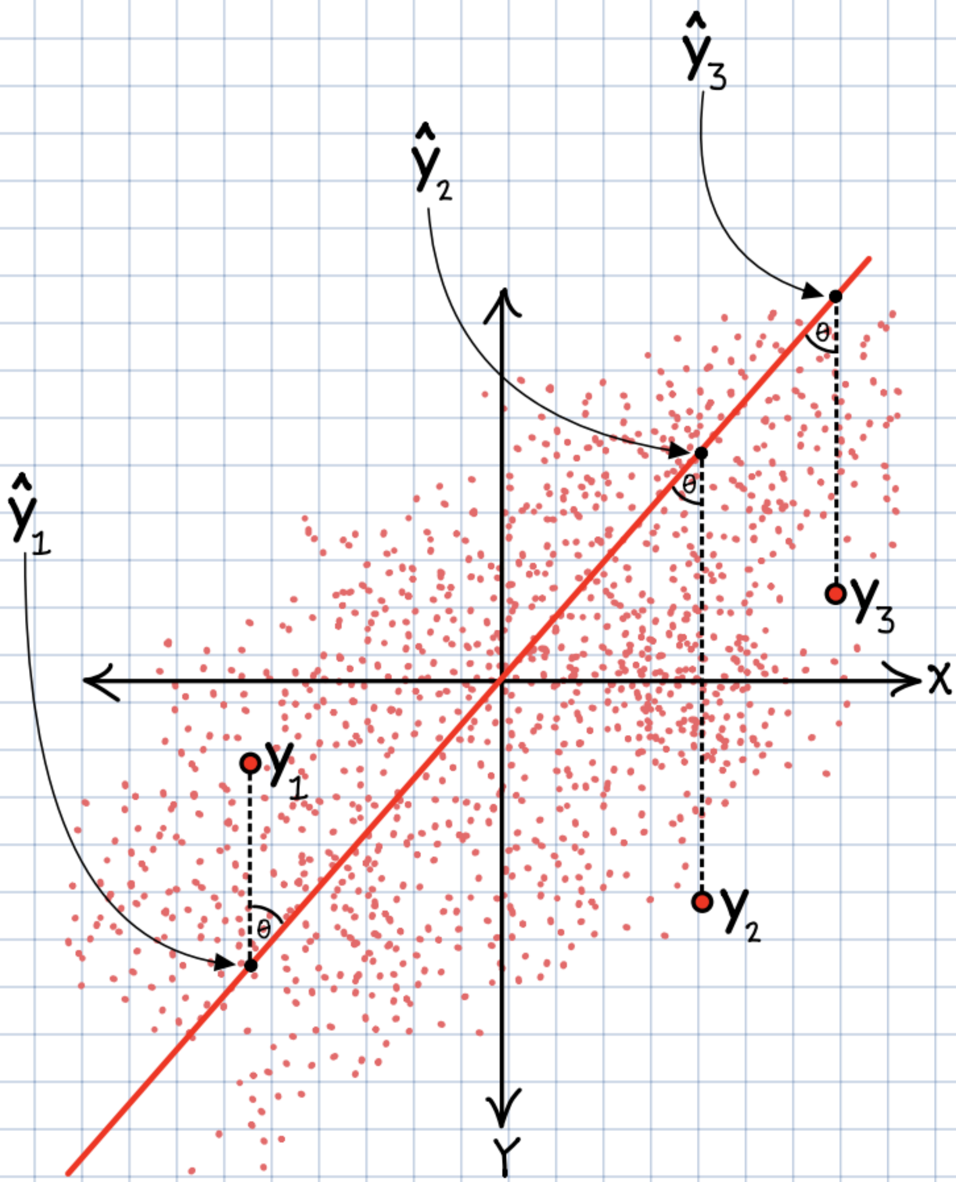
OLS REGRESSION

objective function: $\min \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$



PRINCIPAL COMPONENTS

objective function: $\max \vec{h}_1' \Omega \vec{h}_1$ s.t. $\vec{h}_1'\vec{h}_1 = 1$,

