SSID: WWCode
Password: password

WOMEN WHO CODE MANILA

Machine Learning & AI Study Group

Twitter: @wwcodemanila
FB: fb.com/wwcodemanila

#WWCodeManila
#YourProgrammingLanguage
#StudyGroup

Issa Tingzon

Research Fellow

Philippine-California Advanced Research Institutes

OUR MISSION

Inspiring women to excel in technology careers.

WOMEN WHO CODE MANILA

OUR VISION

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

WOMEN WHO CODE MANILA

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** promote your recruit or promote your business

New Member's Introduction

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

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.

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:

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

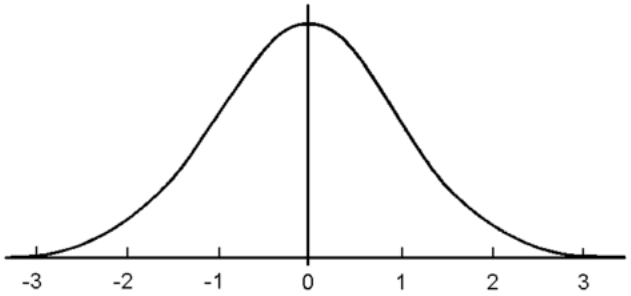
Example:

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

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

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

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

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

Standardization

Rescaling feature vector X:

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

where μ is the mean (average)

σ is the standard deviation

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

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

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

Normalization

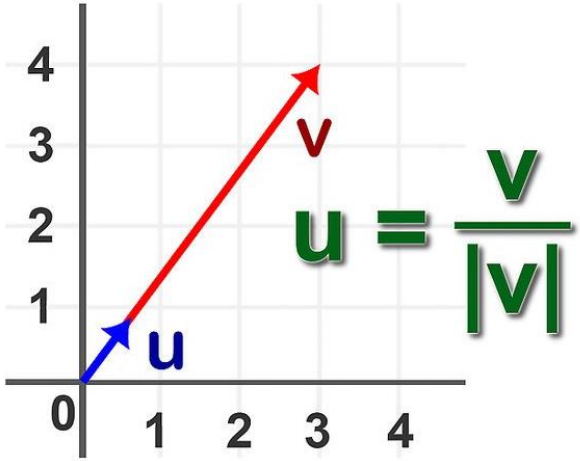
Scales the feature vector to a unit vector

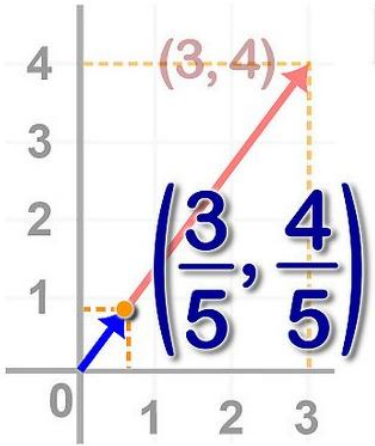
Normalization

ID	Age	Age_Scaled
1	3	$^3/_5 = 0.6$
2	4	$^4/_5 = 0.8$

(Norm) A vector with n elements has length:

$$||v|| = \sqrt[2]{v_1^2 + v_2^2 + \cdots + v_n^2}$$





$$|v| = \sqrt{3^2 + 4^2}$$
$$= \sqrt{9 + 16}$$
$$= \sqrt{25}$$
$$= 5$$

$$u = \frac{(3, 4)}{5}$$

Feature Scaling

Min-max Scaling

Standardization

Normalization

Binarization

Partner/Group/Individual Exercise:

1. DESCRIPTIVE STATISTICS REVIEW
2. WINE DATA CLASSIFICATION

Note: First-timers/beginners are recommended to start with the Introduction to ML Tutorial (see official github repo)

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.

THANK YOU :)