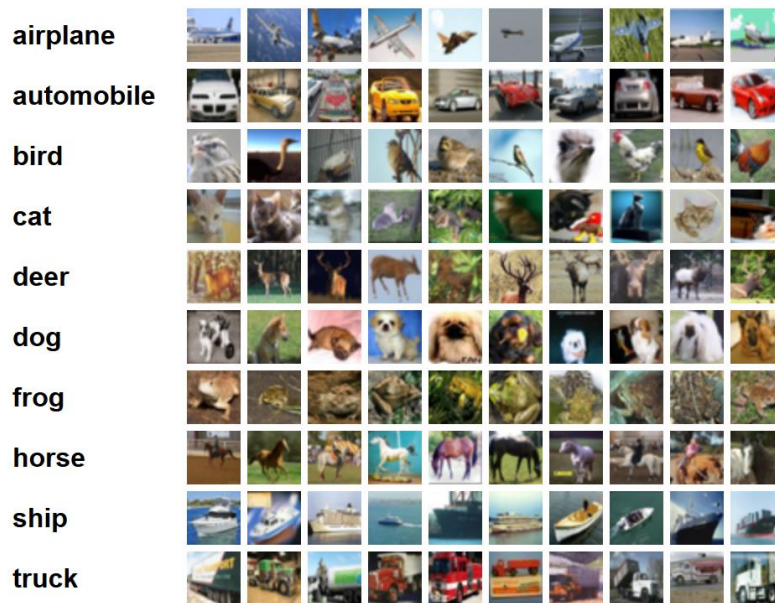


## استفاده از شبکه‌های عصبی به منظور دسته‌بندی داده‌های موجود در مجموعه داده CIFAR-10 با بالاترین دقت ممکن

کارگاه یادگیری ماشین با پایتون  
آزمایشگاه هوش محاسباتی و یادگیری ماشین (همیاد)

مهلت تحویل: ۲۲ آبان ۱۳۹۶



در جلسه سوم کارگاه، کتابخانه Scikit Learn معرفی شد و شیوه استفاده از آن برای حل مسائل گوناگون، به ویژه مسئله تشخیص چهره و مسئله تشخیص ارقام دست‌نویس، بررسی شد. در این تمرین شما باید از کدهای مشابه ارائه شده در کلاس، برای طراحی و اجرای یک دسته‌بند با بالاترین دقت ممکن بر روی مجموعه داده CIFAR-10 استفاده نمایید.

در ادامه، مراحل انجام تمرین و کارهایی که شما باید انجام دهید، آورده شده است.

### مرحله اول) آماده‌سازی داده‌ها

در اینجا مجموعه کدهای لازم برای بارگذاری این مجموعه ارائه شده است. ابتدا باید توابع زیر را تعریف نمایید:

```
import os, pickle

def load_cifar10_batch(filename):
    """ Load a single batch from CIFAR10 """
    with open(filename, 'rb') as f:
        datadict = pickle.load(f, encoding='bytes')
        X = datadict[b'data']
```

```

Y = datadict[b'labels']
X = X.reshape(10000, 3, 32, 32)
X = X.transpose(0, 2, 3, 1).astype('float')
Y = np.array(Y)
return X, Y

def load_cifar10(dir):
    """ Load all batches of CIFAR10 """
    # load train batch file
    xs = []
    ys = []

    for i in range(1, 6):
        filename = os.path.join(dir, 'data_batch_%d' % i)
        X, Y = load_cifar10_batch(filename)
        xs.append(X)
        ys.append(Y)

    Xtr = np.concatenate(xs)
    Ytr = np.concatenate(ys)
    del X, Y

    # load test batch
    Xte, Yte = load_cifar10_batch(os.path.join(dir, 'test_batch'))
    return Xtr, Ytr, Xte, Yte

```

پس از تعریف این دو تابع، با دستور زیر، این مجموعه داده را خوانده و در ماتریس‌های مربوطه ذخیره نمایید.

```
X_train, y_train, X_test, y_test = load_cifar10('datasets/cifar-10-batches-py')
```

اکنون ماتریس `X_train` شامل ۵۰۰۰۰ داده از مجموعه CIFAR-10 است به گونه‌ای که هر داده در یک سطر از این ماتریس ذخیره شده و شامل ۳۰۷۲ مقدار است ( $3 \times 32 \times 32$ ). شما باید از این داده‌ها به منظور آموزش شبکه عصبی خود استفاده کنید.

علاوه بر این، پس از اجرای دستورات بالا، ماتریس `X_test` نیز شامل ۱۰۰۰۰ داده از مجموعه داده CIFAR-10 است و شما باید از این داده‌ها به منظور ارزیابی دقت شبکه عصبی خود استفاده نمایید.

### مرحله دوم) ایجاد و تعریف شبکه عصبی

پس از ایجاد مجموعه آموزشی و مجوعه آزمایشی، باید شبکه عصبی را تعریف و ایجاد نمایید. برای این منظور می‌توانید از دستوراتی مشابه دستورات ارائه شده در ادامه استفاده نمایید

```

from sklearn.neural_networks import MLPClassifier
clf = MLPClassifier()

```

وظیفه شما این است که هنگام ایجاد دسته‌بند خود با استفاده از دستور بالا، آرگومان‌های مورد نیاز را با مقادیر مناسب مقداردهی نمایید (به جای استفاده از مقادیر پیش‌فرض).

### مرحله سوم) آموزش شبکه عصبی بر روی مجموعه آموزشی

پس از ایجاد شبکه عصبی، می‌توانید شبکه را بر روی مجموعه آموزشی به صورت زیر آموزش دهید:

```
clf.fit(X_train, y_train)
```

### مرحله چهارم) پیش‌بینی کلاس مربوط به داده‌های مجموعه آزمایشی

پس از آموزش شبکه عصبی، می‌توانید به شکل زیر از شبکه آموزش یافته به منظور انجام دسته‌بندی بر روی داده‌های مجموعه آزمایشی استفاده نمایید و نتایج دسته‌بندی را در آرایه `y_pred` ذخیره نمایید.

```
y_pred = clf.predict(X_test)
```

### مرحله پنجم) محاسبه دقت شبکه عصبی بر روی داده‌های مجموعه آزمایشی

اکنون می‌توانید دقت دسته‌بندی شبکه عصبی خود را به سادگی و با استفاده از دستورات زیر محاسبه و چاپ نمایید.

```
accuracy = np.mean(y_test == y_pred)  
print("accuracy = %.2f" % accuracy)
```

**راهنمایی.** برخی از مهم‌ترین پارامترهایی که باید مقادیر مختلفی را برای آنها آزمایش کنید تا بهترین دقت دسته‌بندی ممکن را به دست آورید، به شرح زیر است:

تعداد لایه‌های مخفی و اندازه هر یک از این لایه‌ها

- نرخ یادگیری
- الگوریتم بهینه‌سازی
- حداکثر تعداد تکرارها
- روش تنظیم و ضریب تنظیم

### نحوه ارسال پاسخ

پس از حل تمرین و رسیدن به دقت مورد نظر، لطفاً **تنها** فایل نوت‌بوک پایتون را در سایت پیازا به صورت خصوصی (به منظور عدم مشاهده توسط افراد دیگر)، برای اینجانب ارسال فرمایید.

یادآوری. لطفاً برای انجام تمرین‌ها از پایتون نسخه ۳ استفاده نمایید.