Assignment V (2015.2 - T01)

Submission Deadline: November, 30th, 2015.

Instructions:

- Submission deadline: November, 30th, 2015.
- This assignment must be delivered as a report (with Introduction, Methodology, Results, Conclusion and References).
- Source codes must be delivered as attachments.
- When answering the questions, books and papers citations are allowed as long as they are listed in "references" section of the report.
- When submitting this assignment, please inform if you are registered or not in SIGAA.
- If you are enrolled in SIGAA, provide your university ID number when submitting the assignment.

Exercises

Exercise 01

Use *Nadaraya-Watson model*

$$y(\mathbf{x}) \;=\; \sum_n k(\mathbf{x}, \mathbf{x}_n) t_n$$

to predict the values of `data_A_testing_output.txt` using `data_A_testing_input.txt` as input. To build the model, use `data_A_learning_input.txt` and `data_A_learning_output.txt` as training data. Use the gaussian kernel

$$k(\mathbf{x}, \mathbf{x}') \;=\; \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}\right)$$

setting $\sigma = 1$. Plot results (measured data and predicted data). Inform the mean squared error as well. Remember that in *Nadaraya-Watson model*

$$\sum_{n=1}^{N} k(\mathbf{x}, \mathbf{x}_n) = 1.$$

Exercise 02

Using the same data set of the previous exercise, use *Radial Basis Functions* (RBF) to compare results with the ordinary *Nadaraya-Watson* model (used in the previous exercise). Use a clustering algorithm (for example: *k-means*) to determine the radial basis function centers. To project RBF, follow the steps described below.

2.1 Let q be the number of desired basis functions. Determine q with any value. It is a good practice to determine $q \leq N$, where N is the number of samples in the data.

2.2 Determine the centers values of each basis function. You can use a clustering algorithm such as $k - means$ to determine the centers.

2.3 Calculate the activation of each function basis. Use a Gaussian kernel setting $\sigma = 1$.

2.4 Normalize all Gaussian basis function output in a way that

$$\sum_n k(\mathbf{x}, \mathbf{x}_n) \;=\; 1$$

2.5 Store all outputs in a matrix

$$\mathbf{\Phi} \;=\; \begin{pmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \dots & \phi_M(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \dots & \phi_M(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \dots & \phi_M(\mathbf{x}_N) \end{pmatrix},$$

and use *Regularized Least Squares* to determine the weights

$$\mathbf{w} \;=\; \left(\lambda \mathbf{I} + \mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{t},$$

where $0 < \lambda \leq 1$.

Exercise 03

This exercise explores *Gaussian Processes for Regression*. Recall that

$$t_n = y(\mathbf{x}_n) + \epsilon_n$$

where ϵ_n is a random noise variable whose value is chosen independently for each observation n. The joint distribution of the target values $\mathbf{t} = (t_1, \ldots, t_N)^T$ conditioned on the values of $\mathbf{y} = (y(\mathbf{x}_1), \ldots, y(\mathbf{x}_N))^T$ is given by an isotropic Gaussian of the form

$$p(\mathbf{t}|\mathbf{y}) = \mathcal{N}(\mathbf{t}|\mathbf{y}, \beta^{-1}\mathbf{I}_N)$$

where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. Recall that the marginal distribution of $\mathbf{t}$ is given by

$$p(\mathbf{t}) = \mathcal{N}(\mathbf{t}|\mathbf{0}, \mathbf{C}_N)$$

where the covariance matrix $\mathbf{C}_N$ has elements

$$C_N(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \beta^{-1}\delta_{nm}$$

where β^{-1} is the noise precision and

$$\delta_{nm} = \begin{cases} 1, & \text{if (n=m)} \\ 0, & \text{otherwise.} \end{cases}$$

Recall that the goal in regression is to make predictions of target variables for new inputs. In order to find the conditional distribution $p(t_{N+1}|\mathbf{t})$ (where t_{N+1} stands for the target of the new input and $\mathbf{t} = (t_1, \ldots, t_N)^T$ stands for a vector of targets available in the training data), begin by writing down the joint distribution $p(\mathbf{t}_{N+1})$.

$$p(\mathbf{t}_{N+1}) = \mathcal{N}(\mathbf{t}_{N+1}|\mathbf{0}, \mathbf{C}_{N+1})$$

where

$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^T & c \end{pmatrix},$$

$\mathbf{t}_{N+1} = (t_1, t_2, \ldots, t_N, t_{N+1})^T$, $\mathbf{k} = (k(\mathbf{x}_1, \mathbf{x}_{N+1}), k(\mathbf{x}_2, \mathbf{x}_{N+1}), \ldots, k(\mathbf{x}_N, \mathbf{x}_{N+1}))^T$ and the scalar $c = k(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) + \beta^{-1}$. Partitioning $p(\mathbf{t}_{N+1})$ into $p(t_{N+1})|\mathbf{t})$ we get the result:

$$\begin{aligned} m_{t_{N+1}|\mathbf{t}} &= \mathbf{k}^T (\mathbf{C}_N)^{-1} \mathbf{t} \text{ and} \\ \sigma^2_{t_{N+1}|\mathbf{t}} &= c - \mathbf{k}^T (\mathbf{C}_N)^{-1} \mathbf{k}. \end{aligned}$$

Using the dataset pair `data_B_learning_input.txt` and `data_B_learning_output.txt` as the training data pairs $\{\mathbf{x}_n, t_n\}$, implement *Gaussian Processes for Regression* to predict the outputs of `data_B_testing_input.txt`.
Then, compare the predicted values with `data_B_testing_output.txt`. Notice that, in this exercise, ϵ_n is a gaussian noise of standard deviation equal 2.

Use

$$k(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 \exp\left\{\theta_1 ||\mathbf{x}_n - \mathbf{x}_m||^2\right\}$$

as a kernel function. In this kernel, θ_0 and θ_1 are adjustable parameters. Based on *mean squared error*, choose the best values of θ_0 and θ_1 knowing that $(\theta_0, \theta_1) \in \mathbb{Z}^2$ such that

$$\begin{aligned} -10^3 \leq \theta_0 \leq 10^3 \text{ and} \\ -10^3 \leq \theta_1 \leq 10^3 \end{aligned}$$

is true.

Good luck!