Coding Exercise

https://github.com/GabbySuwichaya/Machine-Learning-Course/tree/main/SVM

Exercises: **Multiclass Classification Using SVM.**

SVM does not support multiclass classification natively. Two commonly used approaches that extend SVM for multiclass classification are 1) One-vs-One and 2) One-vs-Rest.

In this exercise, we would like you to apply multiclass classification using SVM to classify number 0-10 from MNIST dataset. Specifically, we would like you to explore the following:

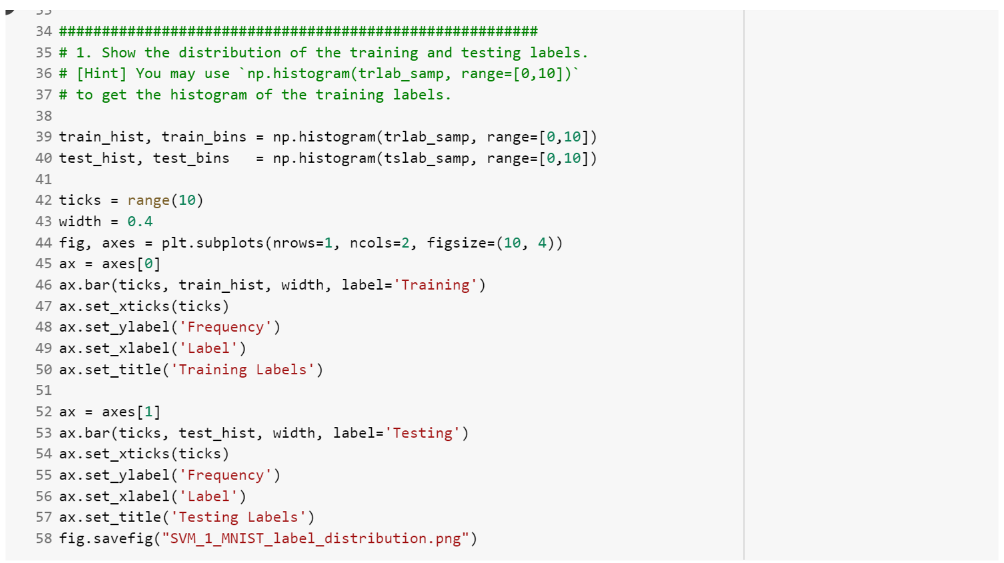
1. MNIST contains around 60,000 samples for training and 10,000 for testing (which can be too large for some computing units).
   a) You may randomly select 6000 samples for training and 1000 sample for testing. Ensure that you have chosen the samples evenly from each class. Then, show us the distribution of labels in the selected training and testing samples.
2. Let's assume that we choose the RBF kernel for SVM. *You may separate your training set for tuning and validation.*
   a) Show the accuracy (or loss) curves across of the validation set across different kernels and model parameters.
   b) Pick the best set of parameters and verify the final performance on the testing dataset.
3. To see the differences between One-vs-one and One-vs-the rest. Let's observe the positive and negative samples.
   a) For One-vs-Rest, what is the number of boundary decision w.r.t. the number of classes? What is the number of binary classifiers in Sklearn (RBF kernel)?
      - Observe the positive and negative samples of the first and the last separation, and any where in the middle.
   b) For One-vs-One, same questions for the number of classifiers and classes.
      - Also, give the same observation.
   c) Can you tell the differences between the observation in (3.a) and (3.b)?
   ➢ For each observation, you may plot the mean shapes of the positive and negative samples & the histogram of the labels associated with the positive and negative samples.

Coding & solutions

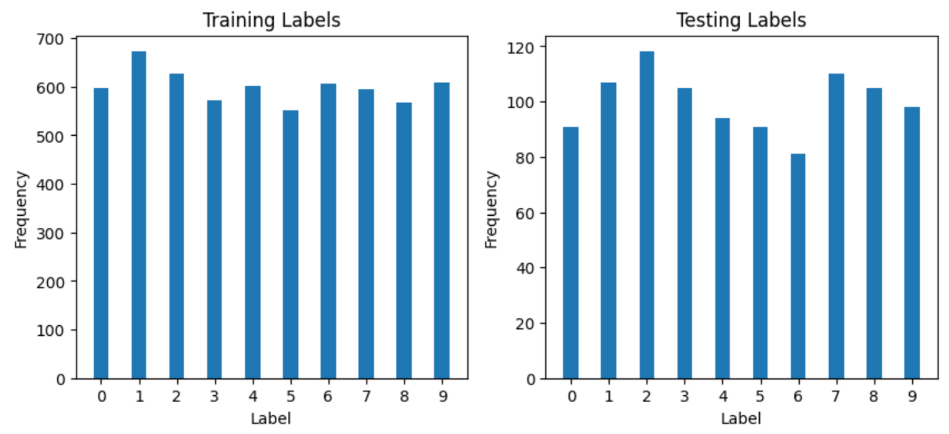
1. MNIST contains around 60,000 samples for training and 10,000 for testing.

You may randomly select 10000 samples for training and 1000 sample for testing.

