Natural Computing Assignment 6 (Domain Adaptation)

Katrin Bujari(s1005213) Simge Ekiz(s4706757), Luca Parola(s1009497)

1-)

We applied Ad-REM algorithm to office-caltech dataset with vgg features and we get the following results. Source dataset was Webcam dataset and target dataset was Caltech256.

```
>> y = predict_adrem(x_src, data.y{1}, x_tgt);
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
Cross Validation Accuracy = 95.3027%
>> mean(y == data.y{2})

ans =

    0.8914
```

We also applied Ad-REM algorithm to office-caltech dataset with surf features and we get the following results. Again, source dataset was Webcam dataset and target dataset was Caltech256.

```
y = predict_adrem(x_src, data.y{1}, x_tgt);
Cross Validation Accuracy = 73.904%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
Cross Validation Accuracy = 68.9979%
>> mean(y == data.y{2})

ans =

    0.4292
```

2-)

We could only run the code with the Ad-REM algorithm on the Office-caltech with VGG and SURF features, due to errors we were not able to fix.
We trained the model on the webcam and we tried to predict the Clatech256.
The Ad-REM algorithm in the first case gave a constant cross validation accuracy of 95.3027% . When predicting on the test set we had a mean score of 0.8914 on the test set. On the other hand when applied the same algorithm on the very same datasets , we had a cross validation score between 68.99 and 73.90 % and we had a much lower mean score of 0.4292

From these results we can see how the difference in the model used in the training phase, affects the final score. We can also see how using the SURF , although it looks decently trained while cross validating, has quite poor performance on the test set compared with the VGG.

3-)

a)We change the Ad-REM code to take the data from another domain and we worked on the office-caltech data with vgg features. We change the input of the Ad-REM as following;

```
addpath src/adrem
addpath src/evaluation

feature = 'vgg';
percentage = 10;

data = load_dataset('office-caltech', feature);

[x_src, x_tgt] = preprocess(data.x{1}, data.y{1}, data.x{2}, 'joint-std');

% Find how many rows we should remove
num_removed_rows = int16((size(x_tgt,1)/100)*percentage);
% Randomly select indices
removed_idx = randperm(size(x_tgt,1), num_removed_rows);

% Add the selected portion from the target to the source
x_src = [x_src; x_tgt(removed_idx,:)];
y_src = [data.y{1}; data.y{2}(removed_idx,:)];

% Remove the selected portion from the target
[x_tgt,~] = removerows(x_tgt,'ind',removed_idx);
[y_tgt,~] = removerows(data.y{2},'ind',removed_idx);

% Predict
y = predict_adrem(x_src, y_src, x_tgt);
mean(y == y_tgt)
```

b) We take the 10% of the Caltech256 data and add it to the Webcam data. And then we train it again. We get the following results;

```
y = predict_adrem(x_src, y_src, x_tgt);
Cross Validation Accuracy = 95.4206%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
Cross Validation Accuracy = 95.1402%
>> mean(y == y_tgt)
 0.8932
```

c)

1st apply	2nd apply	3rd apply	4th apply	5th apply	Average	Standard Deviation

0.8932	0.8922	0.8922	0.8922	0.8922	0.8924	0.0004

d) We repeated the same steps for 20%, 40% and 60% and we got the results in section e.
e)

Random sample size	Average	Standard Deviation
10%	0.8924	0.0004
20%	0.9042	0
40%	0.8872	0
60%	0.8953	0