DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

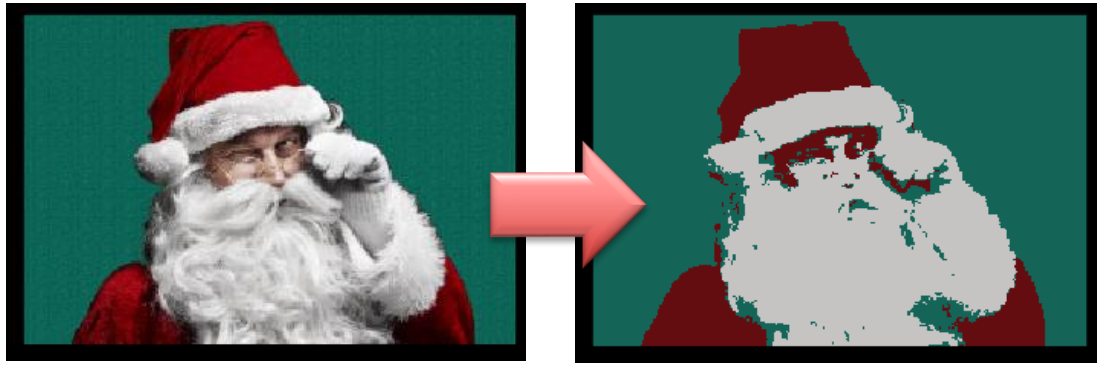
Homework 3

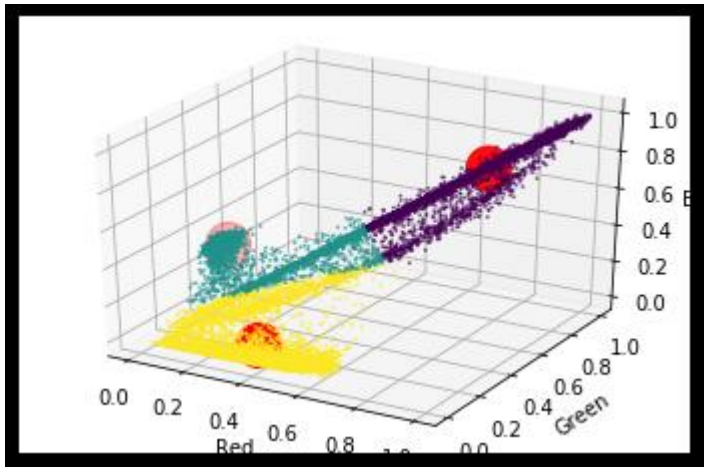
Machine Learning 2023

(P. Zanuttigh – ICT+Physics of Data)

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Clustering and Image Segmentation





- Implement the k-means algorithm
  - The initialization is critical (use numpy randn function with mean and standard deviation of the data, good to scale std.dev, e.g. divide by 4)
  - Avoid empty clusters
- Use it for image segmentation
- See the results with a different number of clusters
- Compare with sklearn implementation ("inertia" is the error/cost function)
- Test linkage based approaches (called "agglomerative") in sklearn
- Evaluate the differences between k-means and linkage based approaches

Your Task

- Complete the jupyter notebook
  - FIRST THING TO DO: you need to put your name and ID number in the notebook
  - You can use the ID also as seed for random number generators, try different seeds
  - The notebook has missing code: need to fill in what is missing
  - You must write the answer to all the questions in the notebook
  - But do not change the structure or the input data files, they will not be submitted
- Check that the notebook run properly from the beginning with the provided data
  - use the `"restart kernel&run all"` command
- Save them as `surname_name_lab3.ipynb`
- Submit on elearning

jupyter

Timeline

- Fri 1/12: Homework released and Lab 3 (room Ue)
- ***Thu 14/12: Delivery deadline***
- The outcome is an on-off mark (i.e., +1 for the exam mark if the homework is reasonably done)