

## Wine dataset

<https://archive.ics.uci.edu/ml/datasets/wine>

Datan finns antingen att ladda hem från länken eller ännu lättare importera den: from sklearn.datasets import load\_wine

```
data = load_wine(as_frame = True)

print(type(data))
✓ 0.5s

<class 'sklearn.utils.Bunch'>

# Convert to DataFrame
data1 = pd.DataFrame(data= np.c_[data['data'], data['target']],
                    columns= data['feature_names'] + ['target'])
data1.head()
```

Target är kolumnen 'target' och det är tre klasser 0, 1, 2 som berättar vilken kvalitet det är på vinet.

Använd en valfri classifier för att göra predictions på datasetet. Dela upp i test/train set med sklearn och gör 0.2-0.3 av datan till testset.

- Spara resultaten av dina predictions i en confusion matrix.
- Använd UMAP eller t-SNE för att klustra dina feature vectors och ge dem labels.
- Tolka dina visuella resultat och se ifall något par av klasser kan misstas för varandra.
- Avgör ifall dina resultat ifrån UMAP/t-SNE och din confusion matrix hänger ihop.

Gör en PCA-transform på datasetet och plotta den totala förklarade variansen för summan utav PC1 + PC2 + PC3... PCx för alla x mellan 2 och 14.

Gör slutligen en grid search med flera kombinationer av hyperparametrar och antal komponenter från PCA. Ser modellen ut att bli bättre av dimensionsreduceringen?