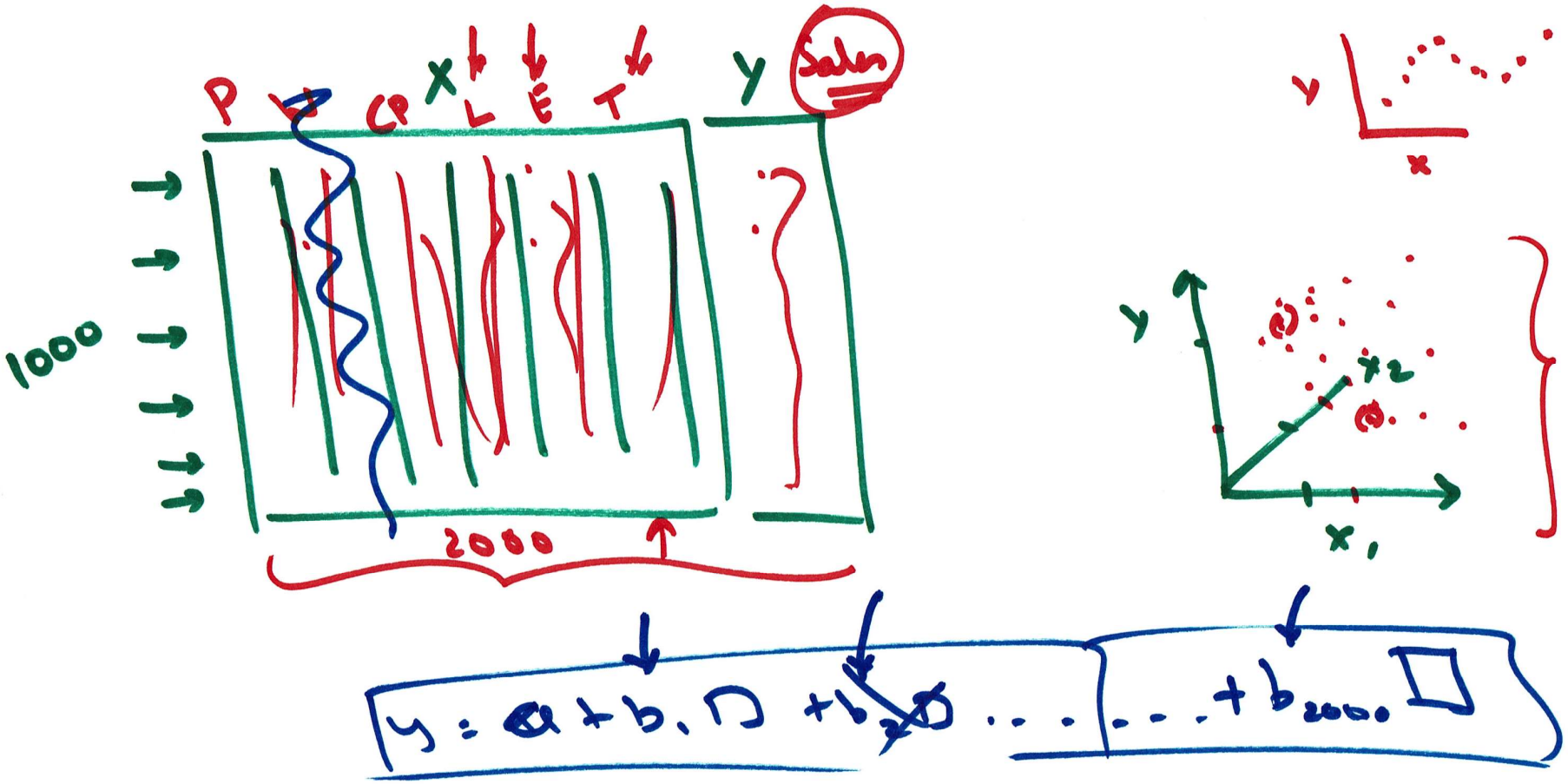
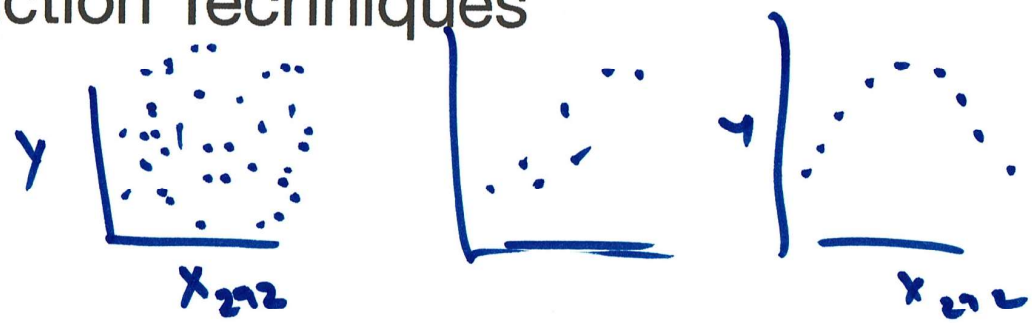


