

1. Preprocessing

Before running, please download "train.zip", "test.zip", "sample_submission.zip" and "labels.csv.zip" from [Kaggle \(https://www.kaggle.com/c/dog-breed-identification/data\)](https://www.kaggle.com/c/dog-breed-identification/data). Create a folder named "data" in the directory storing this Python notebook and unzip the content of the four files inside "data" folder.

Please note that the training dataset from Kaggle is not used in the later steps. To enhance learning performance, Stanford dogs dataset instead of the training set provided by Kaggle is used.

Loading packages

```
In [1]: import shutil
import os
import pandas as pd
```

Setting parameters

```
In [2]: data_dir      = "data"
old_train_path = 'train'
new_train_path = 'for_train'
old_test_path  = 'test'
new_test_path  = 'for_test'
fakeBreed      = '0'
```

Code

```
In [3]: df = pd.read_csv(os.path.join('.', data_dir, 'labels.csv'))
old_train_path = os.path.join('.', data_dir, old_train_path)
new_train_path = os.path.join('.', data_dir, new_train_path)
if os.path.exists(new_train_path):
    shutil.rmtree(new_train_path)
os.makedirs(new_train_path)

for i, (fname, breed) in df.iterrows():
    new_train_path2 = os.path.join('%s', '%s') % (new_train_path, breed)
    if not os.path.exists(new_train_path2):
        os.makedirs(new_train_path2)
    shutil.copy(os.path.join(old_train_path, '%s.jpg') % fname,
                os.path.join(new_train_path2, '%s.jpg') % fname)
```

```
In [4]: df = pd.read_csv(os.path.join('.', data_dir, 'sample_submission.csv'))
old_test_path = os.path.join('.', data_dir, old_test_path)
new_test_path = os.path.join('.', data_dir, new_test_path)
if os.path.exists(new_test_path):
    shutil.rmtree(new_test_path)
os.makedirs(new_test_path)

if os.path.exists(os.path.join(new_test_path, fakeBreed)):
    shutil.rmtree(os.path.join(new_test_path, fakeBreed))
os.makedirs(os.path.join(new_test_path, fakeBreed))

for test_file in os.listdir(old_test_path):
    shutil.copy(os.path.join(old_test_path, test_file),
                os.path.join(new_test_path, fakeBreed))
```