Then, show us the distribution of the labels in training and testing samples that you selected. Please choose the samples from each class evenly.



```python
#########################################################
# 1. Show the distribution of the training and testing labels.
# [Hint] You may use `np.histogram(trlab_samp, range=[0,10])`
# to get the histogram of the training labels.

train_hist, train_bins = np.histogram(trlab_samp, range=[0,10])
test_hist, test_bins   = np.histogram(tslab_samp, range=[0,10])

ticks = range(10)
width = 0.4
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(10, 4))
ax = axes[0]
ax.bar(ticks, train_hist, width, label='Training')
ax.set_xticks(ticks)
ax.set_ylabel('Frequency')
ax.set_xlabel('Label')
ax.set_title('Training Labels')

ax = axes[1]
ax.bar(ticks, test_hist, width, label='Testing')
ax.set_xticks(ticks)
ax.set_ylabel('Frequency')
ax.set_xlabel('Label')
ax.set_title('Testing Labels')
fig.savefig("SVM_1_MNIST_label_distribution.png")
```

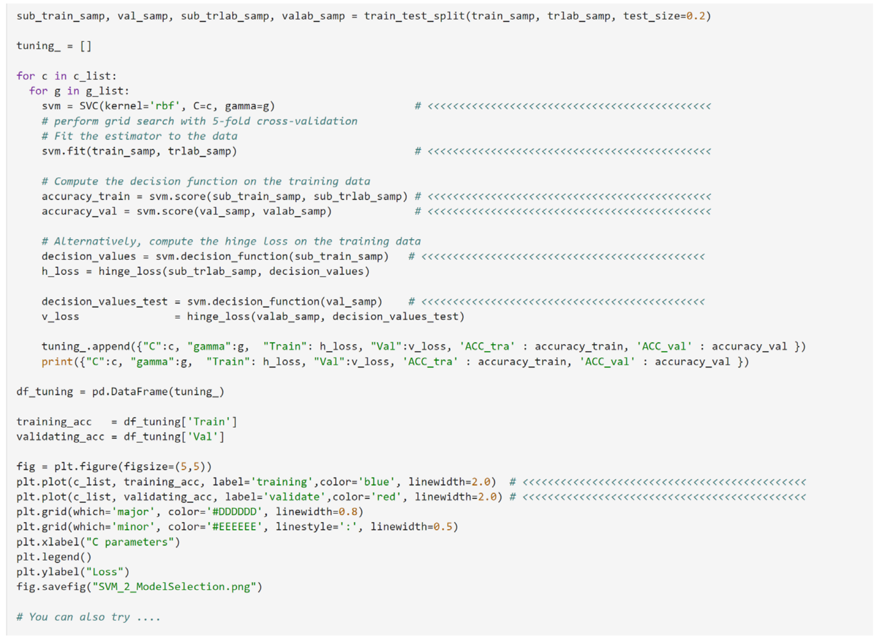


Coding & solutions

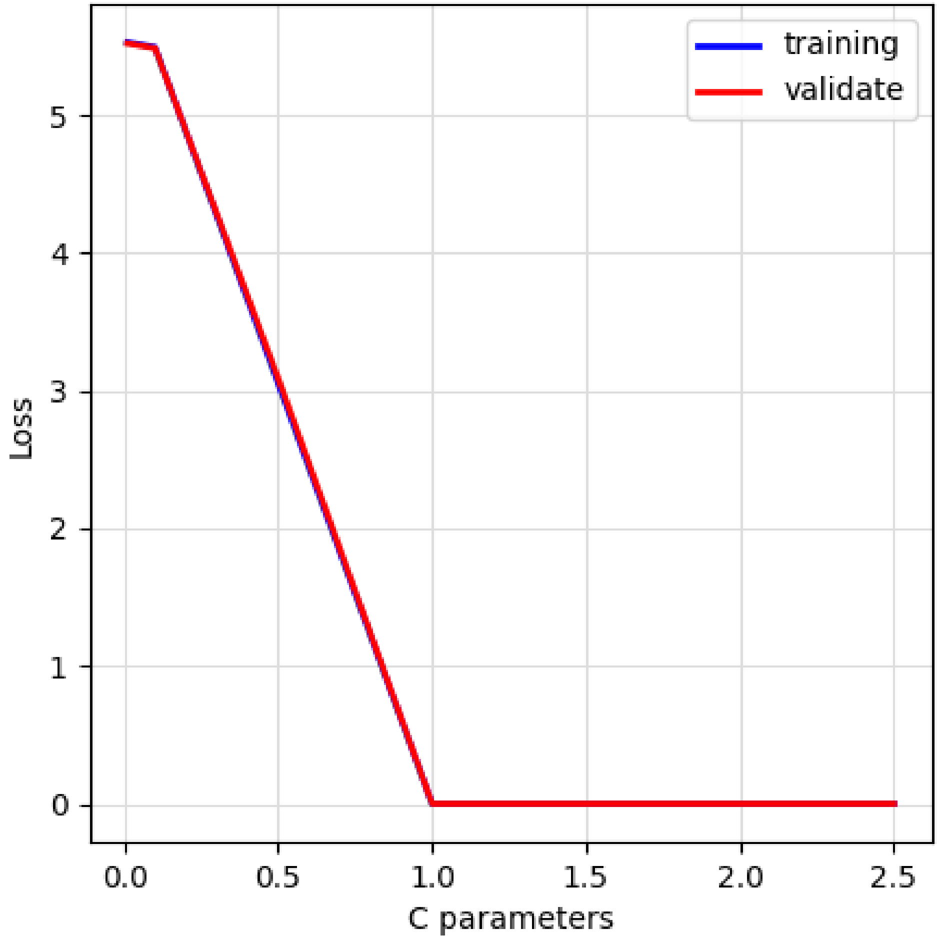
2. Let’s say we choose the RBF kernel SVM. You may separate your training set for tuning and validation. Please show the following results:

a) Show the accuracy (or loss) curves across of the validation set across different kernels and model parameters.

