Assignment Sheet - V

Stylianos Sygletos
s.sygletos@aston.ac.uk

May 17, 2021

Exercise 1 [5 points]

Consider a two-class, two-dimensional classification problem, where the first class (ω_1) is modelled by a Gaussian distribution with mean $\boldsymbol{\mu}_1 = [0, 2]^T$ and covariance matrix $\Sigma_1 = \begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix}$, while the second class ω_2 is modelled by a Gaussian distribution with mean $\boldsymbol{\mu}_2 = [0, 0]^T$ and covariance matrix $\Sigma_2 = \begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix}$

i. Generate and plot a training set $\mathcal{X}$ and a test set $\mathcal{X}_{test}$, each one consisting of 1500 points from each distribution. Use different random number seeds when generating the training and test sets. [1.0pt]

ii. Classify the data vectors of X test test using the Bayesian classification rule.[1.0pt]

iii. Perform logistic regression and use the data set X to estimate the involved parameter vector $\boldsymbol{\theta}$. Evaluate the classification error of the resulting classifier based on $\mathcal{X}_{test}$.[1.0pt]

iv. Comment on the results obtained by (ii) and (iii).[1.0pt]

v. Repeat the previous steps (i)-(iv), for the case where $\Sigma_2 = \begin{bmatrix} 4 & -1.8 \\ -1.8 & 1 \end{bmatrix}$ and compare the obtained results with those produced by the previous setting. Draw your conclusions.[1.0pt]

Hint: For the estimation of $\boldsymbol{\theta}$ in (iii), perform steepest descent with learning parameter μ equal to 0.001 to minimize the log-likelihood function $L(\boldsymbol{\theta}) = -\sum_{n=1}^{N}(y_n \ln s_n + (1 - y_n)\ln(1 - s_n))$, where $s_n := \sigma(\boldsymbol{\theta}^T \boldsymbol{x}_n)$ the estimated output value for training sample $\boldsymbol{x}_n$ at the input.

Problem 2 [5 points]

In this problem we will examine the prediction power of the kernel ridge regression in the presence of noise and outliers. We consider original data were samples are generated from a music recording, such as of the *Blade Runner* by Vangelis Papathanasiou. Read the audio file using the read function from SoundFile [1] package and take 2000 data samples (starting from the 150,000th sample). Then add white gaussian noise at a 10 dB level and randomly “hit” 100 of the data samples y_n with outliers ($y_n \pm 0.9y_{max}$, i.e. set the outlier induced deviation is 90% of the maximum value of all the data samples).

[1] https://pypi.org/project/SoundFile/

i. Find the reconstructed data samples using the kernel ridge regression method (see eq. 6.9 of PRML textbook). Employ the Gaussian kernel with $\sigma = 0.004$ and set $\lambda = 0.0001$. Plot the fitted curve of the reconstructed samples together with the data used for training. [1pt]

ii. Repeat the first step (i) using $\lambda = 10^{-6}, 10^{-5}, 0.0005, 0.001, 0.01$ and 0.05 [1pt]

iii. Repeat the first step (i) using $\sigma = 0.001, 0.003, 0.008$ and 0.05 [1pt]

iv. Comment on the results [2pt]

Problem 3 [5 points]

Consider a two-dimensional class problem that involves two classes ω_1 $(+1)$ and ω_2 (-1). Each one of them is modelled by a mixture of equiprobable Gaussian distributions. Specifically, the means of the Gaussians associated with ω_1 are $[-5, 5]^T$ and $[5, -5]^T$, while the means of the Gaussians associated with ω_2 are $[-5, -5]^T$, $[0, 0]^T$ and $[5, 5]^T$. The covariances of all Gaussians are $\sigma^2 \boldsymbol{I}$ where $\sigma^2 = 1$.

i. Using of the *mixt_model* python function that you will be given during one of the lab sessions, generate and plot a data set X_1 (training set) containing 100 points from ω_1 (50 points from each associated Gaussian) and 150 points from ω_2 (again 50 points from each associated Gaussian). In the same way, generate an additional set X_2 (test set). Make sure that you su Plot the generated sets designating the different points of each class. [0.5pt]

ii. Based on X_1, train a two-layer neural network with two nodes in the hidden layer having the hyperbolic tangent as activation function and a single output node with linear activation function [2], using the standard back-propagation algorithm of 9000 iterations and step-size equal to 0.01. Compute the training and test errors, based on X_1 and X_2, respectively. Also, plot the test points as well as the decision lines formed by the network. Finally, plot the training error versus the number of iterations.[0.5pt]

iii. Repeat step (ii) for step-size equal to 0.0001 and comment on the results.[0.5pt]

iv. Repeat step (ii) for $k = 1, 4, 20$ hidden layer nodes and comment on the results. [0.5pt]

v. Repeat the above steps (i)-(iii) for the case where the covariance matrix for the Gaussians is $6I$ for 2, 20 and 50 hidden layer nodes, compute the training and the test errors in each case and draw the corresponding decision regions. Draw your conclusions. [3.0pt]

[2]The number of input nodes equals to the dimensionality of the feature space, while the number of output nodes is equal to the number of classes minus one.