Lab 6
Learning in Reproducing Kernel Hilbert Spaces
Advanced Machine Learning
DATA 442/642

Exercise 1

Test the the prediction power of the kernel ridge regression in the presence of noise and outliers. The original data are samples from a music recording of Blade Runner by Vangelis Papathanasiou `https://en.wikipedia.org/wiki/Vangelis`.

(a) Read the audio file, BladeRunner.wav, using the Python SoundFile library (`https://pypi.org/project/SoundFile/`). Then take 100 data samples starting from the 100,000th sample. Add white Gaussian noise at a 15 dB level and randomly "hit" 10% of the data samples with outliers (set the outlier values to 80% of the maximum value of the data samples).

(b) Find the reconstructed data samples using the unbiased kernel ridge regression method, that is,

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{y}^\top (\mathbf{K} + C\mathbf{I})^{-1} \boldsymbol{\kappa}(\boldsymbol{x}).$$

Employ the Gaussian kernel with $\sigma = 0.004$ and set $C = 0.0001$. Plot the fitted curve of the reconstructed samples together with the data used for training.

(c) Repeat step (b) using $C = 10^{-6}, 10^{-5}, 0.0005, 0.001, 0.01, 0.05$.

(d) Repeat step (b) using $\sigma = 0.001, 0.003, 0.008, 0.01, 0.05$.

(e) Comment on the results.