b) Pick the best set of parameters and verify the final performance on the testing dataset.



```
sub_train_samp, val_samp, sub_trlab_samp, valab_samp = train_test_split(train_samp, trlab_samp, test_size=0.2)

tuning_ = []

for c in c_list:
  for g in g_list:
    svm = SVC(kernel='rbf', C=c, gamma=g)                       # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    # perform grid search with 5-fold cross-validation
    # Fit the estimator to the data
    svm.fit(train_samp, trlab_samp)                             # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

    # Compute the decision function on the training data
    accuracy_train = svm.score(sub_train_samp, sub_trlab_samp) # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    accuracy_val = svm.score(val_samp, valab_samp)              # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

    # Alternatively, compute the hinge loss on the training data
    decision_values = svm.decision_function(sub_train_samp)    # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    h_loss = hinge_loss(sub_trlab_samp, decision_values)

    decision_values_test = svm.decision_function(val_samp)     # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    v_loss               = hinge_loss(valab_samp, decision_values_test)

    tuning_.append({"C":c, "gamma":g,  "Train": h_loss, "Val":v_loss, 'ACC_tra' : accuracy_train, 'ACC_val' : accuracy_val })
    print({"C":c, "gamma":g,  "Train": h_loss, "Val":v_loss, 'ACC_tra' : accuracy_train, 'ACC_val' : accuracy_val })

df_tuning = pd.DataFrame(tuning_)

training_acc   = df_tuning['Train']
validating_acc = df_tuning['Val']

fig = plt.figure(figsize=(5,5))
plt.plot(c_list, training_acc, label='training',color='blue', linewidth=2.0)  # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
plt.plot(c_list, validating_acc, label='validate',color='red', linewidth=2.0) # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
plt.grid(which='major', color='#DDDDDD', linewidth=0.8)
plt.grid(which='minor', color='#EEEEEE', linestyle=':', linewidth=0.5)
plt.xlabel("C parameters")
plt.legend()
plt.ylabel("Loss")
fig.savefig("SVM_2_ModelSelection.png")

# You can also try ....
```



Coding & solution

3. To see the differences between One-vs-one and One-vs-Rest. Let’s observe the positive and negative supports.

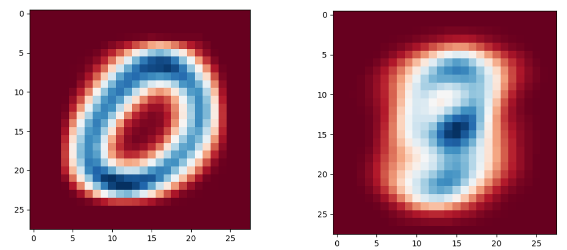
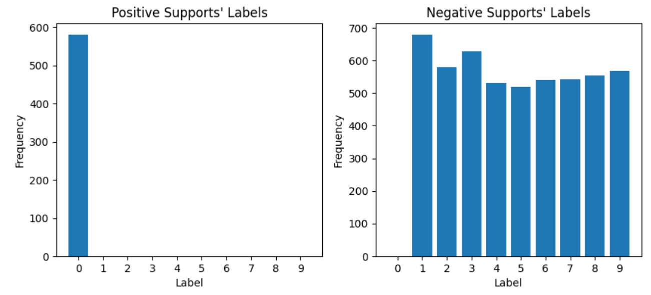
   a) For One-vs-Rest, what is the number of boundary decision with respect to classes? What is the number of binary classifiers in Sklearn (RBF kernel)?

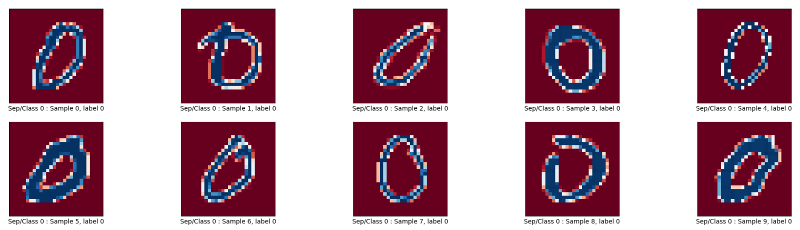
      - Also, observe the positive and negative supports of the first separation, the last separation, and any where inbetween..

ANSWER: The number of binary classifiers is equal to the number of classes. Then the number of boundary decision is **k**. That is, they use the hyperplane to separate one class from the others.

