CK0146: Exercise 01 (2017.1)

Exercise Q.1 (Polynomial regression). Generate a training data set for regression consisting of $N = 100$ points. The input values $\{x_n\}_{n=1}^N$ are generated uniformly in the interval $(0, 1)$ and the corresponding target values $\{t_n\}_{n=1}^N$ are obtained by first computing the corresponding values of the function $abx^{(a-1)}(1-x^a)^{b-1}$, with $a = 2$ and $b = 5$, and, then, by adding some random noise ε with a Gaussian distribution having mean $\mu = 0$ and standard deviation $\sigma = 0.3$, $\varepsilon \sim \mathcal{N}(\mu, \sigma)$:

$$t_n = abx_n^{(a-1)}(1-x_n^a)^{b-1} + \varepsilon_n, \quad \text{with } a = 2, b = 5 \text{ and } \varepsilon_n \sim \mathcal{N}(\mu, \sigma).$$

Fit the data using a polynomial function of the form

$$y_M(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j,$$

in which M is the order of the polynomial. The polynomial coefficients are collected in vector $\mathbf{w}$.

The fit must be done by minimising, independently, two error functions that measure, for any given value of $\mathbf{w}$, the misfit between function $y_M(x, \mathbf{w})$ and the training target data: That is,

- $\triangleright$ $E_{\text{OLS}}(\mathbf{w}, M) = \frac{1}{2}\sum_{n=1}^{N}\Big(y_M(x_n, \mathbf{w}) - t_n\Big)^2$, ordinary least squares regression
- $\triangleright$ $E_{\text{RIDGE}}(\mathbf{w}, M, \lambda) = \frac{1}{2}\sum_{n=1}^{N}\Big(y_M(x_n, \mathbf{w}) - t_n\Big)^2 + \frac{\lambda}{2}||\mathbf{w}||^2$, ridge regression

In the first case, you must select the order M of the polynomial using S-fold cross-validation, for $S = 3$ and for $S = N$ (leave-one-out cross-validation, LOO). In the second case, let M be fixed and select the regularisation coefficient λ, again using K-fold cross-validation, for $K = 3$ and for $K = N$.

Derive/report closed-form solutions $\mathbf{w}^*_{\text{OLS}}$ and $\mathbf{w}^*_{\text{RIDGE}}$, write code for performing all experiments (data generation and model cross-validation) and report on the evolution of cross-validation results.

Comment on the used procedure and on the obtained results.