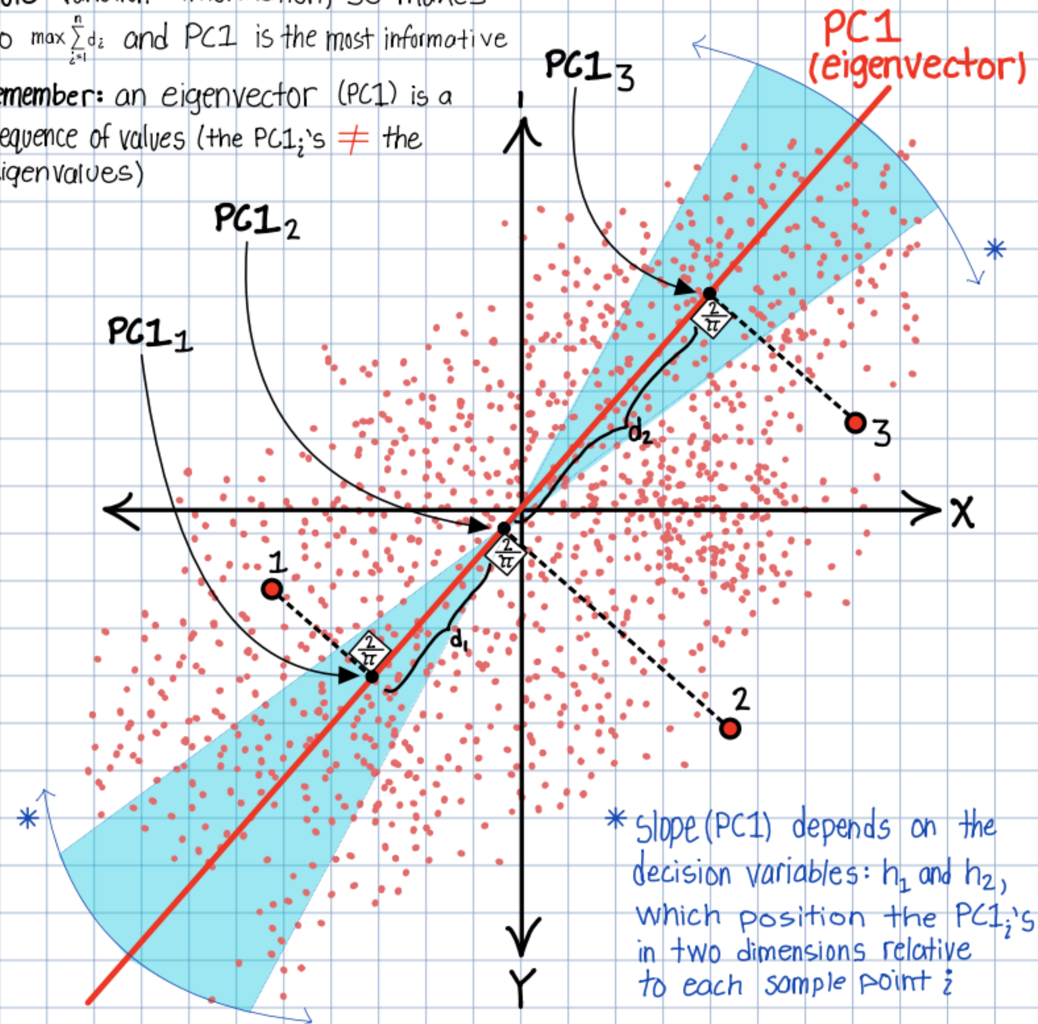
$$h_1 x_1 + h_2 y_1 = PC1_{i=1}$$
$$h_1 x_2 + h_2 y_2 = PC1_{i=2}$$
$$h_1 x_3 + h_2 y_3 = PC1_{i=3}$$

$$\vec{h}_1' \Omega \vec{h}_1 = [h_1\ h_1\ h_1] \begin{bmatrix} & & \\ & \Omega & \\ & & \end{bmatrix} \begin{bmatrix} h_1 \\ h_1 \\ h_1 \end{bmatrix}$$

$\max \vec{h}_1' \Omega \vec{h}_1 \equiv \max \sum_{i=1}^{n} d_i$: $\underbrace{PC1_1 \text{———} PC1_2}_{d_1}\underbrace{\text{————} PC1_3}_{d_2}$

