_F^2$$
$$\text{such that } B^\top B = I$$

minimize sum-of-squares of orthogonal component

reconstruction of X in subspace spanned by B

Summary

- PCA recap
- PCA = fitting an ellipse to data.
  (PCs = major axes of ellipse;
  singular values = amount of variance captured by each PC)
- Centered vs. non-centered PCA
- Plotting data projected onto PCs.

Least Squares regression
also known as
Ordinary Least Squares (OLS)

(on board)