Lecturer: Barbara Re **TAs:** Davide Torlo, Stefano Colombo

Exercise session 6, 23 October, 2019

Numerical Methods in Informatics (L + E)

Exercise 1 The weekly and the daily values of the stock prices of the company *Facebook Inc.*[1] from Oct. 1, 2017 to Sep. 30, 2018 are reported in the attached files `StockPrices.csv` and `StockPrices_day.csv`, respectively. In particularly, the former contains 52 entries, one for each week of the year; the latter includes less than 365 entries, since stock prices are available only on working days.
First of all, load both data-sets in MATLAB, <u>without</u> modifying the input files. For instance, you can load the weekly data-set by means of `csvread('StockPrices.csv',2,0)`, which skips the first two lines of the input file (see the `help`).

(a) For the weekly data-set, compute the least-squares approximations of degree $m = 1$, $m = 2$, and $m = 4$. Then, for each approximation, answer the following questions:

- Which are the coefficients $a_0, a_1, \dots, a_m$ of the approximating $\tilde{f}(x) = a_0 + a_1 x + \dots + a_m x^m$?
- Which is the maximum squared residual $(\tilde{f}(x_i) - y_i)^2$?
- Which is the sum of square of the residual? Why is this number important for this method?

(b) Compute also the cubic spline interpolation of the weekly data-set. Produce a figure, that includes the original points, the cubic spline, and the three least-square approximations.

(c) Compute the least-square approximations of degree $m = 1$, $m = 2$, and $m = 4$, and the cubic spline also for the daily data-set. Produce a figure as in point **(b)** including all four approximations.

(d) Look at the figures obtained in points **(b)** and **(c)**. If you want to forecast the stock price 15 days (or 2 weeks) after the last day in the data-sets, which approximation would you use, and why? What have you obtained? The actual value on Oct. 15, 2018 was 153.32, how can you explain the differences from the data predicted and the actual value?

Exercise 2 Overdetermined systems

(a) Compute the condition number of the pseudo inverse of the matrix used in the least square method. Given a rectangular system the equation

$$\mathbf{X}\mathbf{a} = \mathbf{y}, \quad \mathbf{X} \in \mathbb{R}^{n \times m+1}, \mathbf{a} \in \mathbb{R}^{m+1}, \mathbf{y} \in \mathbb{R}^n, \tag{1}$$

where $\mathbf{a}$ are the unknown coefficients, $\mathbf{y}$ the data values and

$$\mathbf{X}_{ij} = (x_i)^j, \; j = 0, \dots, m, \; i = 1, \dots, n. \tag{2}$$

Compute the matrix $\mathbf{X}^T\mathbf{X}$, that must be inverted to get to the pseudo inverse in $\mathbf{a} \approx (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Compute the condition number of $\mathbf{X}^T\mathbf{X}$. Take $m = 2$ as an example.

(b) Take $m = 2$. Instead of solving the overdetermined system with least-square method, use the QR decomposition to solve the rectangular system (1). The decomposition will be $\mathbf{X} = QR$ and hence we have to solve the two systems

$$\begin{cases} Qz = \mathbf{y} \iff z = Q^T\mathbf{y}, \\ R\mathbf{a} = z. \end{cases} \tag{3}$$

Check the condition number of Q and R.

[1] Data from *Investopedia*

(c) Use different basis functions to approximate the data. In supervised learning tasks, Gaussian functions centered in different points x_j are often used

$$\varphi_j = e^{-\left(\frac{x-x_j}{s}\right)^2}. \tag{4}$$

Use the basis functions $\varphi_1(x) = 1$, $\varphi_2(x) = e^{-\left(\frac{x-120}{50}\right)^2}$, $\varphi_3(x) = e^{-\left(\frac{x-240}{50}\right)^2}$. The approximation functions will be in this case

$$y_i \approx \sum_{j=1}^{3} a_j \varphi_j(x_i). \tag{5}$$

Denoting with

$$\mathbf{X}_{ij} = \varphi_j(x_i),\ j = 1, \ldots, 3,\ i = 1, \ldots, n, \tag{6}$$

solve the least-square problem (1).