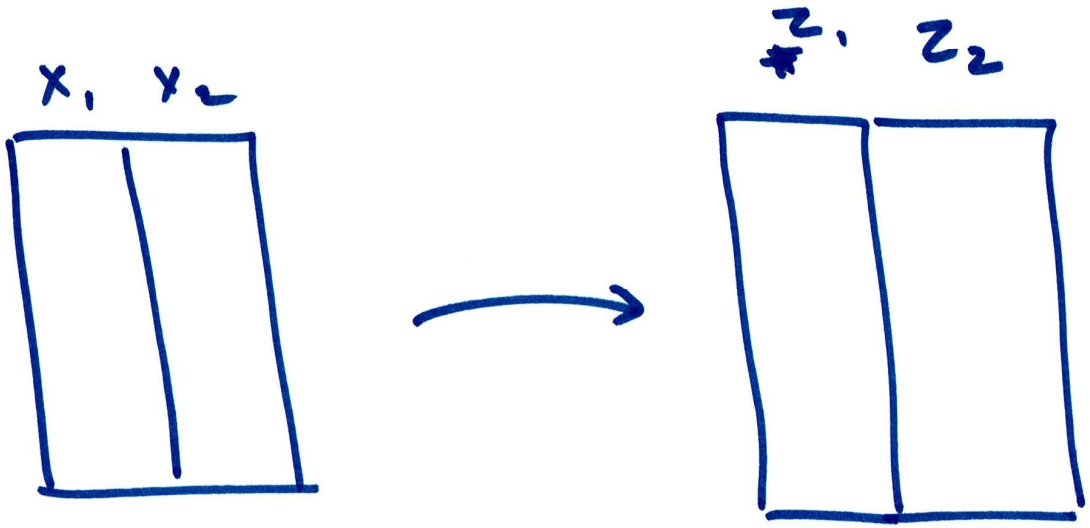
This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Dim. Reduction Techniques

