

WOMEN WHO

SSID: WWCode

Password: password





WOMEN WHO
CODE
MANILA



Machine Learning and AI Study Group

Twitter: @wwcodemanila
FB: fb.com/wwcodemanila

#WWCodeManila
#WWCode.ML.AI
#StudyGroup



Issa Tingzon

Research Fellow II
PCARI



OUR MISSION

Inspiring women to excel in technology careers.



OUR VISION

A world where women are representative as technical executives, founders, VCs, board members and software engineers.



STUDY GROUP

Study groups are events where women can come together and help each other learn and understand a specific programming language, technology, or anything related to coding or engineering.

GUIDELINES

- Study groups are **not** a class/lecture
- If you have a question, just **ask**
- If you have an idea, **share it**
- **Make friends** and learn from your study groupmates
- **Do not** recruit or promote your business

WOMEN WHO

New Member's Introduction

CODE[®]



I am <name>

<your current profession>

<why did you join this study group?>

<background in machine learning and AI>

AGENDA

1. **Review:** Descriptive Statistics
2. **Today's Topic:** Feature Scaling
3. Exercise
4. Presentations

REVIEW

DESCRIPTIVE STATISTIC IN PYTHON



WOMEN WHO CODE

TODAY'S TOPIC

DATA PRE-PROCESSING: FEATURE SCALING



FEATURE SCALING

Different features → measured on different scales.

- height – centimetres
- weight – kilograms
- blood pressure – mmHg

Some classifiers combine and compare feature values (e.g. Euclidean distance).



FEATURE SCALING

Features with a broad range of values → dominate features with a smaller range of values:

- percentage of unemployment in a city - ranges from 0.0 to 1.0
- population of the city - can range up to 500,000

Feature scaling transforms the data so that the features have a uniform range.



WOMEN WHO CODE

Min-max Scaling

Scales values to a range of $[0, 1]$.



Min-max Scaling

Rescaling feature vector X :

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

Example:

$$z_1 = \frac{22 - 22}{42 - 22} = 0.0$$

ID	Age	Age_Scaled
1	22	0.00
2	25	
3	30	
4	42	

Min-max Scaling

Rescaling feature vector X :

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

Example:

$$z_2 = \frac{25 - 22}{42 - 22} = 0.15$$

ID	Age	Age_Scaled
1	22	0.00
2	25	0.15
3	30	
4	42	

Min-max Scaling

Rescaling feature vector X :

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

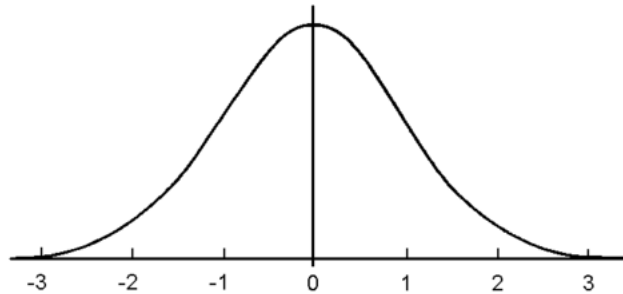
Example:

$$z_3 = \frac{30 - 22}{42 - 22} = 0.4$$

ID	Age	Age_Scaled
1	22	0.00
2	25	0.15
3	30	0.40
4	42	1.00

Standardization

Scales features so that they are centered around 0 with a standard deviation of 1.



Standardization

Rescaling feature vector X :

$$z_i = \frac{x_i - \mu}{\sigma}$$

where μ is the mean (average)

σ is the standard deviation

Exercise 1: Check that $\mu = 29.75$, $\sigma = 8.81$ and that the values in the Age_Scaled are correct.

ID	Age	Age_Scaled
1	22	-1.01
2	25	-0.62
3	30	0.03
4	42	1.60

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

Standardization

Rescaling feature vector X :

$$z_i = \frac{x_i - \mu}{\sigma}$$

where μ is the mean (average)

σ is the standard deviation

Exercise 2: Check that after scaling, the values for μ and σ are approximately 0 and 1, respectively.

ID	Age	Age_Scaled
1	22	-1.01
2	25	-0.62
3	30	0.03
4	42	1.60

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

WOMEN WHO

Normalization

Scales the feature vector to a unit vector

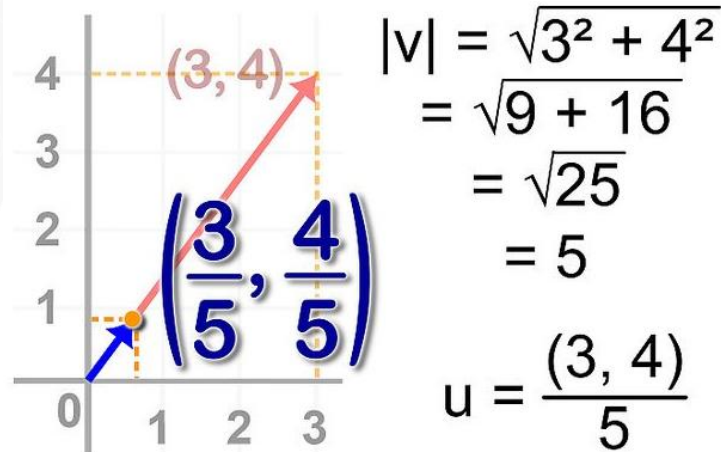
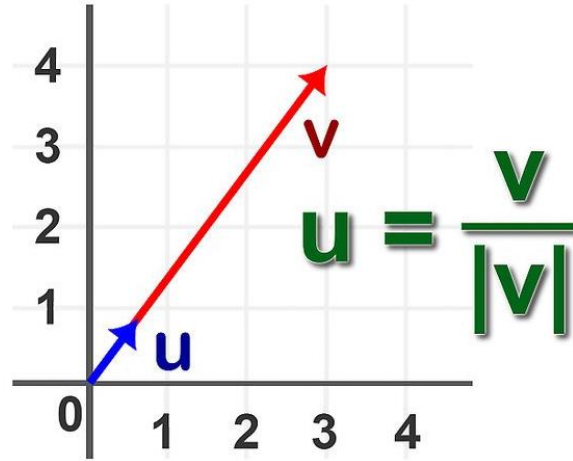


Normalization

ID	Age	Age_Scaled
1	3	$3/5 = 0.6$
2	4	$4/5 = 0.8$

(Norm) A vector with n elements has length:

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$



Feature Scaling

Min-max Scaling

Standardization

Normalization

Binarization



Partner/Group/Individual Exercise:

**1. DESCRIPTIVE STATISTICS
REVIEW**

2. WINE DATA CLASSIFICATION



EXERCISES

- **ML beginners:** Start with the ‘Introduction to ML Tutorial’ and work your way up.
- If you haven’t finished an exercise/assignment, try finishing that one first before moving to the next.
- Have a question? Post it here:
<https://gitter.im/WWCodeManila-ML-AI/>
- When you’re done, share your work! :D

WOMEN WHO

Partner/Group/Individual Presentation



Assignment

Binarize Features in the Handwritten Digit Recognition Exercise



References:

WWCodeLondon Slides

<http://scikit-learn.org/stable/modules/preprocessing.html>

http://sebastianraschka.com/Articles/2014_about_feature_scaling.html

T.I.L.

SHARE IT!
In front!

On Twitter: @wwcodemanila
Or FB: fb.com/wwcodemanila

Don't forget to tag WWCodeManila so we can retweet or share it.

WOMEN WHO

THANK YOU :)

CODE®