

Carlos Granillo

September 28, 2024

ITAI 1378

Lab 05

Professor McManus

Throughout this assignment I was able to learn many new things that I didn't know about coding. For example, I learned how to load the dataset to my new notebook, and this is one of those steps you need to understand from the beginning to get started on your lab. As I continued to work on the lab, I realized that not only is it important to know how to code, but it is also very important to comprehend each code you're inputting so that you know what output or outcome to expect when doing so. For instance, if you want to load a specific dataset or library it must be installed beforehand or it will always give you an error message when trying to run it. This is because each one of those datasets and libraries provide you with different information, therefore each one may be used to perform different tasks. In addition, all of the libraries you wish to install may be installed individually or all together by using the '`!pip install`' command.

When it comes to preparing for the dataset, it is very important to understand the importance of having everything you want to load and use for your notebook. For example, if you wish to load images of different cars and trucks into a dataset, you must first save the images to a file or folder in your machine. By doing so, your notebook can now load them all up when necessary and there are no error codes when you need to use them. Additionally, without properly preparing your dataset you will run into many error codes because the information will not be understood by your machine and your notebook will not run either.

In the CIFAR-10 dataset, everything was already loaded up and ready to run. There were no issues or errors when running all the codes, and I was able to see and understand why each one of the inputted code cells resulted in the expected outcome. One example of this would be how I was able to comprehend the code cells for displaying all the images. The images were all loaded from the computer file we saved and installed onto the new notebook starting from the beginning. Therefore, if no images were already in the file, we would've received an error code because it would not be able to understand what its being asked to do.

On the other hand, I did run into some more trouble when trying to complete the 'Image Processing ML for CV' dataset. This was because I didn't have all the necessary tools to complete it, and you need to have them ready to run before starting to work on the dataset. Because of this, I received an error message that said "no 'Data' file or directory was found" and I was unable to continue working on it. Regardless of not completing it, I was still able to learn some more about OpenCV and SVMs and how they are very important to each other as well. In summary, I learned more about how a dataset really works and how it relates to one another regardless of what they're about. Additionally, every single dataset you ever work on will consists of libraries and preparation to run, or they will always result in some type of error because the machine and notebook will be clueless of what it is being asked to do.


[Copy of 0924Subset_Classical_ML_Image_Classification_with_CIFAR_10_Subset_of_Dataset.ipynb - Colab \(google.com\)](#)

Completed, understood and ran all of the CIFAR 10 dataset / Error code shown on the 2024 Image Processing for DV dataset

Copy of 0924Subset_Classical_ML_Image_Classification_with_CIFAR_10_Subset_of_Dataset.ipynb

File Edit View Insert Runtime Tools Help Last saved at 4:45 PM

+ Code + Text



Step 4: Conclusion

In this notebook, we:

1. Loaded and preprocessed the CIFAR-10 dataset.
2. Converted the images to grayscale and flattened them to use as features.
3. Trained an SVM classifier on the extracted features.
4. Evaluated the model's performance.

Summary of SVM

Support Vector Machines (SVM) are a powerful tool for classification tasks. They work by finding the optimal hyperplane that maximizes the margin between different classes. The key points include:

- **Hyperplane:** The decision boundary.
- **Support Vectors:** Critical data points that define the hyperplane.
- **Margin:** The gap between the hyperplane and the nearest data points from any class.

SVMs are effective in high-dimensional spaces and are versatile due to the use of different kernel functions. However, they can be computationally intensive for large datasets and less effective for overlapping classes.

This exercise provided a basic introduction to image classification using classical machine learning techniques. Next few modules will explore advanced applications, using deep learning models like Convolutional Neural Networks (CNNs) using libraries such as TensorFlow or PyTorch.

```
[ ] print('Carlos Granillo')
```

Carlos Granillo

2024 Image Processing and ML for CV with creating dataset.ipynb

File Edit View Insert Runtime Tools Help Last saved at 2:06 PM

+ Code + Text

```
def load_images(folder):
    images = []
    labels = []
    for label in os.listdir(folder):
        if label == 'cats':
            label_id = 0
        elif label == 'dogs':
            label_id = 1
        else:
            continue
        for filename in os.listdir(os.path.join(folder, label)):
            img_path = os.path.join(folder, label, filename)
            img = cv2.imread(img_path, cv2.IMREAD_GRAYSCALE) # Load image in grayscale
            img = cv2.resize(img, (64, 64)) # Resize image to 64x64
            images.append(img.flatten()) # Flatten the image to 1D array
            labels.append(label_id)
    return np.array(images), np.array(labels)

# Load images from the dataset folder
X, y = load_images('data')

# Display a sample image
plt.imshow(X[0].reshape(64, 64), cmap='gray')
plt.title('Sample Image')
plt.show()
```

```
FileNotFoundError                                Traceback (most recent call last)
<ipython-input-2-a40b3391ecf1> in <cell line: 21>()
     19
     20 # Load images from the dataset folder
--> 21 X, y = load_images('data')
     22
     23 # Display a sample image

<ipython-input-2-a40b3391ecf1> in load_images(folder)
      3 images = []
      4 labels = []
----> 5     for label in os.listdir(folder):
          6         if label == 'cats':
              7             label_id = 0

FileNotFoundError: [Errno 2] No such file or directory: 'data'
```