

# Package ‘torchdatasets’

January 19, 2021

**Title** Ready to Use Extra Datasets for Torch

**Version** 0.0.1

**Description** Provides datasets in a format that can be easily consumed by torch 'dataloaders'.  
Handles data downloading from multiple sources, caching and pre-processing so  
users can focus only on their model implementations.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Imports** torch, fs, zip, pins, torchvision, stringr

**Suggests** testthat, readr

**URL** <https://mlverse.github.io/torchdatasets/>,  
<https://github.com/mlverse/torchdatasets>

**BugReports** <https://github.com/mlverse/torchdatasets/issues>

**NeedsCompilation** no

**Author** Daniel Falbel [aut, cre],  
RStudio [cph]

**Maintainer** Daniel Falbel <daniel@rstudio.com>

**Repository** CRAN

**Date/Publication** 2021-01-19 09:40:02 UTC

## R topics documented:

bird_species_dataset . . . . .	2
dogs_vs_cats_dataset . . . . .	3
guess_the_correlation_dataset . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

bird\_species\_dataset *Bird species dataset*

---

## Description

Prepares the bird species dataset available in Kaggle [here](#)

## Usage

```
bird_species_dataset(  
  root,  
  token = NULL,  
  split = "train",  
  download = FALSE,  
  ...  
)
```

## Arguments

root	path to the data location
token	a path to the json file obtained in Kaggle. See <a href="#">here</a> for additional info.
split	train, test or valid
download	wether to download or not
...	other arguments passed to <code>torchvision::image_folder_dataset()</code> .

## Details

We use pins for downloading and managing authentication. If you want to download the dataset you need to register the Kaggle board as described in [this link](#). or pass the token argument.

## Value

A `torch::dataset()` ready to be used with dataloaders.

## Examples

```
if (torch::torch_is_installed() && FALSE) {  
  birds <- bird_species_dataset("./data", token = "path/to/kaggle.json",  
                               download = TRUE)  
  length(birds)  
}
```

---

dogs\_vs\_cats\_dataset *Dog vs cats dataset*

---

### Description

Prepares the dog vs cats dataset available in Kaggle [here](#)

### Usage

```
dogs_vs_cats_dataset(root, token = NULL, download = TRUE, ...)
```

### Arguments

root	path to the data location
token	a path to the json file obtained in Kaggle. See <a href="#">here</a> for additional info.
download	wether to download or not
...	other arguments passed to <code>torchvision::image_folder_dataset()</code> .

### Value

A `torch::dataset()` ready to be used with dataloaders.

### Examples

```
if (torch::torch_is_installed() && FALSE) {  
  dogs_cats <- dogs_vs_cats_dataset("./data", token = "path/to/kaggle.json",  
                                   download = TRUE)  
  length(dogs_cats)  
}
```

---

guess\_the\_correlation\_dataset

*Guess The Correlation dataset*

---

### Description

Prepares the Guess The Correlation dataset available on Kaggle [here](#) A copy of this dataset is hosted in a public Google Cloud bucket so you don't need to authenticate.

### Usage

```
guess_the_correlation_dataset(  
  root,  
  split = "train",  
  transform = NULL,  
  indexes = NULL,  
  download = FALSE  
)
```

**Arguments**

root	path to the data location
split	string. 'train' or 'submission'
transform	function that receives a torch tensor and return another torch tensor, transformed.
indexes	set of integers for subsampling (e.g. 1:140000)
download	whether to download or not

**Value**

A torch dataset that can be consumed with `torch::dataloader()`.

**Examples**

```
if (torch::torch_is_installed() && FALSE) {  
  gtc <- guess_the_correlation_dataset("./data", download = TRUE)  
  length(gtc)  
}
```

# Index

`bird_species_dataset`, 2

`dogs_vs_cats_dataset`, 3

`guess_the_correlation_dataset`, 3

`torch::dataloader()`, 4

`torch::dataset()`, 2, 3

`torchvision::image_folder_dataset()`, 2,  
3