Machine Learning – Exercise 3

Lavi Lazarovitz – 065957383

Aharon Sharim – 052328523

# General notes

Part2\_bonus1.py is running an automated process to train and select a model for classification.

Part2\_bonus3.py is showing the manipulations result on the prediction of the trained model.

# Extra question a

We automated the process we used to select the model in base exercise:

1. Train the classifiers (Svm, Linear svc, One vs One, Guassian NB, Decision tree and K-nearest neighbors) on a 20 split of the training data.
2. Calculate average score over the training
3. Take the best 2 and retrain them, using a 30 split.
4. Take the best one and fully train it.

# Extra question c

We decided to train a tree model, and checked the most important features it recognized, as a tree model tends to select first the features with most "impact" on the classification.

Then we tried to manipulate those features on the validation data to check the manipulated results.

The tree model gave back this array of feature importance:

array([ **0.13763966,**

0.04784999,

**0.26060406,**

0.09384121,

**0.19467166,**

**0.14047433,**

0.08807282,

0.00821176,

0.00247503,

0.00115261,

0.00353883,

0.00047635,

0.01749513,

0.00188829,

0.00160827])

The bolded numbers are the most important ones.

On those we tried several manipulations on the data to see if we can get different predictions.

The original predicted winner is Purples.

After increasing Overall\_happiness\_score by 1.75 (All values are normalized), the predicted winner is Greens.

After decreasing Overall\_happiness\_score by 0.25, the predicted winner is Browns.

After changing Will\_vote\_only\_large\_party\_int to 1, the predicted winner is Yellows.

After changing Will\_vote\_only\_large\_party\_int to 0, the predicted winner is Oranges.

After increasing Garden\_sqr\_meter\_per\_person\_in\_residancy\_area by 2, the predicted winner is Whites.

After decreasing Number\_of\_valued\_Kneset\_members by 0.25, the predicted winner is Browns.