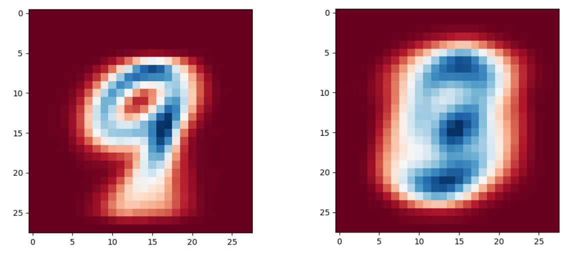
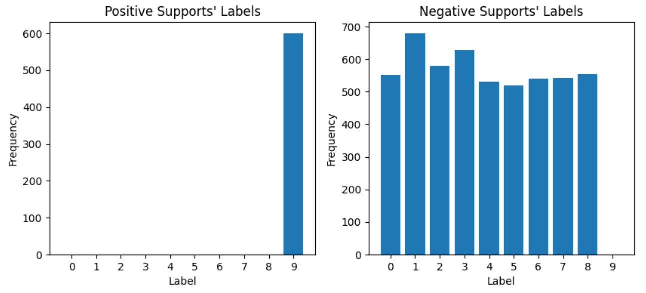
Mean Positive **Mean Negative** **Examples of Positive Supports**

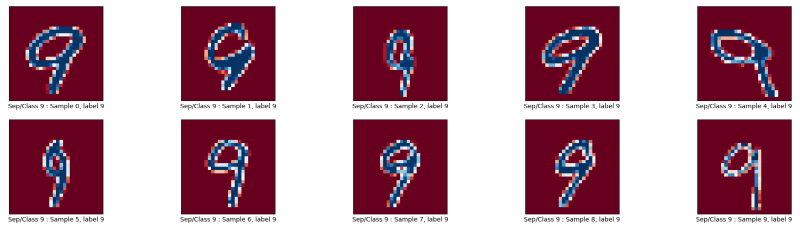
The first separation (the first binary classifier)

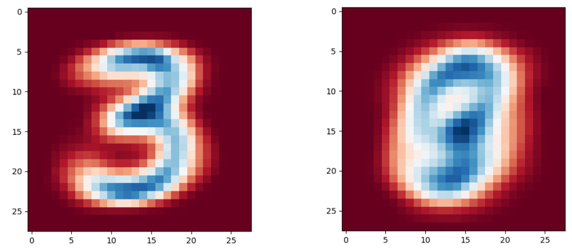
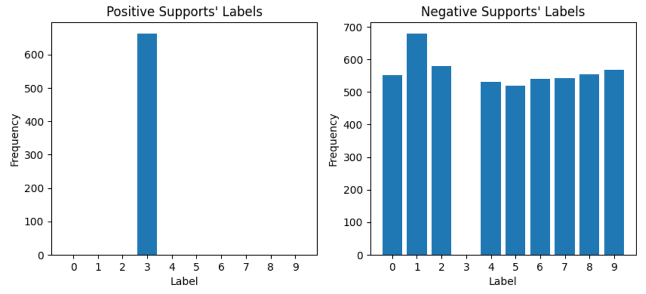


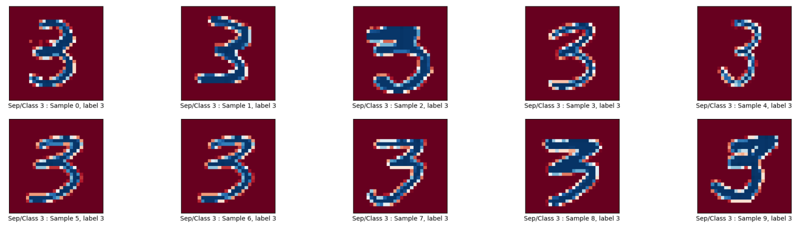
The last separation (the last binary classifier)



The 3rd separation (the 3rd binary classifier)



Coding & solution

3. To see the differences between One-vs-one and One-vs-the rest. Let's observe the positive and negative supports.
   b) For one-vs-one, **what is the number of boundary decisions in association with the number of classes? What is the number of binary classifiers for Sklearn?**
      - Observe the positive and negative supports of the first separation, the last separation, and any where in the middle.

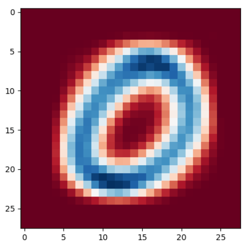
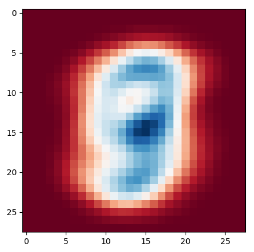
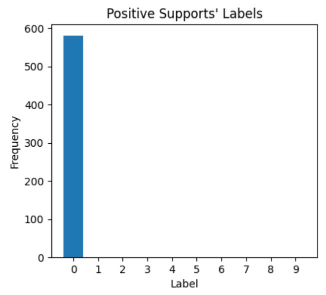
ANSWER: Theoretically, the number of **boundary decisions** for One-vs-One is equal to the total number of possible pairwise combinations between classes (k choose 2), i.e., k*(k-1)/2 = 10*(10-1)/2=5*9=45 for k is the number of classes.

