My GitHub repository https://github.com/saimasano123/L05-image-classification-using-SVM.git

**Objective:** Document your learning and insights from the lab on image classification using the Support Vector Machine (SVM) algorithm with the CIFAR-10 dataset.

**First Impression:**

Honestly, I was intimidated by this lab because I thought that I must create a program where I must use SVM to classify the image. Even when I downloaded the file, I was not able to open it because for some reason the Jupyter lab was not working. Then, Juan one of the classmates suggested and shared the video on how to download anaconda and opened the file in there. Also, I thought I have to individually follow all the codes, like installing the library and importing it. I search on tutorials on why my code is not working or doing anything when I am trying to install library. Then I noticed I can run the code in the notebook and see the result. After that I watched the recording of Q & A session and found out what is required for me to do in this lab.

**Reflection on Learning:**

* SVM algorithm is used to classify and regressed the tasks. It works best by finding the division boundary or hyperplane. Also, to find the best division boundary, we must see who is giving us a bigger or wider margin. The margin is drawn when we draw two parallel lines to hyperplane. I read the definition from the PowerPoint provided in the class and in ChatGPT, but I was not able to grasp the concept. Then, I searched for YouTube videos, but it also did not clear my concept. Also, I was getting headache because there are many new terms, and it is hard to remember all of them or their functionality. So, finally I started watching the video in Hindi like my language. English is not my first language I figured out if I understand these concept in my language, I might excel in this.
* The video I found also helped me since he was explaining it through drawing. Like he explained the whole process by making flowchart.

Data Set

Model Training

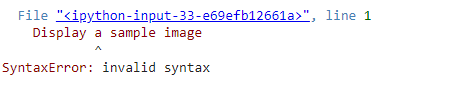
New Data

Prediction

Output

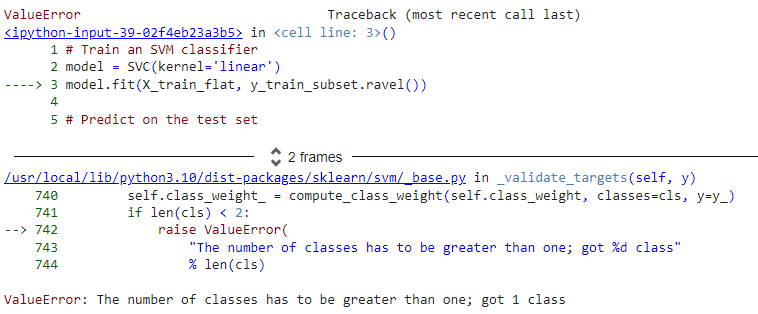
* Like in lab we used CIFAR-10 data set, this data set has about 600 images for each of the ten class. So, by giving that data we are training our SVM, since it is supervised learning. Then at the prediction stage when new data is given, the algorithm will decide whether the item belong to class 1 or class 2 and then it will display the output.
* How the data is split into classes and margin is drawn is also important for SVM to function. Two parallel lines are drawn to values of opponent class that are closer to division boundary. Then, you add the distance of both lines from division boundary. The hyperplane can be drawn vertical, horizontal or diagonal, but the best hyperplane will be the one who margin is greater. This will provide us with more accuracy in classification.

**Refreshing the concept:**

* I know it was not required but I wanted to copy all the codes in new notebook and want to see how it will run. This thing made me revise the concept of writing the code in Jupyter notebook and uploading in GitHub. When I was copying the code in Jupyter, I encountered some issues that help me understand the code better. For example: 

At this point, I searched it on ChatGPT and it told me maybe I am missing “#”.

* Also, I was playing with subset, I decrease the subset to just one class which was frog, so when I run the code to train the SVM classifier, I got an error.



* So, then I came to realize that I need at least two subsets to train the algorithm. Also, when I gathered all the codes in separate file, I realize if I made any changes in the code, I must run the codes from beginning to end to get the accurate result.
* I also noticed that the kernel takes some time to run the last command while processing images from the dataset at different stages. At first, I thought I had made a mistake, but then I realized that in the top right corner, there’s a circle that indicates how many cells have been executed and the total time it took to run the entire code.

**Inclusion of Visuals:**

* At the end of this journey, I was proud of myself because of the result I have achieved.

A close-up of a person's body

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A collage of images of trucks

Description automatically generated.

**Critical Analysis & Referencing:**

* This lab built my curiosity of learning about the numbers it was providing like



* I asked ChatGPT about it and it informed me that this algorithm has 15000 samples or data points for training the model and there are 1024 features or attributes for each sample. For testing, this model has 3000 examples, and each example have 1024 feature.
* When I was doing classification report, I was not able to understand the recall, precision and F1-Score. So, I when I watched the video, I realize it is like a DNA chart we make in bio, where we predict what gene will the offspring get from the parents. So, this code correctly predicted the “frog” 58% of the time and recall is saying that the algorithm was able to identify the “frog” 62% in actual instances. So why the precision and recall are different, the reason is that recall is how many times the algorithm correctly predicted the actual instances while precision is how many predicted instances were correct.

A white paper with writing on it

Description automatically generated

* F1-score is the mean of precision and recall that balances between both metrics. I used all these resources to understand the algorithm

OpenAI. (2023). *Support Vector Machine and dataset dimensions in machine learning*. ChatGPT. Retrieved September 26, 2024, from <https://www.openai.com/chatgpt>

5 Minutes Engineering. (2019a, April 25). *Support Vector Machine (SVM) Part-1 ll Machine Learning Course Explained in Hindi* [Video]. YouTube. <https://www.youtube.com/watch?v=xLkk6MUrvrw>

5 Minutes Engineering. (2019b, April 25). *Support Vector Machine (SVM) Part-2 ll Machine Learning Course Explained in Hindi* [Video]. YouTube. <https://www.youtube.com/watch?v=0MJTaPoHv-g>

5 Minutes Engineering. (2023, September 22). *Precision and Recall Explained in Hindi l Machine Learning* [Video]. YouTube. <https://www.youtube.com/watch?v=n5pD4CglwPk>