




< Previous	 	 	 						Next >
------------	---	---	---	--	---	---	---	---	--------

Feature Scaling - Normalization and Standardization

 Bookmark this page

Normalization

Normalization, sometimes called Min-Max scaling, transforms the features to a predefined range. Typically, we normalize to between 0 and 1. This sets 0 for the lowest value in our data for that variable, sets 1 for the highest, and all other values to be a proportional float between 0 and 1. The formula for normalization is as follows:

 NORMALIZATION

$$X^{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Note that we can also use [sklearn's MinMaxScaler\(\)](#) to normalize our data.

Video


HarvardX

Copyright © President and Fellows of Harvard College



2:10 / 2:10

1.0x

Video


 [Descargar archivo de video](#)

Transcripciones

-  [Descargar SubRip \(.srt\) file](#)
-  [Descargar Text \(.txt\) file](#)

Standardization

Standardization transforms the features by subtracting the mean and dividing by the standard deviation. This sets the new mean to 0 with a standard deviation of 1.

 STANDARDIZATION

$$X^{stand} = \frac{X - \bar{X}}{\sigma_X}$$

Note that we can also use sklearn's StandardScaler() to standardize our data.

the data also follows a Gaussian distribution. Most data does follow a Gaussian distribution but this is not necessarily true. Standardization is also less impacted by outliers since there is no predefined range of features.

Normalization is recommended for data that already has a fixed range. Such as a survey question with values that can only be between 1 and 10. Data that does not follow a Gaussian distribution may work better with Normalization as well. Normalization is more vulnerable to outliers than standardization.

However, these are just rules of thumb to provided you with some intuition. It is often fine to simply try both and figure out what works best for your data.

Video

⋮

[COMPUTER KEYBOARD TYPING]

▶ 0:00 / 0:00

▶ 1.0x

🔊

🔍

CC

🗣️

◀

[Inicio de la transcripción. Saltar al final.](#)

[COMPUTER KEYBOARD TYPING]

[MOUSE CLICKING]

CHRIS G: They suggested that I scale my predictor first

before creating a squared term and adding it to the model.

OK, it's worth a shot.

I just create an instance of the standard scalar

Video

📄 [Descargar archivo de video](#)

Transcripciones

- 📄 [Descargar SubRip \(.srt\) file](#)
- 📄 [Descargar Text \(.txt\) file](#)

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

< Previous

Next >



edX

- [About](#)
- [Affiliates](#)
- [edX for Business](#)
- [Open edX](#)
- [Careers](#)
- [News](#)

Legal

- [Terms of Service & Honor Code](#)
- [Privacy Policy](#)
- [Accessibility Policy](#)
- [Trademark Policy](#)
- [Sitemap](#)
- [Cookie Policy](#)
- [Your Privacy Choices](#)

Connect

- [Idea Hub](#)
- [Contact Us](#)
- [Help Center](#)
- [Security](#)
- [Media Kit](#)



© 2024 edX LLC. All rights reserved.
深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)