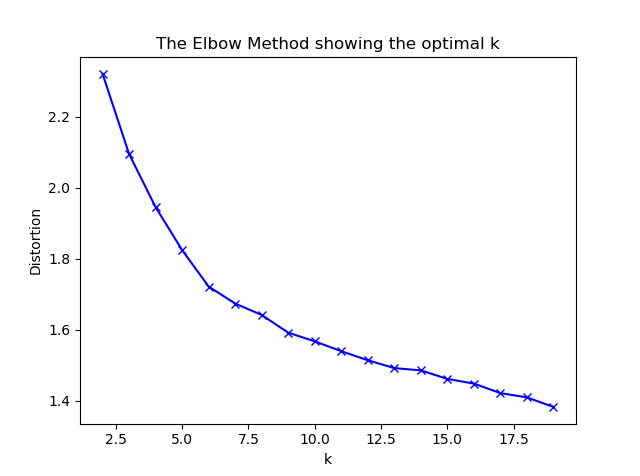
# Exercise 4

# Stable coalition

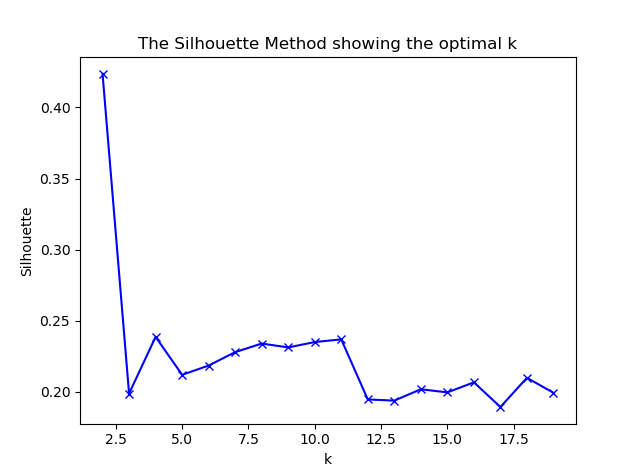
First, we try to find clustering in the data. By identifying clusters which have a "good" mix of two or more parties, we might be able to identify parties which are 'close together' in their features.

We first decided to train k-mean. For that, we want to figure out how many clusters we should pick, as it's not guaranteed that each party is clustered around one point.

For that we use both the elbow method:



And the Silhouette method:



From this we conclude we should probably use 8-11 clusters.

After training the k-means model with 11 clusters, we want to see the spread of the parties across the clusters:

## Example output of count of parties in each cluster:

'0': {'Reds': 320, 'Yellows': 36, 'Greys': 322, 'Oranges': 307},

'1': {'Pinks': 8, 'Yellows': 39, 'Purples': 13, 'Greens': 983},

'2': {'Whites': 78, 'Browns': 665, 'Pinks': 241, 'Yellows': 13, 'Purples': 68},

'3': {'Blues': 116, 'Yellows': 9},

'4': {'Pinks': 394, 'Yellows': 18},

'5': {'Blues': 126, 'Yellows': 1},

'6': {'Whites': 109, 'Browns': 434, 'Pinks': 2, 'Purples': 185},

'7': {'Blues': 111, 'Yellows': 2},

'8': {'Whites': 13, 'Yellows': 25, 'Oranges': 1, 'Browns': 4, 'Pinks': 13, 'Purples': 967}

'9': {'Blues': 70, 'Yellows': 104},

'10': {'Blues': 122, 'Yellows': 81},

## Example output of spread of parties across clusters:

'Reds': {'0': 320},

'Greens': {'1': 983},

'Whites': {'8': 13, '2': 78, '6': 109},

'Yellows': {'10': 81, '1': 39, '0': 36, '3': 9, '2': 13, '5': 1, '4': 18, '7': 2, '9': 104, '8': 25},

'Greys': {'0': 322},

'Oranges': {'0': 307, '8': 1},

'Browns': {'8': 4, '2': 665, '6': 434},

'Blues': {'9': 70, '10': 122, '3': 116, '5': 126, '7': 111},

'Pinks': {'1': 8, '8': 13, '2': 241, '4': 394, '6': 2},

'Purples': {'1': 13, '8': 967, '2': 68, '6': 185}}

We also calculated the distances between the clusters' centers, to figure out which clusters were close to each other and which were far (by using the median distance as a yardstick).

## Example output clusters distances:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster distances** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **0** | Close | Close | Close | Far | Close | Far | Close | Far | Close | Close | Close |
| **1** | Close | Close | Close | Far | Far | Far | Close | Far | Close | Close | Far |
| **2** | Close | Close | Close | Far | Far | Far | Close | Far | Close | Close | Far |
| **3** | Far | Far | Far | Close | Far | Far | Far | Far | Far | Far | Far |
| **4** | Close | Far | Far | Far | Close | Far | Close | Far | Close | Far | Far |
| **5** | Far | Far | Far | Far | Far | Close | Far | Far | Far | Close | Far |
| **6** | Close | Close | Close | Far | Close | Far | Close | Far | Close | Far | Close |
| **7** | Far | Far | Far | Far | Far | Far | Far | Close | Far | Far | Far |
| **8** | Close | Close | Close | Far | Close | Far | Close | Far | Close | Far | Close |
| **9** | Close | Close | Close | Far | Far | Close | Far | Far | Far | Close | Close |
| **10** | Close | Far | Far | Far | Far | Far | Close | Far | Close | Close | Close |

