LELEC2870: Exercise Session 2
Data Parsing

1 Dataset and Notations

The datasets for this session are available on the Moodle website. After loading the `diabetes.mat` file, you will be able to work on a total of 442 instances coming from a diabetes study. For each one, the values of 10 features are given, as well as the target value. The features include the age (feature 1), the sex (feature 2), the body mass index (feature 3) and the blood pressure (feature 4) of the patient, as well as the result of several serum measurements (features 5 to 10). During this exercise session, it may be interesting to link your results with the feature's interpretation.
In this session, the learning set is $\{(\mathbf{x}_p, t_p)|p = 1...P\}$ where P is the number of samples, $\mathbf{x}_p$ is an input row vector of dimension D

$$\mathbf{x}_p = \begin{pmatrix} x_p^1 & x_p^2 & \cdots & x_p^D \end{pmatrix} \tag{1}$$

and t_p is the scalar output. For computations, inputs are placed in the matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_P \end{pmatrix} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^D \\ x_2^1 & x_2^2 & \cdots & x_2^D \\ \vdots & \vdots & \ddots & \vdots \\ x_P^1 & x_P^2 & \cdots & x_P^D \end{pmatrix} \tag{2}$$

and targets are placed in the column vector

$$\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_P \end{pmatrix}. \tag{3}$$