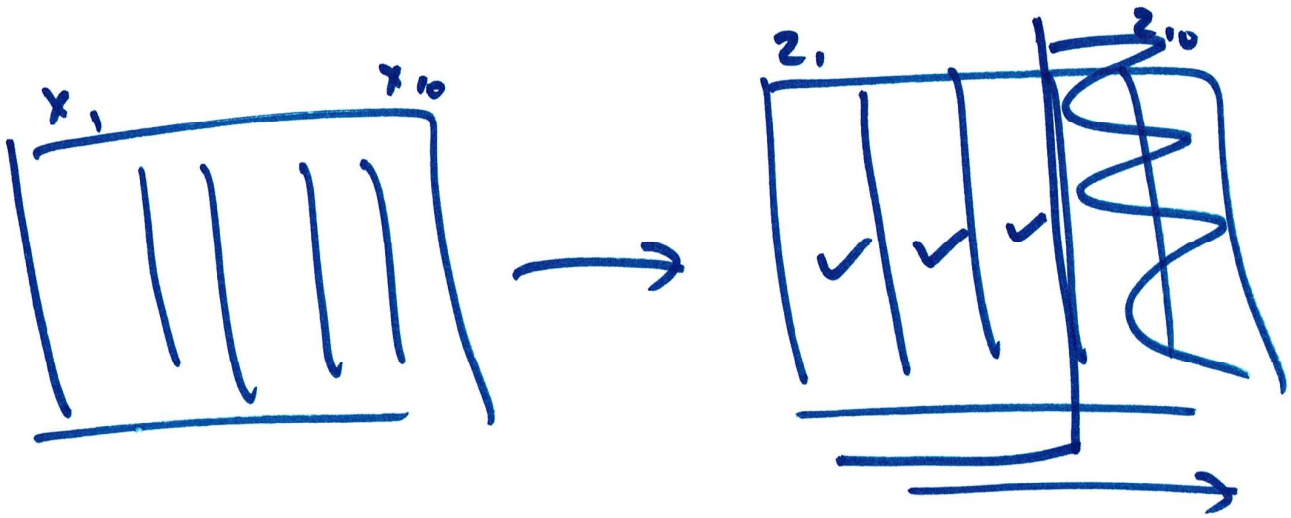


- Feature elimination
  - Simply identify and remove variables (columns) that are not important
  - The disadvantage is that we would gain no insight from those dropped variables and loose any information they contain
- Feature extraction
  - Create a few new variables from the old variables
  - **PCA** Principal Component Analysis: is the most popular feature extraction technique (linear)
  - t-SNE (non-linear)

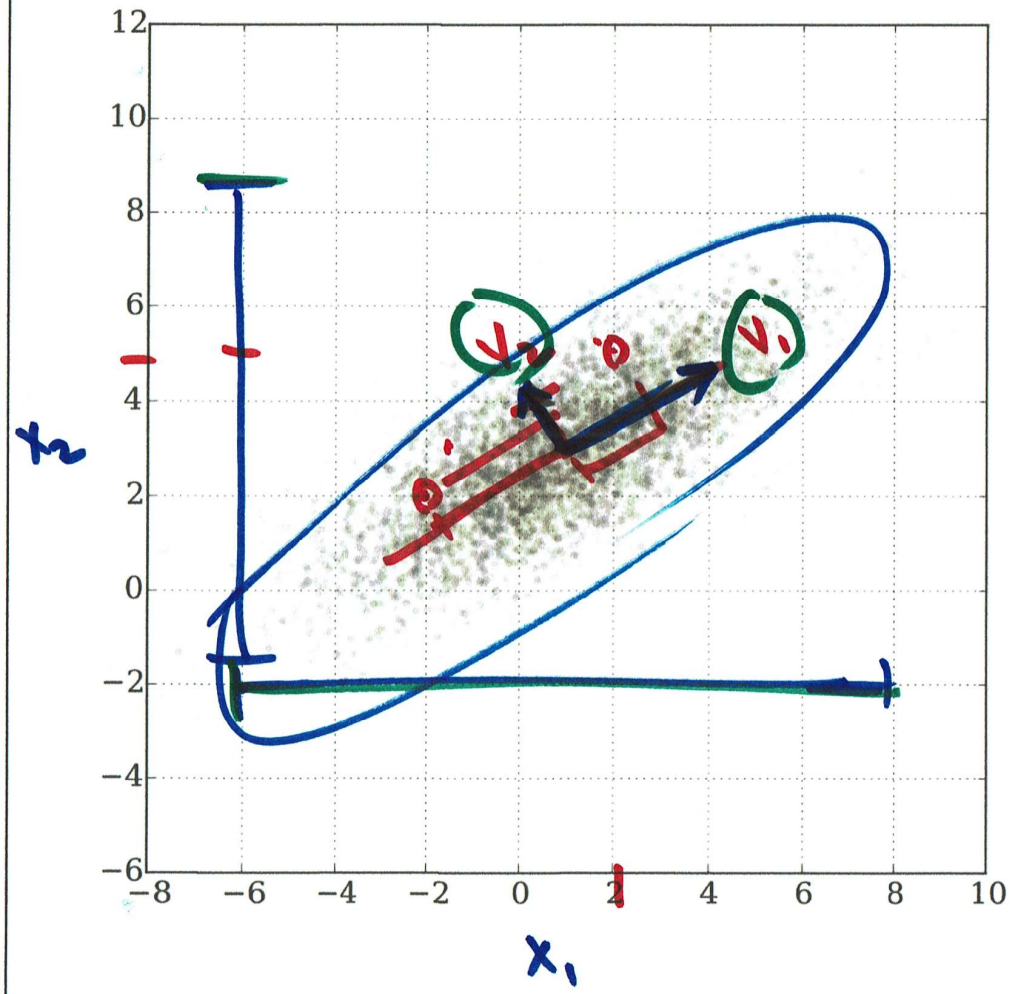
This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

