

Homework 5

Shao-Ting Chiu (UIN:433002162)

1/20/23

Table of contents

Description	1
Problem 9.8	1
(a)	2
(b)	2
(c)	2
(d)	2
Problem 11.11	2
(a)	2
(b)	2

Description

Problems from the Book

9.8

11.11

Both problems are coding assignments, with starting code provided. Each is worth 40 points.

Problem 9.8

This assignment concerns the application of PCA to the soft magnetic alloy data set (See section A8.5).

(a)

Reproduce the plots in Figure 9.5 by running `c09_PCA.py`

(b)

Plot the percentage of variance explained by each PC as a function of PC number. This is called the *scree plot*. Now plot the cumulative percentage of variance explained by the PCs as a function of PC number. How many PCs are needed to explain 95% of the variance.

Coding hint: use the attribute `explained_variance_ratio_` and the `cusum()` method.

(c)

Print the loading matrix W (this is the matrix of eigenvectors, ordered by PC number from left to right). The absolute value of the coefficients indicate the relative importance of each original variable (row of W) in the corresponding PC (column of W).

(d)

Identify which two features contribute the most to the discriminating first PC and plot the data using these top two features. What can you conclude about the effect of these two features on the coercivity? This is an application of PCA to feature selection.

Problem 11.11

Apply linear regression to the stacking fault energy (SFE) data set.

(a)

(b)