ENGS 108 Fall 2022 Assignment 1

Due September 15, 2022 at 11:59PM on Canvas

Instructors: George Cybenko

TAs: Chase Yakaboski and Clement Nyanhongo

In [ ]:
```
#TODO: Put any import statements in this code block.
```

Question 1

Go to https://www.kaggle.com/datasets/coloradokb/dandelionimages and download the dandelion and other images.

In [ ]:
```

```

(a) Define a reasonable distance metric between images and explain it with an example.

TODO: Explain your distance metric with text, equations or both.

In [ ]:
```
import numpy as np
import cv2
import os
from matplotlib import pyplot as plt

from google.colab import drive
drive.mount('/content/drive')

Name = []
Name1 = []
Name2 = []
HH1 = []
HH2 = []
HH = []

for i in os.listdir('/content/drive/MyDrive/Images/dandelion/'):
  Name1.append(i)
  image1 = cv2.imread('/content/drive/MyDrive/Images/dandelion/'+i)
  #image1 = cv2.resize(image1,(1000,1000))
  H1 = cv2.calcHist([image1], [1], None, [256], [0, 256])
  #distance metric: Images with more pixels of the same color are more simi
  H1 = cv2.normalize(H1, H1, 0, 1, cv2.NORM_MINMAX, -1)
  HH1.append(H1)

for j in os.listdir('/content/drive/MyDrive/Images/other/'):
  Name2.append(j)
```

```
    Name2.append(j)
    image2 = cv2.imread('/content/drive/MyDrive/Images/other/'+j)
    #image2 = cv2.resize(image2,(1000,1000))
    H2 = cv2.calcHist([image2], [1], None, [256], [0, 256])
    H2 = cv2.normalize(H2, H2, 0, 1, cv2.NORM_MINMAX, -1)
    HH2.append(H2)

Name.extend(Name1)
Name.extend(Name2)
HH.extend(HH1)
HH.extend(HH2)
```

```
Mounted at /content/drive
```

In [ ]:

```
for k in range(len(HH2)):
  print(Name2[k])
```

(b) Which two images are most similar according to your metric?

In [ ]:

```
b = 0
for b1 in range(len(HH)):
  for b2 in range(len(HH)):
    similarity = cv2.compareHist(HH[b1], HH[b2], 0)
    if(similarity > b and Name[b1] != Name[b2]):
      b = similarity
      name1b = Name[b1]
      name2b = Name[b2]
      sim_b = similarity
    similarity = 0
print(name1b,"and",name2b,"are most similar.")
```

```
IMG_3908.jpg and IMG_3909.jpg are most similar.
```

(c) Which two images are most different according to your metric?

In [ ]:

```
c = 1
for c1 in range(len(HH)):
  for c2 in range(len(HH)):
    similarity = cv2.compareHist(HH[c1], HH[c2], 0)
    if(similarity < c and Name[c1] != Name[c2]):
      c = similarity
      name1c = Name[c1]
      name2c = Name[c2]
      sim_c = similarity
    similarity = 0
print(name1c,"and",name2c,"are most different.")
```

```
IMG_1337.jpg and IMG_5484.jpg are most different.
```

(d) What dandelion image is the most representative of the dandelion images?

In [ ]:

```
d = 0
sim_d = []
for d1 in range(len(HH1)):
  d0 = 0
  for d2 in range(len(HH1)):
```

```
    similarity = cv2.compareHist(HH1[d1], HH1[d2], 0)
    d0 = d0 + similarity
  d0 = d0 / len(HH1)
  sim_d.append(d0)
for d3 in range(len(HH1)):
  if(d < sim_d[d3]):
    named = Name1[d3]
    d = sim_d[d3]
print(named,d)
```

```
IMG_6268.jpg 0.653401143544912
```

(e) Which other image is most representative of the other images?

