## 实验报告

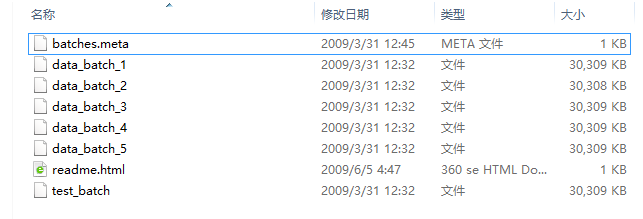
### 数据预处理

官网下载的数据集，包含训练集，data\_banch\_n 5万张图片；测试集，test\_banch 1万张图片；还有10个类别名称，见banches.meta.

编写load\_data()加载数据集的图片向量的numpy array,

编写load\_label\_names()加载数据集的图片类标的numpy array,

编写mat\_to-img把向量转成图片，进行查看图片



### 编写分类器

见 KKN.py

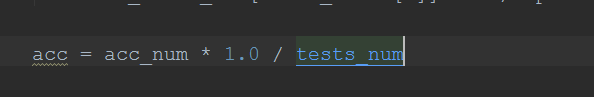
编写Class KNN分类器，包含predict(self, x, k=1, L='L1')方法。

通过计算图片向量之间的差异，对目标图片进行分类。

对比不同K值和距离函数对准确率的影响。

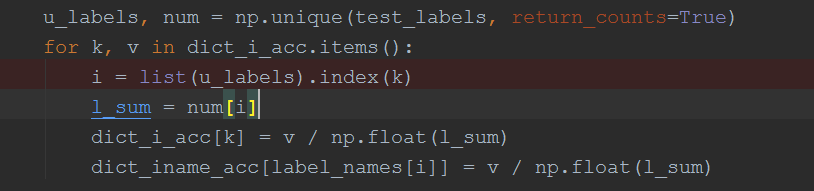
K=5时，增加阈值参数，处理预测类别不唯一的情况

### 测试准确率

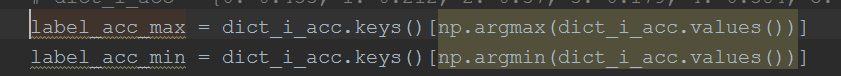


### 计算各类别的准确率

公式 i\_acc = i\_acc\_num / i\_num

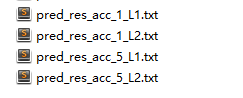


### 寻找准确率最高的类别



### 保存实验结果

不同的K和距离函数得到的结果见下图中的文件：KNN\data目录下



50000条训练数据，1000条测试数据的结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K | L | 准确率 | 个类别准确率 | label\_acc\_max |
| 1 | L1 | 0.3859 | {"1": 0.297, "0": 0.523, "3": 0.264, "2": 0.407, "5": 0.306, "4": 0.452, "7": 0.343, "6": 0.371, "9": 0.276, "8": 0.62} | 8 |
| 1 | L2 | 0.3539 | {"1": 0.218, "0": 0.485, "3": 0.24, "2": 0.384, "5": 0.29, "4": 0.457, "7": 0.294, "6": 0.353, "9": 0.199, "8": 0.619} | 8 |
| 5 | L1 | 0.377 | {"1": 0.288, "0": 0.582, "3": 0.246, "2": 0.456, "5": 0.266, "4": 0.489, "7": 0.267, "6": 0.288, "9": 0.226, "8": 0.662} | 8 |
| 5 | L2 | 0.3398 | {"1": 0.205, "0": 0.537, "3": 0.217, "2": 0.452, "5": 0.22, "4": 0.514, "7": 0.21, "6": 0.248, "9": 0.14, "8": 0.655} | 8 |
|  |  |  |  |  |