data_for_machine_learning

March 3, 2019

```python
In [3]: import pickle
        import urllib
        from sklearn.datasets import fetch_mldata, load_svmlight_file, fetch_covtype, fetch_20r
```

```python
In [6]: # mnist = fetch_mldata("MNIST original") # [WinError 10060] A connection attempt faile
        from tensorflow.examples.tutorials.mnist import input_data
        mnist = input_data.read_data_sets('MNIST_data', one_hot=True)
        pickle.dump(mnist, open("mnist.pickle",'wb'))
```

```
D:\Python\Lib\importlib\_bootstrap.py:219: RuntimeWarning: numpy.ufunc size changed, may indica
  return f(*args, **kwds)
D:\Python\Lib\importlib\_bootstrap.py:219: RuntimeWarning: numpy.ufunc size changed, may indica
  return f(*args, **kwds)
D:\Python\Lib\importlib\_bootstrap.py:219: RuntimeWarning: numpy.ufunc size changed, may indica
  return f(*args, **kwds)


WARNING:tensorflow:From <ipython-input-6-a1d59a542922>:3: read_data_sets (from tensorflow.con
Instructions for updating:
Please use alternatives such as official/mnist/dataset.py from tensorflow/models.
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\data
Instructions for updating:
Please write your own downloading logic.
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\data
Instructions for updating:
Please use urllib or similar directly.
Successfully downloaded train-images-idx3-ubyte.gz 9912422 bytes.
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\data
Instructions for updating:
Please use tf.data to implement this functionality.
Extracting MNIST_data\train-images-idx3-ubyte.gz
Successfully downloaded train-labels-idx1-ubyte.gz 28881 bytes.
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\data
Instructions for updating:
Please use tf.data to implement this functionality.
Extracting MNIST_data\train-labels-idx1-ubyte.gz
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\data
Instructions for updating:
```

```
Please use tf.one_hot on tensors.
Successfully downloaded t10k-images-idx3-ubyte.gz 1648877 bytes.
Extracting MNIST_data\t10k-images-idx3-ubyte.gz
Successfully downloaded t10k-labels-idx1-ubyte.gz 4542 bytes.
Extracting MNIST_data\t10k-labels-idx1-ubyte.gz
WARNING:tensorflow:From d:\python\lib\site-packages\tensorflow\contrib\learn\python\learn\datas
Instructions for updating:
Please use alternatives such as official/mnist/dataset.py from tensorflow/models.
```

```
In [7]: target_page = 'http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary/ijcnn1.bz2
        with urllib.request.urlopen(target_page) as response:
            with open('ijcnn1.bz2','wb') as W:
                W.write(response.read())
```

```
In [8]: target_page = 'http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression/cadata
        cadata=load_svmlight_file(urllib.request.urlopen(target_page))
        pickle.dump(cadata,open('cadata.pickle', 'wb'))
```

```
In [10]: covertype_dataset = fetch_covtype(random_state=101, shuffle=True)
         pickle.dump(covertype_dataset,open('covertype_dataset.pickle', 'wb'))
```

```
Downloading https://ndownloader.figshare.com/files/5976039
```

```
In [12]: news_groups_dataset = fetch_20newsgroups(shuffle=True, remove=('readers','footers','qu
         pickle.dump(news_groups_dataset, open('news_groups_dataset.pickle','wb'))
```

```
In [ ]:
```