

Assignment 2 (individual): Deadline Saturday 2/2/2019 11:59pm

Dataset description

`sklearn.datasets.fetch_olivetti_faces`

An object with the following attributes::

- `data` : numpy array of shape (400, 4096)
 - Each row corresponds to a ravelled face image of original size 64 x 64 pixels.
- `images` : numpy array of shape (400, 64, 64)
 - Each row is a face image corresponding to one of the 40 subjects of the dataset.
- `target` : numpy array of shape (400,)
 - Labels associated to each face image. Those labels are ranging from 0-39 and correspond to the Subject IDs.
- `DESCR` : string
 - Description of the modified Olivetti Faces Dataset.

Import the olivetti faces dataset

```
from sklearn.datasets import fetch_olivetti_faces
```

fetch the faces data

```
faces = fetch_olivetti_faces()
```

```
print faces.DESCR
```

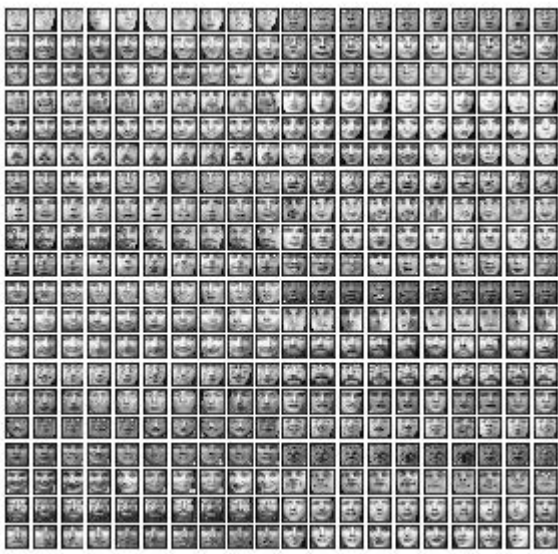
*Dataset consists of 400 images of faces, each one is composed by a matrix of 64x64 pixels.
faces.data has the same data but in rows of 4096 attributes instead of matrices (4096 = 64x64)*

```
print faces.keys()  
print faces.images.shape  
print faces.data.shape  
print faces.target.shape
```

```
['images', 'data', 'target', 'DESCR']
(400L, 64L, 64L)
(400L, 4096L)
(400L,)
```

Please note: you don't have to scale attributes since data is already normalized

```
fig = plt.figure(figsize=(5, 5))
for i in range(400):
    p = fig.add_subplot(20, 20, i + 1,xticks=[], yticks=[])
    p.imshow(faces.images[i], cmap="gray")
```



1. Split the faces.data and faces.target into training and test subsets with test size 25%
2. Choose support vector machine with linear kernel
3. Apply cross validation with 3 folds on the training
4. Print the scores on each testing fold
5. Print the average across all folds
6. Train the svm model on training data (show the accuracy on training data)
7. Predict on the testing data (show the accuracy on testing data)

Here the indexes **ranges** of people with glasses in the array

```
peopleindexwithglass=[(10, 19), (30, 32), (37, 38), (50, 59), (63, 64),
(69, 69), (120, 121), (124, 129), (130, 139), (160, 161),
(164, 169), (180, 182), (185, 185), (189, 189), (190, 192),
(194, 194), (196, 199), (260, 269), (270, 279), (300, 309),
(330, 339), (358, 359), (360, 369)]
```

1. You need to create training training and testing data subsets for glasses with test size 25%

2. Train the svm model on training data (show the accuracy on training data)
3. Predict on the testing data
 1. show the accuracy on testing data)
 2. show the predicted values and corresponding images