Functional and Shape Data Analysis
Weekly assignments 1 autumn 2019

Solutions for the following questions should be returned to Moodle's quiz platform. Platform will be opened within couple of days. The last return time is 17 January at 12.00.

1. The pinch is a dataset included in the fda package. It consists of 151 measurements of pinch force for 20 replications (curves).

```
>library(fda)
> pinch
        [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]  [,10]  [,11]  [,12]  [,13]  [,14]  [,15]  [,16]  [,17]  [,18]  [,19]  [,20]
  [1,] -0.068  0.078 -0.148 -0.065 -0.291 -0.041 -0.059 -0.102 -0.172 -0.224  0.002 -0.032 -0.111 -0.087  0.087 -0.019 -0.141 -0.032 -0.209 -0.102
  [2,] -0.090 -0.032 -0.090  0.045 -0.193 -0.087  0.115  0.005 -0.083  0.017 -0.123  0.042 -0.282 -0.410  0.002 -0.535  0.081 -0.038 -0.038 -0.053
  [3,]  0.100 -0.053 -0.053 -0.093 -0.273 -0.004 -0.032 -0.160 -0.120 -0.184 -0.135 -0.056  0.023 -0.026  0.020 -0.050  0.463 -0.169 -0.187 -0.141
  .
  .
[150,] -0.105  0.213 -0.120 -0.166 -0.227 -0.163 -0.239 -0.267 -0.032 -0.035 -0.181  0.011 -0.318 -0.111 -0.044 -0.361 -0.251 -0.050 -0.172 -0.190
[151,] -0.032 -0.120 -0.093 -0.050 -0.242  0.036 -0.264 -0.224 -0.102 -0.029 -0.047 -0.087 -0.123 -0.050 -0.145 -0.059 -0.215 -0.285 -0.227 -0.050


Matrices of dimension c(151, 20) = 20 replications of measuring pinch force every 2 milliseconds for 300 milliseconds.
pinch selected 151 observations so the maximum of each curve occurred at 0.076 seconds.
```

   (a) Convert the pinch data to functional objects using $K = 15$ B-splines of order four (cubic splines), and plot the 20 smoothed curves on one graph. Do the smoothing by using the least squares method to estimate coefficient vectors $\mathbf{c}_i$.

   (b) Calculate the sample mean curve $\bar{x}(t) = \frac{\sum_{i=1}^{N} \hat{x}_i(t)}{N}$, where now $N = 20$, and plot the sample mean curve $\bar{x}(t)$ to the same graph with smoothed curves.

   (c) Calculate also fitted values $\hat{\mathbf{y}}_i = \mathbf{H}\mathbf{y}_i$, and also consider calculation of the SSE values for each sampling unit i: $SSE_i = (\mathbf{y}_i - \hat{\mathbf{y}}_i)'(\mathbf{y}_i - \hat{\mathbf{y}}_i)$.

2. Let us continue with the pinch dataset.

   (a) Use the function `smooth.basis` from the library fda to convert the pinch data to functional objects with $K = 15$ B-splines of order four (cubic splines), if you haven't already done it.

   (b) Take a look at the help page of `deriv.fd` in library fda. Use that function for creating the first and second derivatives of the smooth curves. Plot the first and second derivatives of the smoothed curves to separate graphs.

   (c) Use the function `eval.fd` to evaluate how the first derivatives are behaving. Particularly, try to approximate time points t_* when the first derivative is zero for the sampling unit $i = 1$, i.e., $\frac{d\hat{x}_1(t)}{dt}|_{t=t_*} = 0$.

3. In example 1.3, it was considered the sea surface temperatures (SST's) available in the file ninoSST.txt. By using the NINO3 column, the data was treated for each calendar year as a functional observation.

```
> data<-read.table("ninoSST.txt", sep="\t", header=TRUE)
> data
    YEAR MON NINO12 ANOM12 NINO3 ANOM3 NINO4 ANOM4 NINO3.4 ANOM3.4
1   1950   1  23.11  -1.57 23.74 -2.03 27.03 -1.29   24.83   -1.84
2   1950   2  24.20  -1.91 24.92 -1.57 27.15 -1.02   25.20   -1.63
3   1950   3  25.37  -1.13 26.33 -0.91 27.06 -1.23   26.03   -1.30
.
.
767 2013  11  21.49  -0.33 24.94 -0.20 28.54 -0.11   26.54   -0.19
768 2013  12  22.61  -0.40 25.01 -0.30 28.28 -0.26   26.20   -0.49
```

(a) Convert the dataset ninoSST.txt to the form that you can smooth it by the function smooth.basis from the library fda.

(b) Take a look at the help page of create.fourier.basis function, and then convert the NINO3 values to functional objects using five Fourier basis functions (Note! You only need to use rangeval, nbasis arguments in function). Do the smoothing by using the least squares method to estimate coefficient vectors $\mathbf{c}_i$.

(c) Plot the smoothed curves on one graph, and plot the sample mean curve $\bar{x}(t)$ to the same graph with smoothed curves.

(d) Convert the NINO3 values to functional objects using 7 Fourier basis functions. Which number of Fourier basis functions works better in your opinion, $K = 5$ or $K = 7$?