MAD Assignment 3

Ask Jensen

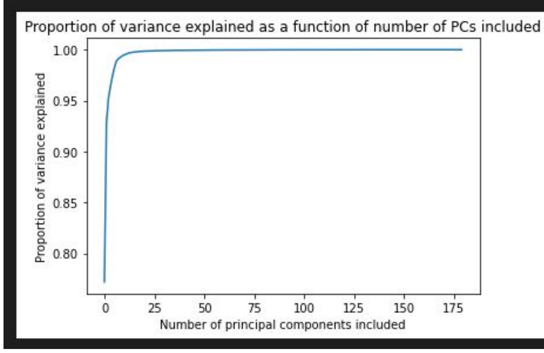
13. december 2021

Indhold

1 Problem 1

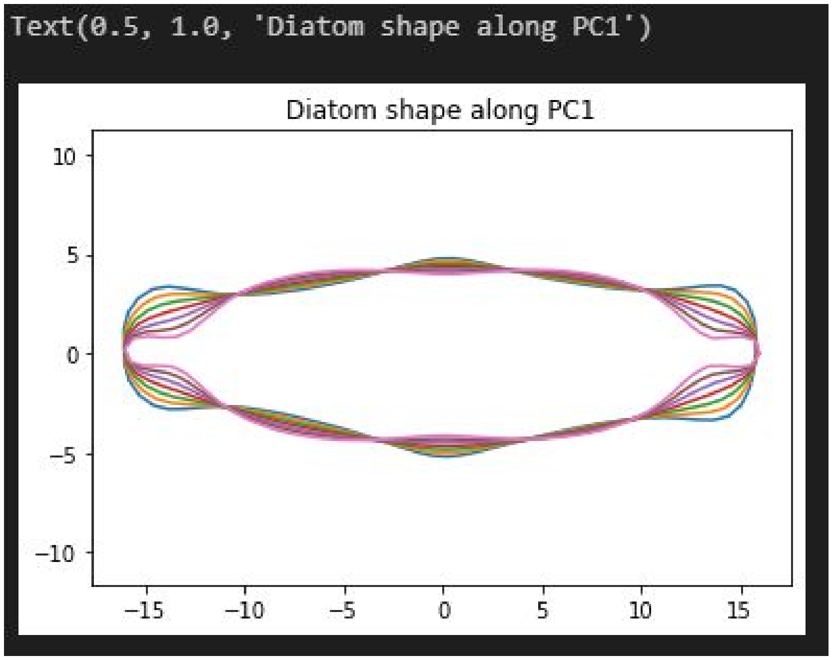
1.1 (a)

```
Proportion of variance explained by the first 1 principal components: 0.7718721493017527
Proportion of variance explained by the first 2 principal components: 0.9276996293043024
Proportion of variance explained by the first 3 principal components: 0.9521198453942007
Proportion of variance explained by the first 4 principal components: 0.9637878603999529
Proportion of variance explained by the first 5 principal components: 0.9739084497954094
Proportion of variance explained by the first 6 principal components: 0.98236065164916
Proportion of variance explained by the first 7 principal components: 0.9889975933245944
Proportion of variance explained by the first 8 principal components: 0.9910287023941854
Proportion of variance explained by the first 9 principal components: 0.9926692113360289
Proportion of variance explained by the first 10 principal components: 0.9939926229665051
```



Figur 1: Values for the first 10 proportions of variance, and the corrosponding graph

1.2 (b)



Figur 2: Plotted diatom

2 Problem 2

assesses the given claim $E[(X-\mu)^4] \geq \sigma^4$

X has the mean μ and the variance σ^4 which can be rewritten as $\sigma^4 = (Var(X))^2$
$E[(X-\mu)^4]$ kan be rewritten as $E(g(x))$ where $g(x) = (x-\mu)^4$
It is possible to Jensen's inequality if $g''(x) \geq 0$
$g(x) = (x-\mu)^4$
$g(x) = 4(x-\mu)^3$
$g(x) = 12(x-\mu)^2$
g(x) is convex, since the second derivative of the function is quadratic. Hence it will always be greater than zero. Which means that it its possible to make use of Jensen's inequality.
$E(cX-\mu)^4 \geq (E(X-\mu))^4$
$(E(X-\mu))^4$
$((E(X-\mu))^2)^2 = (Var(X))^2$

thus the claim is true, and shown by Jensen's inequality

$$E[(X-\mu)^4] \geq \sigma^4$$

3 Problem 3

3.1 (a)

4 Problem 4

4.1 (a)

In this exercise I'm asked to choose the null hypothesis.
My Null hypothesis is:, $H_0 : \mu_0 = 0$ My alternative hypothesis is, $H_A : \mu \neq \mu_0$
Which menas, that I assume that there is no difference in flowering time, since the value for $X_3 - Y_3 = -0.5$ showing that the scientists claim does not hold for all of the samples. With the specified alternative hypothesis, I would have to perform af two-sided t-test
$\mu_0 = 0$ since the assumption is, that there is no difference between the two types of flowers

4.2 (b)

Performing the corrosponding t-test (Assuming that I have to perform the corrosponding thest to my claim in (a)). Following the "six steps"from the lecture. I will be starting from step 3 since both step one and two are defined in question (a). The dataset is $X_i - Y_i$, which gives me

1	0.5	-0.5	1.5	0.5

Calculating the observed mean = $\frac{1+0.5-0.5+1.5+0.5}{5} = 0.6$
Calculating the standard deviation for my sample
$S = \sqrt{Var(Z)}$

$Var(Z) = \sum_{i=1}^{5} \frac{(x_i - \overline{x})^2}{n-1}$

$Var(Z) = \frac{(1-0.6)^2+(0.5-0.6)^2+(-0.5-0.6)^2+(1.5-0.6)^2+(0.5-0.6)^2}{4} = \frac{2.2}{4} = 0.55$

$S = \sqrt{0.55} = 0.7416$

$t = \frac{\overline{x}-\mu_0}{S\sqrt{n}} = \frac{0.6-0}{0.7416\sqrt{5}} = \frac{0.6}{0.3317} \approx 1.81$

I've calculated c_1 and c_2 using the following small code.

```

```