In [ ]:

```
e = 0
sim_e = []
for e1 in range(len(HH2)):
  e0 = 0
  for e2 in range(len(HH2)):
    similarity = cv2.compareHist(HH2[e1], HH2[e2], 0)
    e0 = e0 + similarity
  e0 = e0 / len(HH2)
  sim_e.append(e0)
for e3 in range(len(HH2)):
  if(e < sim_e[e3]):
    namee = Name2[e3]
    e = sim_e[e3]
print(namee,e)
```

```
IMG_6032.jpg 0.7380167688612211
```

(f) What dandelion image and other image are closest to each other?

In [ ]:

```
f = 0
for f1 in range(len(HH1)):
  for f2 in range(len(HH2)):
    similarity = cv2.compareHist(HH1[f1], HH2[f2], 0)
    if(similarity > f):
      f = similarity
      name1f = Name1[f1]
      name2f = Name2[f2]
      sim_f = similarity
    similarity = 0
print(name1f,name2f,sim_f)
```

```
IMG_7520.jpg IMG_7520.jpg 1.0
```

Question 2

Let

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{1}{1 + e^{-2x_1 - 3x_2 - 4x_3 - 5x_4 - 6x_5}}$$

(a) What is

$$\frac{\partial f}{\partial x_i}$$

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \frac{\partial f}{\partial x_4}, \frac{\partial f}{\partial x_5}] \\ &= [\frac{2e^{-2x_1-3x_2-4x_3-5x_4-6x_5}}{(1+e^{-2x_1-3x_2-4x_3-5x_4-6x_5})^2}, \\ &\frac{3e^{-2x_1-3x_2-4x_3-5x_4-6x_5}}{(1+e^{-2x_1-3x_2-4x_3-5x_4-6x_5})^2}, \\ &\frac{4e^{-2x_1-3x_2-4x_3-5x_4-6x_5}}{(1+e^{-2x_1-3x_2-4x_3-5x_4-6x_5})^2}, \\ &\frac{5e^{-2x_1-3x_2-4x_3-5x_4-6x_5}}{(1+e^{-2x_1-3x_2-4x_3-5x_4-6x_5})^2}, \\ &\frac{6e^{-2x_1-3x_2-4x_3-5x_4-6x_5}}{(1+e^{-2x_1-3x_2-4x_3-5x_4-6x_5})^2}]\end{aligned}$$

(b) What is

$$\nabla f(1,0,0,0,0) \cdot ones(5,1)$$

$$\begin{aligned}&= \frac{2e^{-2}}{(1+e^{-2})^2} + \frac{3e^{-2}}{(1+e^{-2})^2} + \frac{4e^{-2}}{(1+e^{-2})^2} + \frac{5e^{-2}}{(1+e^{-2})^2} + \frac{6e^{-2}}{(1+e^{-2})^2} \\ &= 20\frac{e^{-2}}{(1+e^{-2})^2}\end{aligned}$$

(b) What is

$$\nabla f(0,1,0,0,0) \cdot ones(5,1)$$

$$\begin{aligned}&= \frac{2e^{-3}}{(1+e^{-3})^2} + \frac{3e^{-3}}{(1+e^{-3})^2} + \frac{4e^{-3}}{(1+e^{-3})^2} + \frac{5e^{-3}}{(1+e^{-3})^2} + \frac{6e^{-3}}{(1+e^{-3})^2} \\ &= 20\frac{e^{-3}}{(1+e^{-3})^2}\end{aligned}$$

Question 3

Some Tesla batteries have an amp hour rating of 230 but in tests, it has been

determined that the actual amp hours of a new battery is a random variable uniformly distributed between 230 and 250 amp hours. (Tesla is conservative in reporting this.)

(a) Tesla service center is shipped 5 randomly selected batteries. What is the probability that exactly 2 of the 5 batteries have less than 235 amp hours?

$$5select2 * (\frac{235-230}{250-230})^2 * (1-\frac{235-230}{250-230})^3$$

$$= 10(\frac{1}{16}\frac{27}{64})$$

$$= \frac{135}{512}$$

(b) Your Tesla battery is being replaced at that service center. The service center does not measure the amp hours of batteries it receives or installs. What is the probability that the smallest amp hours of the 5 batteries is below 235?

$$\frac{(235-230)}{(250-230)} = \frac{1}{4}$$