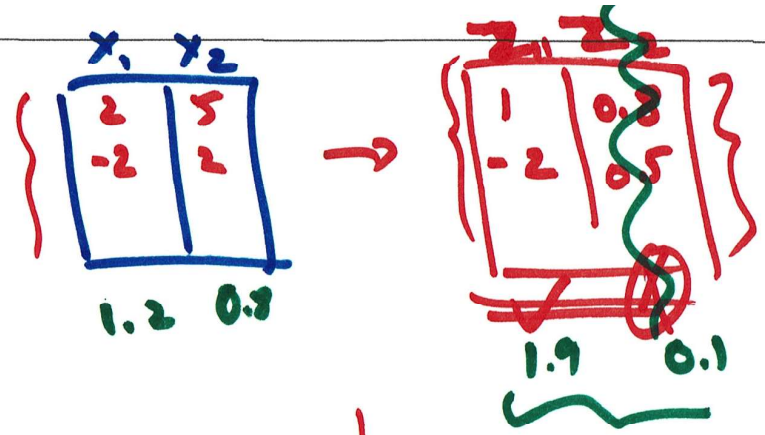


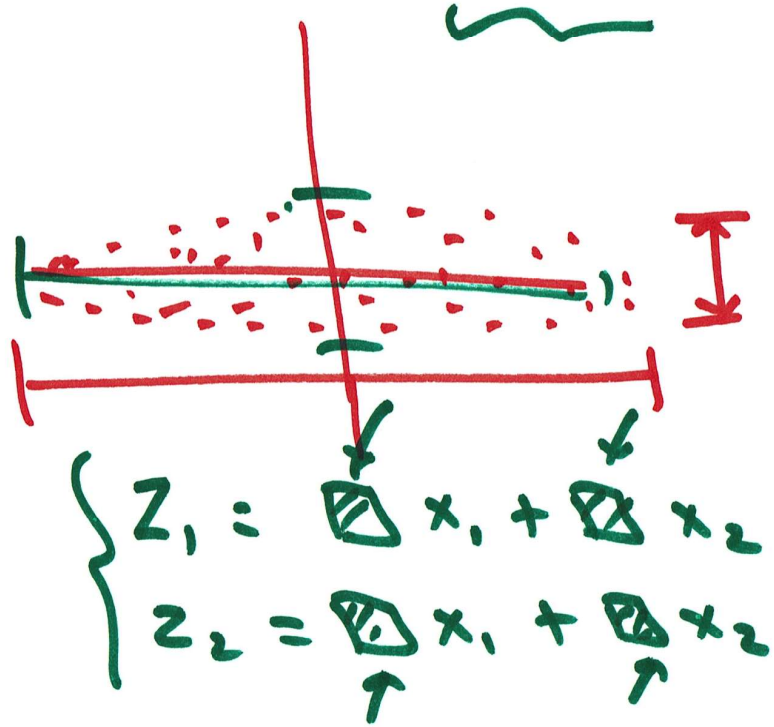
This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.



This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.