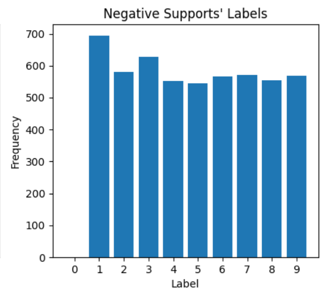
However, non-linear SVM from Sklearn uses **k-1 binary classifiers** that separate a set of classes into positive and negative groups based on the presence or absence of each class in the subsets. For example, if there are 5 classes for {A,B,C,D,E}, it will classify them ({A}, {B,C,D,E}), ({A,B,C,D}, {E}) , ({A,B}, {C,D,E}) , ({A,B,C}, {D,E}). Finally, the multiclass classification result is the intersection of positive and negative group. Therefore, it uses only k-1 binary classifiers.

- We can observe this by selecting each binary classifier and show the negative and positive supports.

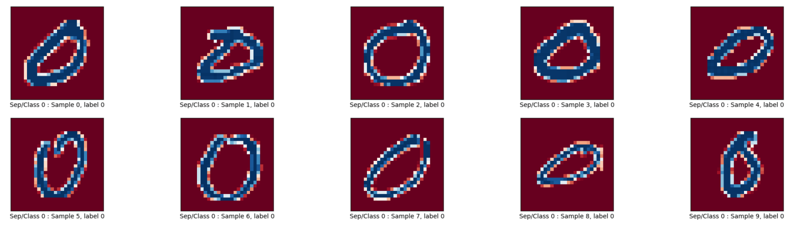
Mean Positive **Mean Negative**

The first separation (the first binary classifier)

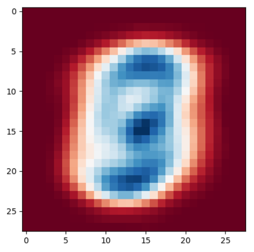
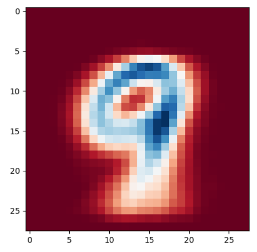
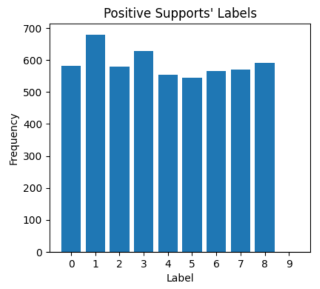




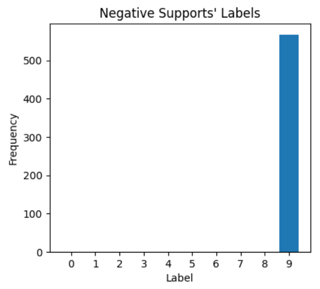


Positive Supports

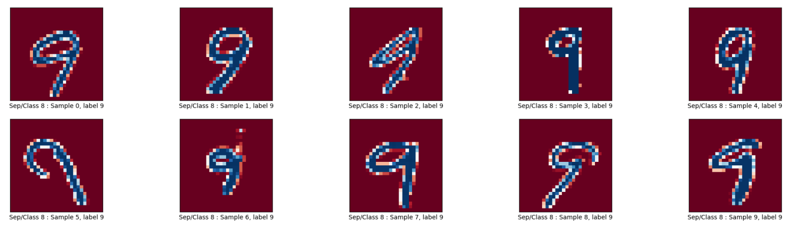


The last separation (the last binary classifier)



Negative Supports



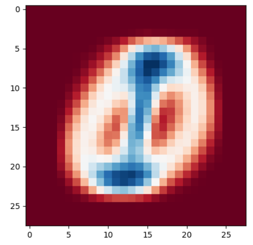
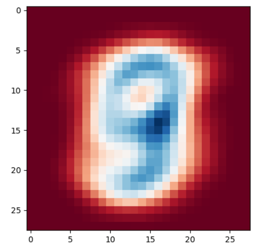
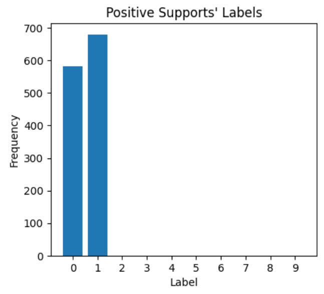
Coding & solution

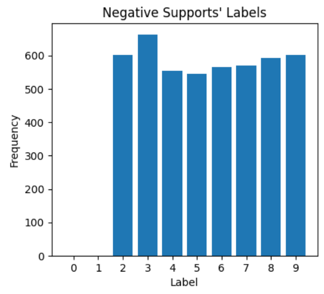
3. To see the differences between One-vs-one and One-vs-the rest. Let's observe the positive and negative supports.
   a) For one-vs-one, **what is the number of boundary decisions in association with the number of classes? What is the number of binary classifiers for Sklearn?**
      - Observe the positive and negative supports of the first separation, the last separation, and any where in the middle.

