coding_hw1_report.md

#Homework 1

Name: Clemence Goh (1002075)

1. Splitting:
   - Splitting has to be done classwise rather than randomly into 50-25-25 due to the idea of preserving the histogram of each set of data such that it matches the original.
   - There is a need to preserve the ratio or frequency of the characteristics.
   - Another way is to split such that there is no spacial bias.

Report

- Randomization of data is done while running the code by shuffling the order of the images in each class
- Code:

```
for groups in flowers_array:
    np.random.shuffle(groups)

    training_data.extend(groups[0:40])
    validation_data.extend(groups[40:60])
    final_test_data.extend(groups[60:80])
```

- regularization constants available: `[0.01, 0.1, 0.1**0.5, 1, 10**0.5, 10, 100**0.5]`
  - From part 1:

```
All accuracies:
[0.9676470588235294, 0.961764705882353, 0.961764705882353,
    0.961764705882353, 0.961764705882353, 0.961764705882353, 0.961764705882353]
Corresponding reg_constants: [0.01, 0.1, 0.31622776601683794, 1, 3.1622776601683795, 10, 10.0]
Highest accuracy: 0.9676470588235294
Best regularizer: 0.01
```

  - Best regularizer: 0.01
  - Final test accuracies for each class:
  -
```
list of accuracies for each class:
[0.9705882352941176, 0.9852941176470589,
0.9911764705882353, 0.9794117647058823,
0.9882352941176471, 0.9852941176470589,
0.9970588235294118, 0.9852941176470589,
0.9970588235294118, 0.9941176470588236,
0.9970588235294118, 0.9852941176470589,
0.9852941176470589, 0.9705882352941176,
0.9852941176470589, 0.9941176470588236,
0.9823529411764705]
```

  - From here it can be seen that in each classifier there are a few datapoints classified wrongly. This is most likely due to the image of the flower looking very similar to another class of flowers.