pca_prediction

April 21, 2022

[ ]:
```
import os

import matplotlib.pyplot as plt
import numpy as np
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
```

[ ]:
```
os.chdir("../..")
IMAGES_FOLDER = "data/Rice_Image_Dataset"
CATEGORIES = os.listdir(IMAGES_FOLDER)
```

[ ]:
```
NUM_IMAGES_FROM_CATEGORY = 1000
categories = []
images = []
for category in CATEGORIES:
    images_path = os.path.join(IMAGES_FOLDER, category)
    category_images = os.listdir(images_path)[:NUM_IMAGES_FROM_CATEGORY]
    categories += [category] * len(category_images)
    print(f"Category: {category}")
    for image in category_images:
        img = plt.imread(os.path.join(images_path, image))
        images.append(img.flatten())
```

```
Category: Jasmine
Category: Basmati
Category: Karacadag
Category: Ipsala
Category: Arborio
```

[ ]:
```
images = np.array(images)
```

[ ]:
```
x_train, x_test, y_train, y_test = train_test_split(
    images, categories, test_size=0.1
)
```

[ ]:
```
mod = Pipeline([("pca", PCA(50)), ("forest", RandomForestClassifier())])
```

[ ]:
```
mod.fit(x_train, y_train)
```

[ ]:
```
Pipeline(steps=[('pca', PCA(n_components=50)),
                ('forest', RandomForestClassifier())])
```

[ ]:
```
y_pred = mod.predict(x_test)
```

[ ]:
```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

     Arborio       0.97      0.96      0.96        96
     Basmati       0.93      0.98      0.96       102
      Ipsala       0.99      1.00      0.99        93
     Jasmine       0.96      0.92      0.94       106
   Karacadag       0.98      0.97      0.98       103

    accuracy                           0.97       500
   macro avg       0.97      0.97      0.97       500
weighted avg       0.97      0.97      0.97       500
```