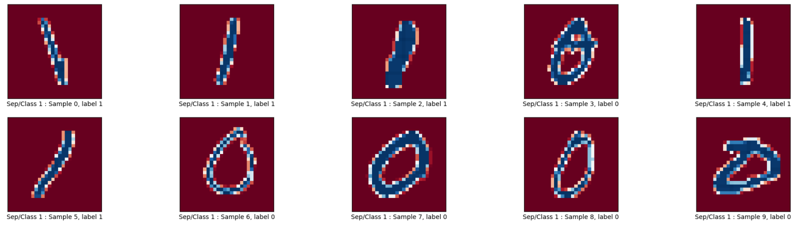
ANSWER: Cont'

Mean Positive **Mean Negative** **Examples of Positive Supports**

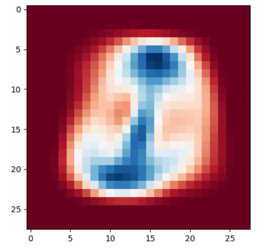
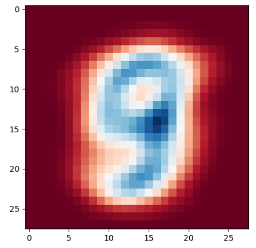
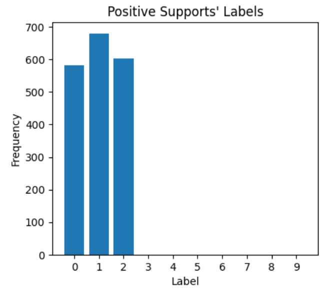
The 2nd separation (the 2nd binary classifier)

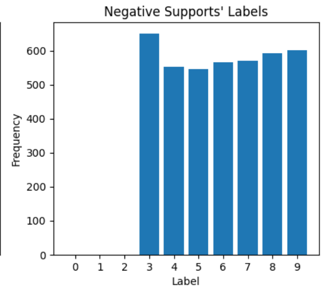




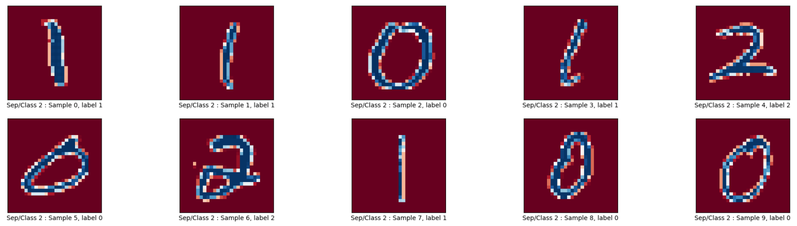



Sep/Class 1 : Sample 0, label 1
Sep/Class 1 : Sample 1, label 1
Sep/Class 1 : Sample 2, label 1
Sep/Class 1 : Sample 3, label 0
Sep/Class 1 : Sample 4, label 1
Sep/Class 1 : Sample 5, label 1
Sep/Class 1 : Sample 6, label 0
Sep/Class 1 : Sample 7, label 0
Sep/Class 1 : Sample 8, label 0
Sep/Class 1 : Sample 9, label 0

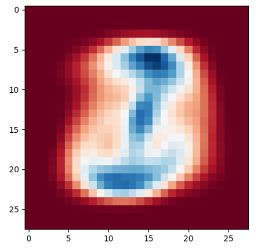
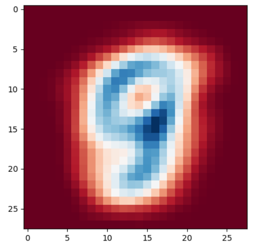
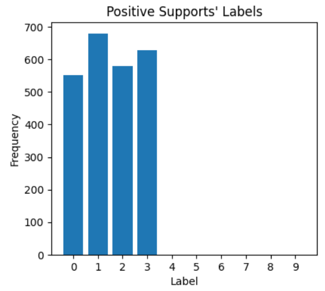
The 3rd separation (the 3rd binary classifier)

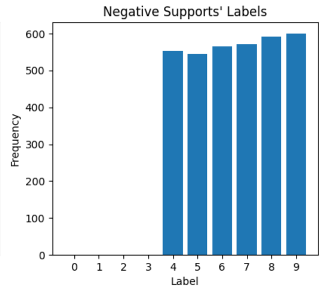




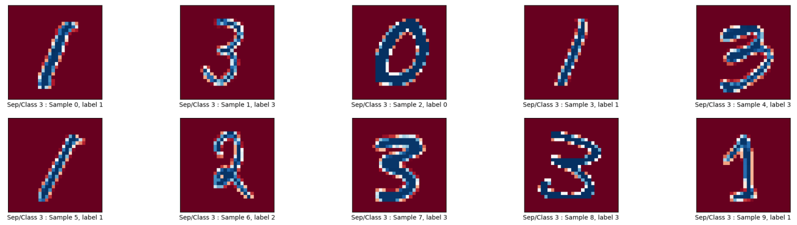



Sep/Class 2 : Sample 0, label 1
Sep/Class 2 : Sample 1, label 1
Sep/Class 2 : Sample 2, label 0
Sep/Class 2 : Sample 3, label 1
Sep/Class 2 : Sample 4, label 2
Sep/Class 2 : Sample 5, label 0
Sep/Class 2 : Sample 6, label 2
Sep/Class 2 : Sample 7, label 1
Sep/Class 2 : Sample 8, label 0
Sep/Class 2 : Sample 9, label 0

The 4th separation (the 4th binary classifier)




Sep/Class 3 : Sample 0, label 1
Sep/Class 3 : Sample 1, label 3
Sep/Class 3 : Sample 2, label 0
Sep/Class 3 : Sample 3, label 1
Sep/Class 3 : Sample 4, label 3
Sep/Class 3 : Sample 5, label 1
Sep/Class 3 : Sample 6, label 2
Sep/Class 3 : Sample 7, label 3
Sep/Class 3 : Sample 8, label 3
Sep/Class 3 : Sample 9, label 1

