

Feature Scaling

Patrick Smith

Feature Scaling

LEARNING OBJECTIVES

- Use the scikit-learn preprocessing module to scale data in various ways

Feature Scaling

PRE-WORK

- Load and manipulate data with Pandas
- Fit models with scikit-learn

Feature Scaling

OPENING

Feature Scaling

Introduction

Feature Scaling

Why do we scale data?

Feature Scaling

Why do we scale data?

There are a number of good reasons why we scale our data:

- To handle disparities in units
- Because many machine learning models require scaling
- It can speed up our models

Feature Scaling

Scaling is good for gradient descent

- **The reason we scale for gradient descent is to prevent major differences in the steps on different axis to be widely different.**
- **This makes it difficult to find a good learning rate since once that is too small will take a long time to move around in the direction of a larger-scale feature, and a learning rate that is too large will not have good resolution on a smaller-scale feature.**

Feature Scaling

- We scale for models to prevent major differences in the steps on different axis from varying widely
- It's rarely a bad idea to scale your data!

Feature Scaling



- ☐ Perform exploratory data analysis
- ☐ Verify the quality of the data

MINE THE DATA

- ☐ Determine sampling methodology and sample data
- ☐ Format, clean, slice, and combine data in Python
- ☐ Create necessary derived columns from the data (new data)

REFINE THE DATA

Feature Scaling

- We scale for models to prevent major differences in the steps on different axis from varying widely
- It's rarely a bad idea to scale your data!
 - Only exception: Tree based models

Feature Scaling

How do we scale our data?

Feature Scaling

How do we scale our data?

We can scale data in a few ways

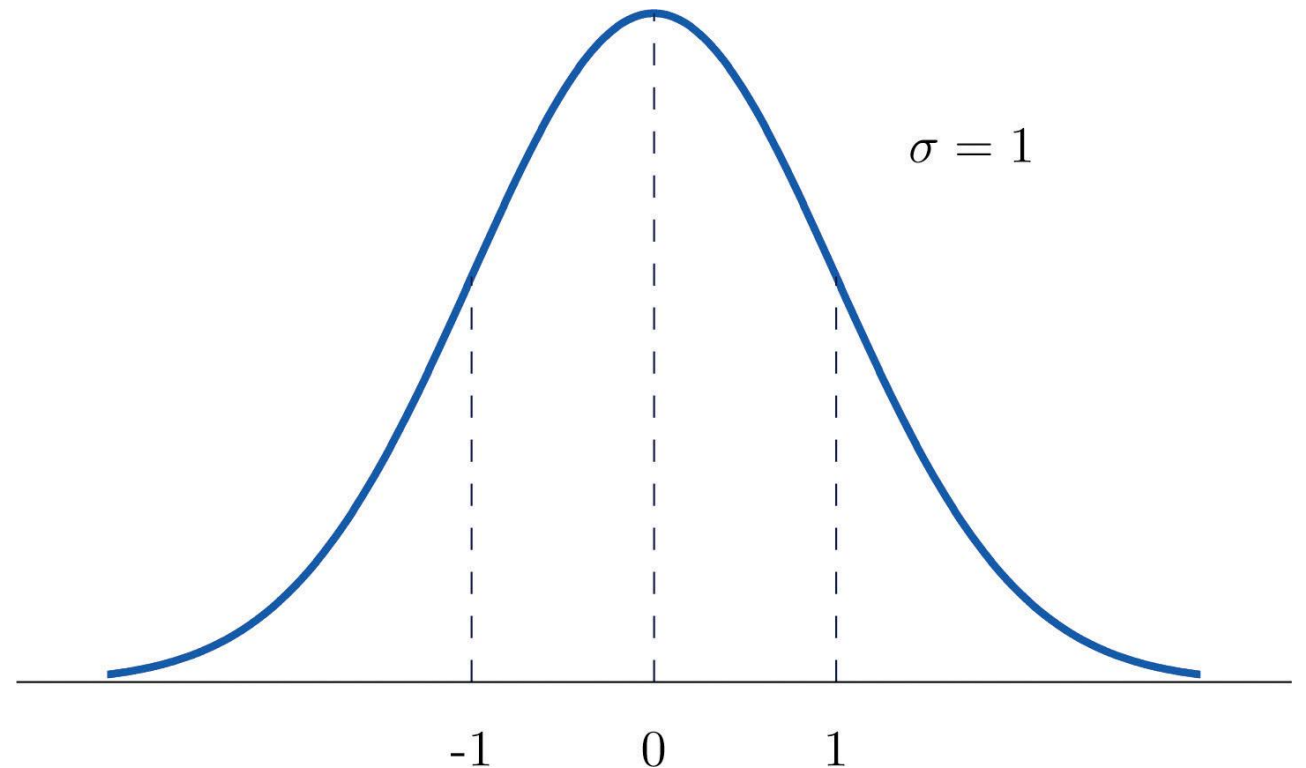
- 1. Standardization: (Also called Z-Score normalization) Takes a feature and rescales it to have a mean of zero and a variance of 1**

Feature Scaling

1. Standardization: Takes a feature and rescales it to have a mean of zero and a variance of 1

- Compute the mean and standard deviation

$$x' = (x - \text{mean}) / \text{std_dev}$$



Feature Scaling

How do we scale our data?

We can scale data in a few ways

- 1. Standardization: Takes a feature and rescales it to have a mean of zero and a variance of 1**
- 2. Min-Max Scaling: Simple rescaling of data to fit a defined interval**

Feature Scaling

How do we scale our data?

We can scale data in a few ways

1. **Standardization:** Takes a feature and rescales it to have a mean of zero and a variance of 1
2. **Min-Max Scaling:** Simple rescaling of data to fit a defined interval (between 0 and 1)

$$x' = (x - \min) / (\max - \min)$$

Feature Scaling

When would we want to standardize?

- **KNN: Scaling is necessary if you want all the features to contribute equally!**
- **K-Means: Since the algorithm works off of computing means, unscaled data will severely affect the model**
- **Logistic Regression, Neural Networks, and SVMs - unscaled data will disproportionately weigh some data points**

Feature Scaling

When would we want to standardize?

- **KNN: Scaling is necessary if you want all the features to contribute equally!**
- **K-Means: Since the algorithm works off of computing means, unscaled data will severely affect the model**
- **Logistic Regression, Neural Networks, and SVMs - unscaled data will disproportionately weigh some data points**

Min-Max scaling will give us a similar result, but smaller standard deviation

Feature Scaling

When would we want to use standardization vs min-max ?

Feature Scaling

Demo: Scaling in Python

Feature Scaling

Guided Practice: Scaling in Python

Feature Scaling

Independent Practice

Conclusion

Feature Scaling

Q & A

Feature Scaling

Review

Feature Scaling

EXIT TICKET

DON'T FORGET TO FILL OUT YOUR EXIT TICKET