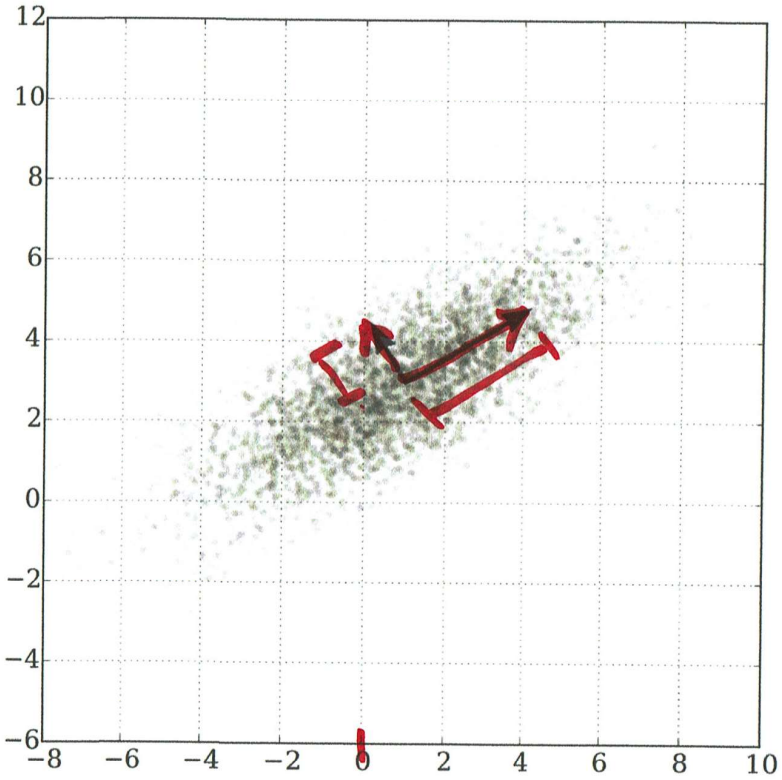




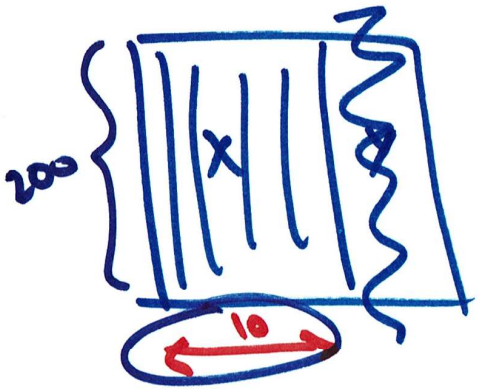
This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

PCA

- creates new variables using linear combinations of old variables
- is designed to create variables that are independent of one another
- also manages to tell us how important each of these new variables are
- this "importance", helps us to choose how many variables we will use



This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.