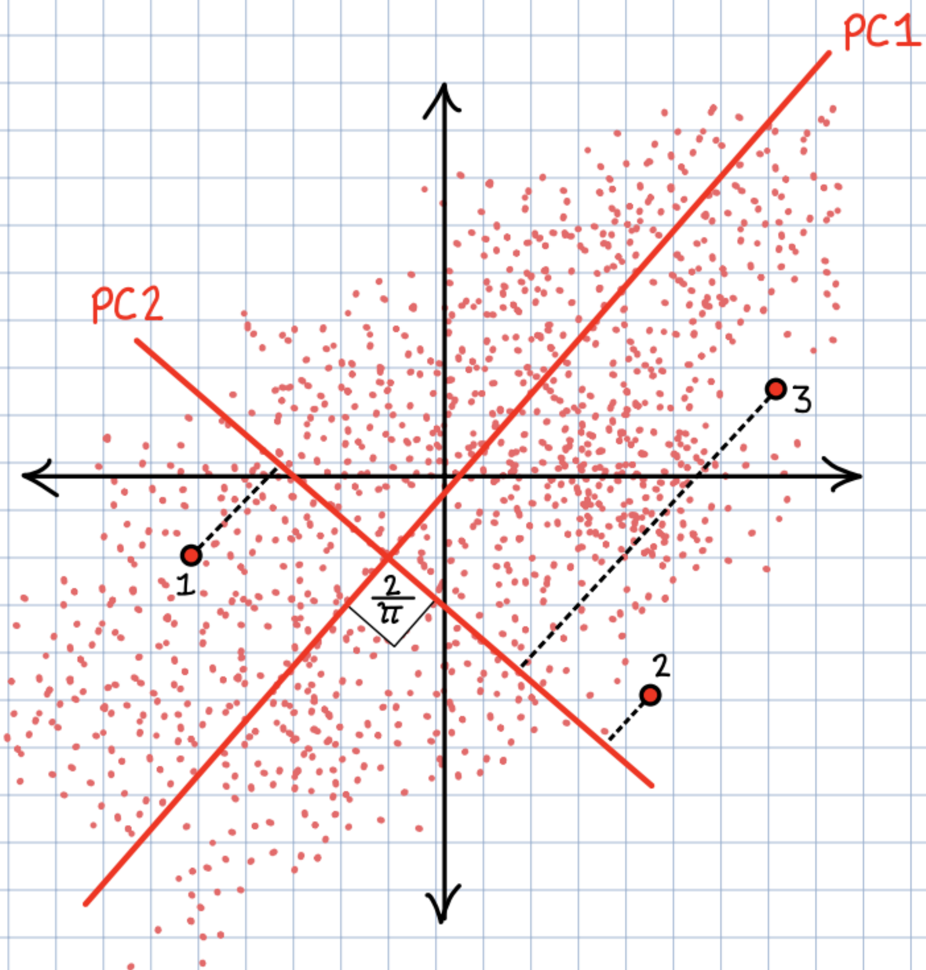
note: variation = information, so makes to $\max \sum_{i=1}^{n} d_i$ and PC1 is the most informative

remember: an eigenvector (PC1) is a sequence of values (the $PC1_i$'s $\neq$ the eigenvalues)



* slope (PC1) depends on the decision variables: h_1 and h_2, which position the $PC1_i$'s in two dimensions relative to each sample point i

the Principal Components algorithm only works in two dimentions, and PC2 must be perpendicular (orthogonal) to PC1:

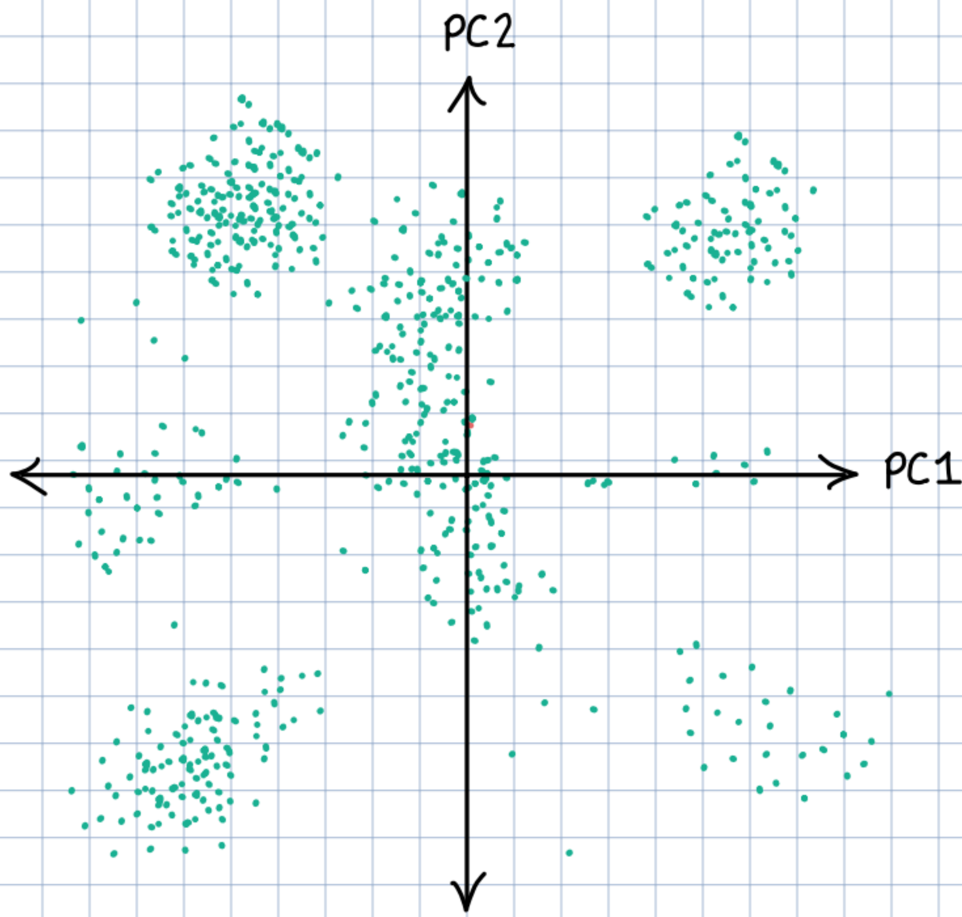


PC2 captures an entirely different set of information than PC2; PC2 gets the "leftover" varition after PC1. The PC2 optimization problem is as follows:

$$\max \vec{h}_2' \boldsymbol{\Omega} \vec{h}_2 \quad \text{s.t.} \quad \vec{h}_2'\vec{h}_2 = 1 \,, \quad \vec{h}_1'\vec{h}_2 \overset{\text{DOT PRODUCT}}{=\!=\!=} \langle \vec{h}_1, \vec{h}_2 \rangle \overset{\text{ORTHOGONAL}}{=\!=\!=} 0$$

PRINCIPAL COMPONENTS

- purpose: dimentionality reduction
- often lends itself to clustering (cluster the PC's):



- you would never fit more PC eigenvectors than underlying variables; should always have at least one more variable than PC eigenvectors

ADDITIONAL PCA CALCULATIONS: THE EIGENDECOMPOSITION OF A MATRIX

PROCESS OF EIGENDECOMPOSITION

step 1 **Standardize** the data via z-score or some other method

- When to standardize? Data is not on the same scale.
- When <u>NOT</u> to standardize? Data is already on the same scale. e.g., a correlation matrix; binary/dummy variables; prices all in the same currency; temperatures all in fahrenheit.

W/o standardization, the PCA will emphasize variables w/ the largest <u>absolute</u> variances not <u>relative</u> variances.

step 2 When you perform eigendecomposition on a set of (hopefully standardized) data—which should be expressed as a matrix— you decompose the matrix into eigenvectors & eigenvalues.

- **eigenvalues:** scalars that indicate the magnitude of the eigenvectors.
- **eigenvectors:** vectors (i.e., a sequence of numbers) that define the directions of maximum variance in each data. Each eigenvector has an eigenvalue and is orthogonal (perpendicular; independent) to all other eigenvectors.

Eigenvalues can give you a sense of how many variables/factors are necessary to capture most of the variability in your data. This is particuarly useful in feature selection and dimensionality reduction.

step 3 Represented mathematically, step 2 looks like this: $A = PDP^{-1}$, where P is a matrix whose columns are the eigenvectors of A, and D is a diagonal matrix whose diagonal elements are the eigenvalues of A. To find the eigenvalues of A, you solve the characteristic equation $\det(A - \lambda I) = 0$ for λ. The proportion of variation by PC1 $= \dfrac{\text{eigenvalue of PC1}}{\text{sum of all eigenvalues}} = \dfrac{\lambda_1}{\sum_i \lambda_i}$