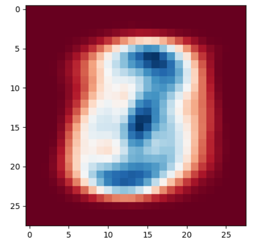
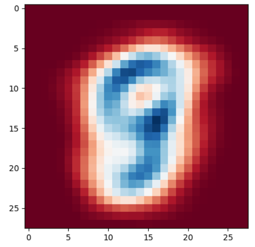
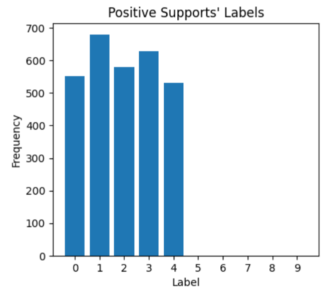
Coding & solution

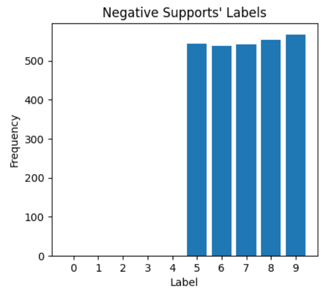
3. To see the differences between One-vs-one and One-vs-the rest. Let's observe the positive and negative supports.
    a) For one-vs-one, **what is the number of boundary decisions in association with the number of classes? What is the number of binary classifiers for Sklearn?**
        - Observe the positive and negative supports of the first separation, the last separation, and any where in the middle.

ANSWER: Cont'

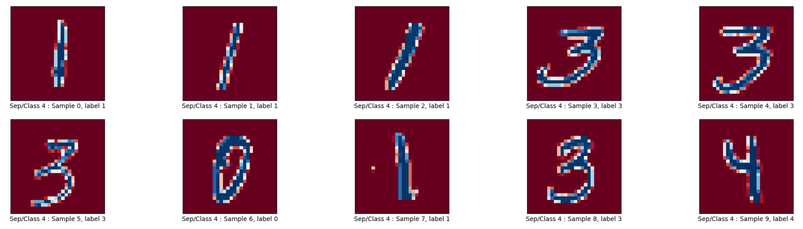
Mean Positive **Mean Negative**

The 5th separation (the 5th binary classifier)

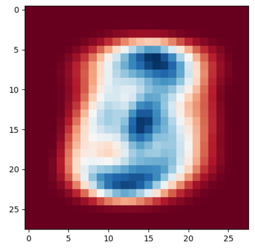
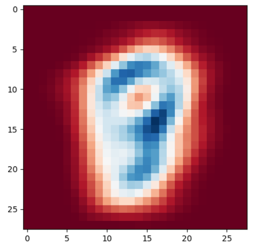
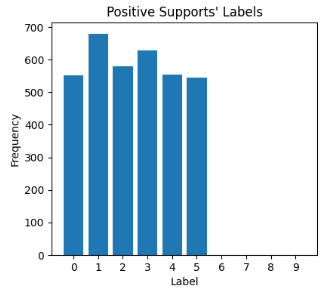




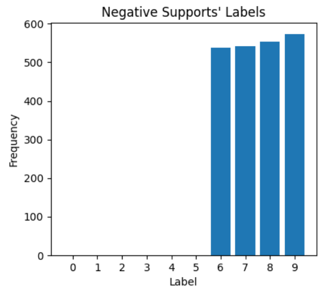


Examples of Positive Supports

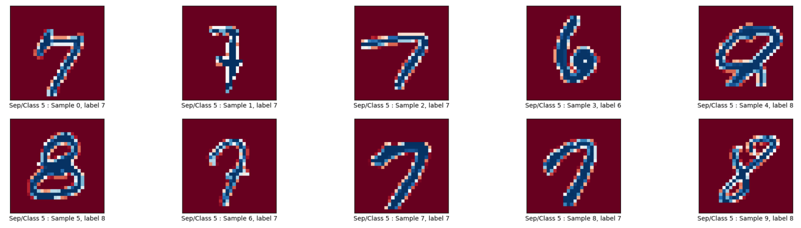

Sep/Class 4 : Sample 0, label 1
Sep/Class 4 : Sample 1, label 1
Sep/Class 4 : Sample 2, label 1
Sep/Class 4 : Sample 3, label 3
Sep/Class 4 : Sample 4, label 3
Sep/Class 4 : Sample 5, label 3
Sep/Class 4 : Sample 6, label 0
Sep/Class 4 : Sample 7, label 1
Sep/Class 4 : Sample 8, label 3
Sep/Class 4 : Sample 9, label 4

The 6th separation (the 6th binary classifier)

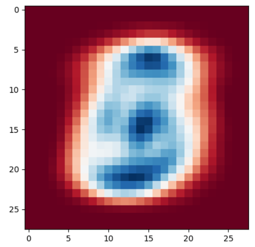
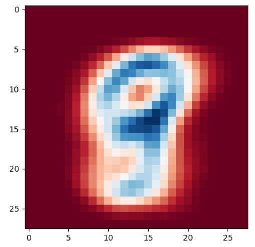
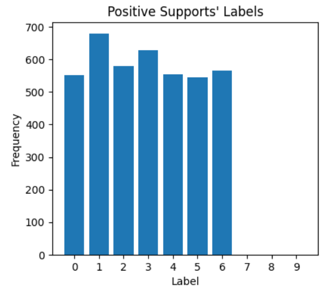




