Excercise 1: Data Normalization

mchuongbk1995

October 2020

1 Introduction

- In Normalization, the attribute values are scaled to a small specified range.
- In this exercises, we have to practice min-max normalization and z-score normalization.

2 Knowledge

2.1 Min-max normalization

Min-max normalization is the simplest method to normalize value to new range $[new\_max_i - new\_min_i]$

$$v^{new} = \frac{v^{old} - min_i}{max_i - min_i}(new\_max_i - new\_min_i) + new\_min_i \tag{1}$$

2.2 Z-score normalization

The general method of calculation is to determine the distribution mean μ_i and standard deviation σ_i for each feature.

$$v^{new} = \frac{v^{old} - \mu_i}{\sigma_i} \tag{2}$$

3 Exercise

Suppose $m = 4$ students have taken some class, and the class had a midterm exam and a final exam. You have collected a data set of their scores on the two exams, which is as follows:

midterm exam	final exam
89	96
72	74
94	87
69	78

What is the normalized feature of third student if you using the following normalization strategies?

1. Min-max normalization with new range (0,1)
2. Z-score normalization

4 Solving

1. Min-max normalization:

$$v^{(2)}_{midterm} = \frac{72 - 69}{94 - 69} * 1 + 0 = 0.15$$

2. Z-score normalization:

$$\mu^{(2)}_{midterm} = \frac{89 + 72 + 94 + 69}{4} = 81$$
$$\sigma_{midterm} = 10.7$$
$$v^{(2)}_{midterm} = \frac{72 - 81}{10.7} = -0.84$$