Machine Learning Homework 3

February 25, 2025

Gabe Morris

```
[1]: # toc
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.datasets import load_breast_cancer

plt.style.use('../maroon_ipynb.mplstyle')

cancer = load_breast_cancer()
features = cancer.data
target = cancer.target
```

Contents

Problem 1	3	
Solution	3	

Problem 1

The sklearn library includes a set of data containing image information on fine needle aspirates used to identify breast cancer. The dataset contains information gathered from the images as well as whether the mass was malignant (target=0) or benign (target=1).

```
[2]: # Showing the features of the dataset
feature_names = cancer.feature_names
temp_df = pd.DataFrame(features, columns=feature_names)
temp_df.head(n=4).transpose()
```

-				
	0	1	2	3
mean radius	17.990000	20.570000	19.690000	11.420000
mean texture	10.380000	17.770000	21.250000	20.380000
mean perimeter	122.800000	132.900000	130.000000	77.580000
mean area	1001.000000	1326.000000	1203.000000	386.100000
mean smoothness	0.118400	0.084740	0.109600	0.142500
mean compactness	0.277600	0.078640	0.159900	0.283900
mean concavity	0.300100	0.086900	0.197400	0.241400
mean concave points	0.147100	0.070170	0.127900	0.105200
mean symmetry	0.241900	0.181200	0.206900	0.259700
mean fractal dimension	0.078710	0.056670	0.059990	0.097440
radius error	1.095000	0.543500	0.745600	0.495600
texture error	0.905300	0.733900	0.786900	1.156000
perimeter error	8.589000	3.398000	4.585000	3.445000
area error	153.400000	74.080000	94.030000	27.230000
smoothness error	0.006399	0.005225	0.006150	0.009110
compactness error	0.049040	0.013080	0.040060	0.074580
concavity error	0.053730	0.018600	0.038320	0.056610
concave points error	0.015870	0.013400	0.020580	0.018670
symmetry error	0.030030	0.013890	0.022500	0.059630
fractal dimension error	0.006193	0.003532	0.004571	0.009208
worst radius	25.380000	24.990000	23.570000	14.910000
worst texture	17.330000	23.410000	25.530000	26.500000
worst perimeter	184.600000	158.800000	152.500000	98.870000
worst area	2019.000000	1956.000000	1709.000000	567.700000
worst smoothness	0.162200	0.123800	0.144400	0.209800
worst compactness	0.665600	0.186600	0.424500	0.866300
worst concavity	0.711900	0.241600	0.450400	0.686900
worst concave points	0.265400	0.186000	0.243000	0.257500
worst symmetry	0.460100	0.275000	0.361300	0.663800
worst fractal dimension	0.118900	0.089020	0.087580	0.173000
	mean texture mean perimeter mean area mean smoothness mean compactness mean concavity mean concave points mean symmetry mean fractal dimension radius error texture error perimeter error area error smoothness error compactness error concavity error concave points error symmetry error fractal dimension error worst radius worst texture worst perimeter worst area worst smoothness worst concave points worst concave points worst concave points worst concave points worst symmetry	mean radius 17.990000 mean texture 10.380000 mean perimeter 122.800000 mean area 1001.000000 mean smoothness 0.118400 mean compactness 0.277600 mean concavity 0.300100 mean concave points 0.147100 mean symmetry 0.241900 mean fractal dimension 0.078710 radius error 1.095000 texture error 0.905300 perimeter error 3.589000 area error 0.006399 compactness error 0.049040 concavity error 0.053730 concave points error 0.015870 symmetry error 0.030030 fractal dimension error 0.006193 worst radius 25.380000 worst texture 17.330000 worst smoothness 0.162200 worst compactness 0.665600 worst concavity 0.711900 worst concave points 0.265400 worst symmetry 0.460100	mean radius 17.990000 20.570000 mean texture 10.380000 17.770000 mean perimeter 122.800000 132.900000 mean area 1001.000000 1326.000000 mean smoothness 0.118400 0.084740 mean compactness 0.277600 0.078640 mean concavity 0.300100 0.086900 mean concave points 0.147100 0.070170 mean symmetry 0.241900 0.181200 mean fractal dimension 0.078710 0.056670 radius error 1.095000 0.543500 texture error 0.905300 0.733900 perimeter error 8.589000 3.398000 area error 153.40000 74.080000 smoothness error 0.006399 0.005225 compactness error 0.049040 0.013080 concavity error 0.053730 0.013400 symmetry error 0.030030 0.013890 fractal dimension error 0.006193 0.003532 worst radius 25.380	mean radius 17.990000 20.570000 19.690000 mean texture 10.380000 17.770000 21.250000 mean perimeter 122.800000 132.900000 130.00000 mean area 1001.00000 1326.00000 1203.00000 mean smoothness 0.118400 0.084740 0.109600 mean compactness 0.277600 0.078640 0.159900 mean concavity 0.300100 0.086900 0.197400 mean symmetry 0.241900 0.070170 0.127900 mean fractal dimension 0.078710 0.056670 0.05990 radius error 1.095000 0.543500 0.745600 texture error 0.905300 0.733900 0.786900 perimeter error 8.589000 3.398000 4.585000 area error 153.40000 74.08000 94.03000 smoothness error 0.049040 0.013080 0.04066 concavity error 0.0553730 0.018600 0.038320 concave points error 0.03030 0.013890

Perform a PCA transformation of the features and plot the individual and cumulative explained variable as a function of number of components.

Solution