Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
February 7th, 2023

Problem n.4

Riccardo is interested in studying his velocity during a 10 km run. The file `10km.txt` reports the recorded velocity during a running session, collected every 50 meters. Considering a functional data analysis approach, answer to the following questions.

a) Perform a smoothing of the data using smoothing splines of order 4 (with penalty of order 2). Choose the value of the smoothing parameter λ in the interval $[10^{-5}, 10]$ using a generalized cross-validation (GCV) criterion. Report the plot of the values of the GCV statistic as a function of the logarithm of λ, the value of λ chosen, and a plot of the smoothed data.

b) Compute an approximation of the first derivative of the curve from the data and the first derivative of the smoothed curve obtained at point (a). Provide a plot to compare the two and comment on the result.

c) Choose and report a value of λ that you deem appropriate to show the effect of oversmoothing. Provide a plot of the the smoothed data and comment the result.

d) Choose and report a value of λ that you deem appropriate to show the effect of overfitting. Provide a plot of the the smoothed data and comment the result.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUOFlMTUdRWEVLVUw5UTNVRlo2RDVYOEQ5Wi4u`