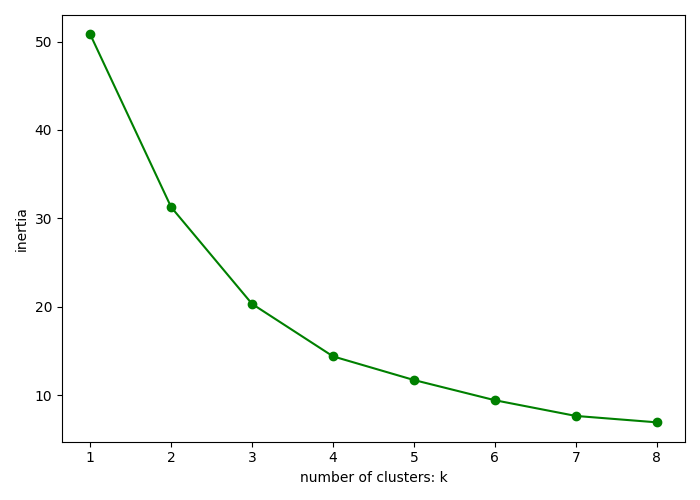
# Assignment 10

## k-means

* Question 1: by visual check the plot below, optimal k is 2



* Question 2:
  1. Cluster 0: percentage of “green”0%, percentage of “red”100%
  2. Cluster 1: percentage of “green”100%, percentage of “red”0%
* Question 3:

Cluster 1 is a pure cluster, all points in cluster 1 are labeled as green.

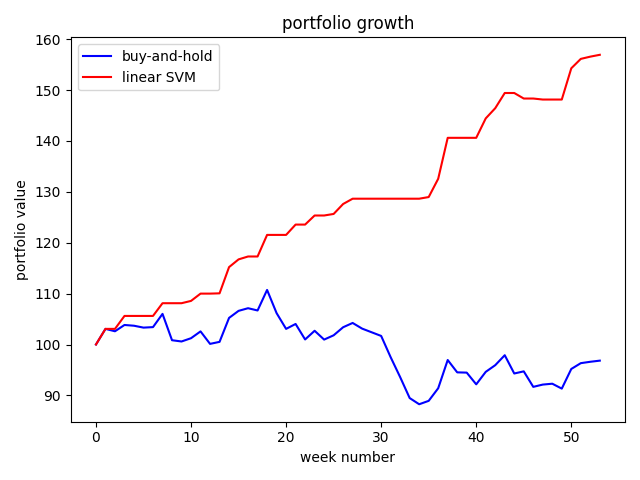
## SVM

* Question 1: SVM accuracy 96.23%
* Question 2: confusion matrix

[[27 2]

[ 0 24]]

* Question 3: TPR 93.1% TNR: 100%
* Question 4: Gaussian SVM accuracy 96.23%, equal to linear SVM
* Question 5: polynomial SVM accuracy 50.94%, lower than linear SVM
* Question 6: plot, of cause the SVM strategy results in a larger amount



## Compare classifier

|  |  |  |
| --- | --- | --- |
| Classifier | Accuracy | Portfolio Value for Year 2 |
| Buy and Hold | 100% | 96.8293 |
| kNN with best k=7 | 94% | 157.1482 |
| logistic | 98% | 158.0046 |
| Naïve bayesian | 94% | 157.9337 |
| Linear models (best degree, 3) | 58% |  |
| Decision tree | 92% | 150.6385 |
| Random Forest | 100% | 157.3832 |
| Linear discriminant | 96% | 156.9372 |
| Quadratic discriminant | 98% | 157.2326 |
| SVM (linear) | 96% | 156.9372 |
| Ada Boost |  |  |

what are the best 3 classifiers by accuracy: random forest, logistic, QDA

what are the worst 3 classifiers by accuracy: linear model, decision tree, kNN and naïve bayesian