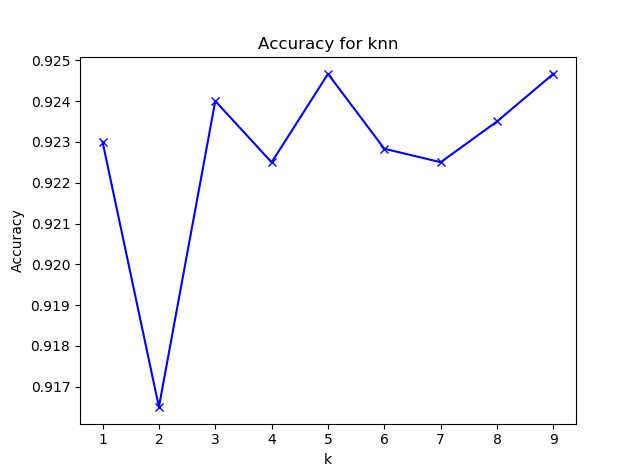
We ran this process several times to see the results.

### From this we conclude the following:

1. Yellows and Blues are pretty spread out, but tend to cluster together.
2. Reds, greys and oranges are very similar, almost always end up in one cluster, and are not too far away from the other two main groups.
3. Whites, browns, pinks and purples are pretty close together.
4. Greens seems to be in their own cluster usually, perhaps due to size, and close to browns.

Next, we try the k-nearest-neighbors clustering algorithm.

First, we run it with in classification mode, with 5-fold cross validation to figure out what would be a good K to set:



We decide to select K = 5, to get good accuracy on one side, but ability to "spread out" the neighbors for each point.

Then, we count for each point how many neighbors it has from each party, and sum over all the parties:

## Example output for KNN:

'Reds': {'Reds': 1146, 'Greys': 64, 'Oranges': 70}

'Greens': {'Greens': 3932}

'Whites': {'Whites': 516, 'Browns': 219, 'Pinks': 40, 'Purples': 25}

'Yellows': {'Blues': 243, 'Yellows': 1069}

'Greys': {'Reds': 69, 'Greys': 1145, 'Oranges': 74}

'Oranges': {'Reds': 145, 'Greys': 113, 'Oranges': 974}

'Browns': {'Whites': 111, 'Browns': 4182, 'Pinks': 98, 'Purples': 21}

'Blues': {'Blues': 1847, 'Yellows': 333}

'Pinks': {'Whites': 50, 'Browns': 345, 'Pinks': 2190, 'Purples': 47}

'Purples': {'Whites': 29, 'Browns': 29, 'Pinks': 64, 'Greens': 13, 'Purples': 4797}

We see that we get similar results to the first method, which we use to strengthen our conclusions.

## Coalition:

We suggest a narrow coalition of browns-purples-pinks-whites which should be higher than 50% of the votes, and contains 4 parties which are pretty close together.

This, in our opinion, would be stronger than the alternatives:

Greens are a large party, and pretty distinct and even their closest "ally" Browns seems to not be that close to them.

Red-Gray-Orange could almost be seen as one party.

Blue and Yellows are medium sized parties, but tend to be spread out in their locations, so it's hard for the parties to be close to them in view. Any coalition containing them would need much more support from bigger, "tighter" parties.

Brown and Purple are large and close together and so can provide a solid core for a coalition, and they have enough of close "allies", clustered with them or very close, to get above 51%.

# Major features

To identify the major features for each party, we trained a Naive Bayes model, and for each feature, checked the mean and the variance for each party.

We tried to understand which features have distinct means for each party, preferably with low variance.

For example:

This most obvious feature is "Will\_vote\_only\_large\_party\_int", where each party has pretty much only one value, and so it is a good separator:

Blues and Yellows have "Yes", Greys/Organges/Reds have "Maybe", other parties rest have "No".

Blues have high Avg\_Satisfaction\_with\_previous\_vote, which is low for everyone else except Yellows, where it is around 0 (normalized).

Features we recognized for each party (For "most important issue" we don’t give the full issue name, as this is a technical analysis anyway):

### Blues

Will\_vote\_only\_large\_party\_int, Avg\_Satisfaction\_with\_previous\_vote, Number\_of\_valued\_Kneset\_members, Weighted\_education\_rank

### Purples

Will\_vote\_only\_large\_party\_int, Number\_of\_valued\_Kneset\_members, Yearly\_IncomeK

### Browns

Will\_vote\_only\_large\_party\_int, Yearly\_IncomeK

### Greens

Will\_vote\_only\_large\_party\_int, Number\_of\_valued\_Kneset\_members, Overall\_happiness\_score, Weighted\_education\_rank, Garden\_sqr\_meter\_per\_person\_in\_residancy\_area

### Pinks

Will\_vote\_only\_large\_party\_int, Weighted\_education\_rank

### Yellows

Will\_vote\_only\_large\_party\_int, Avg\_Satisfaction\_with\_previous\_vote, Number\_of\_valued\_Kneset\_members, Weighted\_education\_rank

### Reds

Will\_vote\_only\_large\_party\_int, Most\_Important\_Issue, Yearly\_IncomeK

### Greys

Will\_vote\_only\_large\_party\_int, Most\_Important\_Issue, Yearly\_IncomeK

### Oranges

Will\_vote\_only\_large\_party\_int, Most\_Important\_Issue, Yearly\_IncomeK

### Whites

Will\_vote\_only\_large\_party\_int, Yearly\_IncomeK

# Feature manipulation

We used two types of inference for this part.

We've shown one option in the extra assignment for the previous exercise, where we trained and used a decision tree to identify the most significant features.

The features we identified were:

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.

The second option

# Strengthening the coalition

# 