$$X_{new} = \frac{X - \mu}{\sigma}$$

$$Cov = Cov(X_{new})$$

Eigen decomposition

$e_1, e_2, \ldots e_{10}$

$v_1, v_2 \ldots v_{10}$

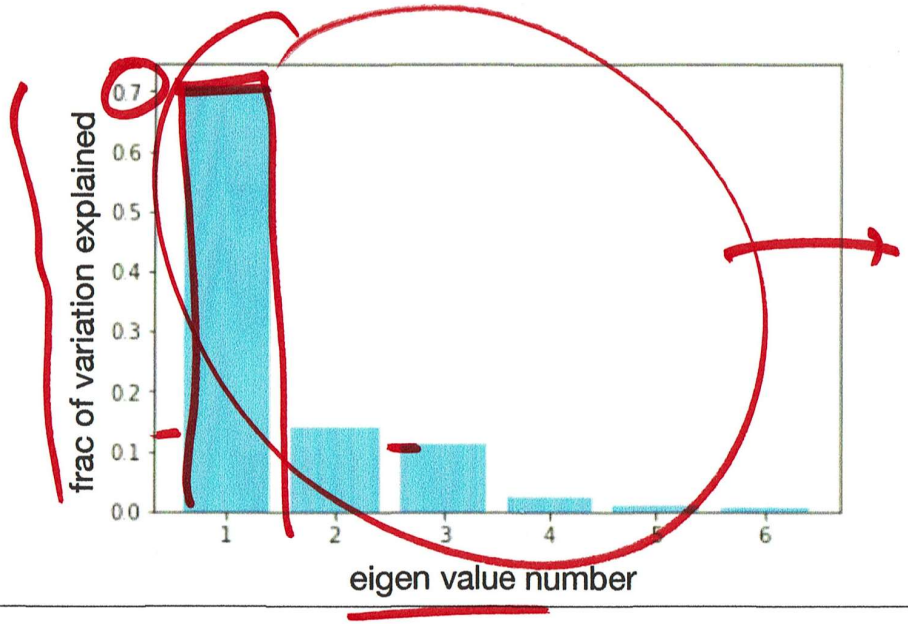
Z

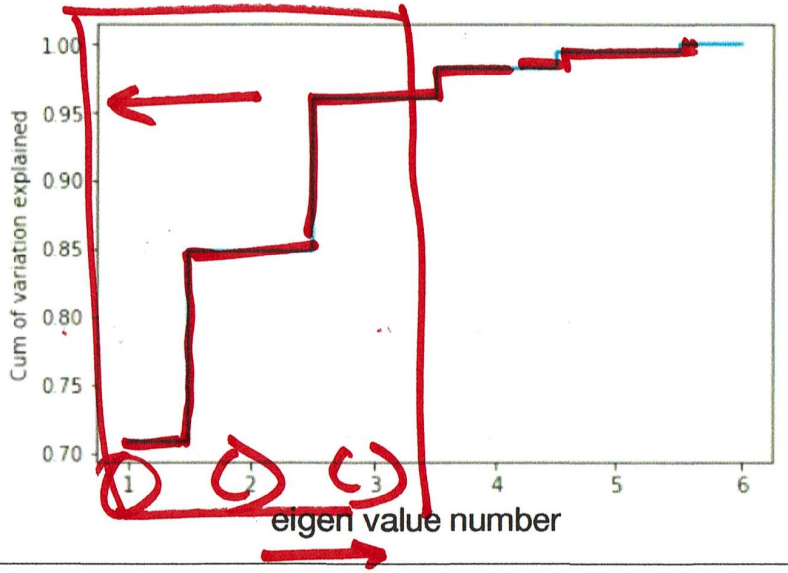
10 x 3

$$W = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}_{10 \times 3}$$

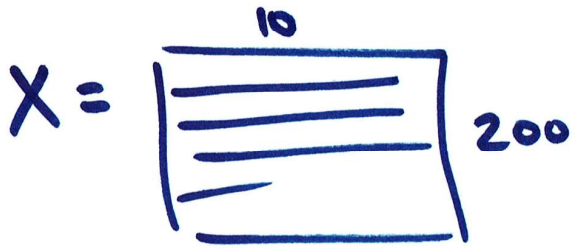
This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

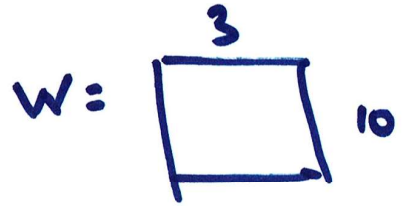
- Scale the data and compute the covariance matrix
- Break the covariance matrix into magnitude and direction. Eigen Vectors and the Eigen Values of the covariance matrix can be thought of as the natural axis/directions and magnitudes along those axis, of the data
  - The eigen values also can be used to calculate the percentage of variation explained by each component
- Sort in the eigen values in desending order and calculate the cumulative percentage of variation explained
- Pick the number of principal components you will use
- Transform to new variables

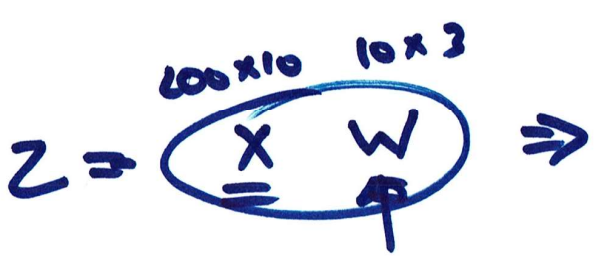


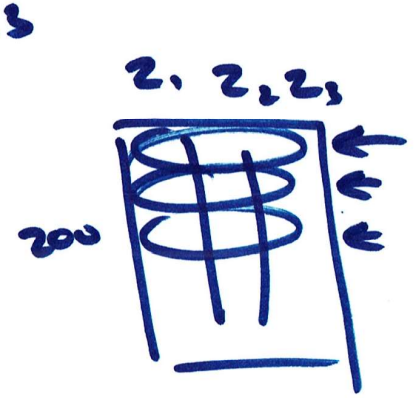


This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.









$S = 100 + 20 z_1 + 15 z_2 - 10 z_3$

$z_1 = \square x_1 + \square x_2 + \square x_3 + \ldots + x_{10}$

$z_2 = \ldots$

$z_3 = \ldots$

This file is meant for personal use by arielighuma4@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.