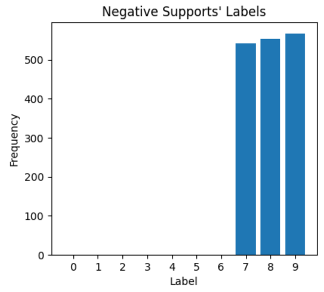


Examples of Negative Supports

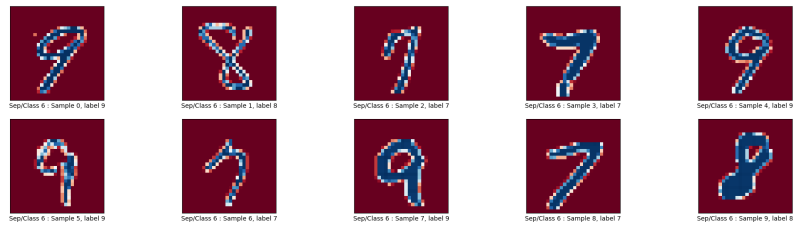

Sep/Class 5 : Sample 0, label 7
Sep/Class 5 : Sample 1, label 7
Sep/Class 5 : Sample 2, label 7
Sep/Class 5 : Sample 3, label 6
Sep/Class 5 : Sample 4, label 8
Sep/Class 5 : Sample 5, label 8
Sep/Class 5 : Sample 6, label 7
Sep/Class 5 : Sample 7, label 7
Sep/Class 5 : Sample 8, label 7
Sep/Class 5 : Sample 9, label 8

The 7th separation (the 7th binary classifier)



Examples of Negative Supports


Sep/Class 6 : Sample 0, label 9
Sep/Class 6 : Sample 1, label 8
Sep/Class 6 : Sample 2, label 7
Sep/Class 6 : Sample 3, label 7
Sep/Class 6 : Sample 4, label 9
Sep/Class 6 : Sample 5, label 9
Sep/Class 6 : Sample 6, label 7
Sep/Class 6 : Sample 7, label 9
Sep/Class 6 : Sample 8, label 7
Sep/Class 6 : Sample 9, label 8

Coding & solution

3. To see the differences between One-vs-one and One-vs-the rest. Let's observe the positive and negative supports.
   a) Can you tell the differences between the observation in (3.a) and (3.b)?
   - ➢ For each observation, you may plot the mean shapes of the positive and negative samples & the histogram of the labels associated with the positive and negative samples.

The difference lies on how the boundary decision is made.

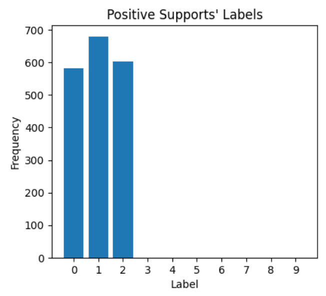
- ➢ For One-vs-rest, the decision is made between one class vs. the rest. Therefore, we can always obtain the number of classifier =k. Also, we can see that the positive support will always corresponding to the class of interest.

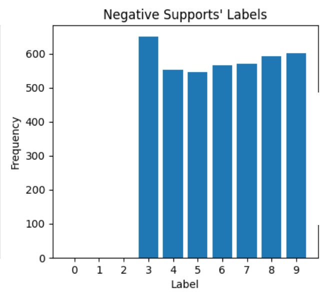
- ➢ Meanwhile, for One-vs-one, the decision is made between pairs of classes. Theoretically, the number of boundary decision will be k*(k-1)/2 for k is the number of classes. In the Sklearn implementation, this is done via the binary classification that separate data into two group of classes, which requires k-1 separation. Finally, the results gathered from the binary classification are used to form the boundary decision.

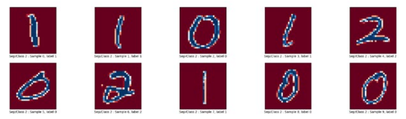
- ➢ We can see this when we look at the positive/negative supports of the binary classifiers. For example, the 2nd to the 2nd to last classifiers can be used to show the differences in their classification mechanisms.

One-vs-One

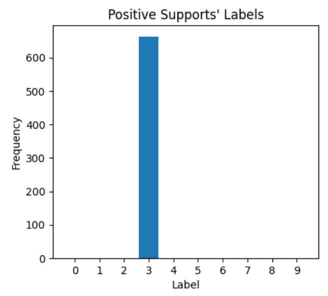
The 3rd separation (the 3rd binary classifier)

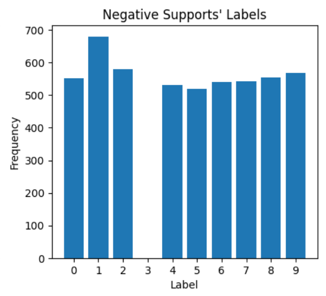






One-vs-Rest


Positive Supports' Labels
Frequency
Label


Negative Supports' Labels
Frequency
Label