

NMDA 2019 第二次作业

推荐安装Anaconda软件包。每道题提交一个Jupyter Notebook文件，其中包括可直接运行的Python Code，以及Markdown文字说明和结果讨论。请勿提交数据

1. Labeled Faces in the Wild (LFW)是 University of Massachusetts 从网络上搜集的自然条件下的人脸图像数据库，其中包括 13233 张人脸图像，5749 个不同的人，其中 1680 人有 2 张以上的脸部图像。针对其中一个结果预处理过的数据集 lfw_funneled，提取了其中每人有 70 张以上照片的一个子集，包含 7 个人的 1288 张人脸），Scikit-learn 网站给出了采用主成分分析降维之后通过支持向量机 SVM 分类人脸的方法代码：http://scikit-learn.org/stable/auto_examples/applications/plot_face_recognition.html#sphx-glr-auto-examples-applications-plot-face-recognition-py

请完成以下任务，作业要求提交文件 lfw. ipynb:

- 1) 自己编写主成分分析PCA函数my_PCA，替换Scikit-learn网站代码中的PCA函数。my_PCA分别采用协方差矩阵求特征值，和奇异值分解(SVD)两种方法，比较它们的计算速度。
- 2) 采用SVM或者其他你认为合适的分类方法，完成下面的任务：
利用提供的训练数据 (ex1.npz, 字典中包含两个key，分别为X和y，X为2d数据矩阵，为(n_faces, n_dims), y为1d矩阵，(n_faces,)), 设计一个分类流程，以合理的验证方法分析结果。
将模型用于提供的测试数据test.npz，提交一个output.txt的包含预测结果的文件。不同样本间预测的类别以空格分隔，例如：0 1 6 3 ...，（注意不要打乱原有的顺序）。

注：*.npz使用numpy.load()读取（<https://docs.scipy.org/doc/numpy-1.15.0/reference/generated/numpy.load.html>），读取得到的结果为一个dict类。例如，要读取ex1.npz中的X数据，则：

```
>> x = np.load('ex1.npz')
>> X = x['X']
```

2. PCA 分类用于神经细胞放电分类：用金属或者硅基微电极插入活体动物的大脑或者神经核团，可以记录到神经细胞的放电，即动作电位 AP(Action Potential)，或者称为峰电位 Spike。通常一个电极记录到的信号来自多个临近的神经细胞，为了把这几个细胞的 AP 区分出来，要进行 Spike Sorting。Sorting 之后的 AP 序列分别归类到单个神经细胞，进行下一步的定量分析，寻找神经细胞放电和外界刺激以及行为的关系。这是神经电生理研究的重要方法，1981 年获得诺贝尔生理与医学奖的 Hubel 和 Wiessel 采用这种神经记录方法开创了视觉神经编码研究的新领域。所附数据文件 spikewave.mat 中包括 5376 个 spike 波形，每个波形长度 40 个采样点。请完成以下任务，提交 spikesorting. ipynb:

- 1) 把这些 spike 波形叠画在一起，观察并描述它们的可分性；
- 2) 用 PCA 方法把这些 spike 分别降维到 2 维和 3 维空间，并画出来

- 3) 在 2 维和 3 维空间分别采用合适的聚类方法把它们分成合理类别，这些类别就对应于临近的神经细胞放电。
- 4) 把 Spike sorting 之后的波形按照神经细胞归类，用不同颜色再次叠画在一起，观察并描述它们的波形在类内和类间的差异

3. 用自己编写的my_PCA函数完成自己头像的PCA压缩实验，说明实验方法和结果。讨论主成分个数、图像分块大小对压缩比和压缩质量影响，画出变化曲线。请选择一种定量评价压缩质量的方法，不要仅做主观比较。提交文件picom.ipynb。

注：建议将图像的一个矩形小块作为一个样本进行操作