Machine Learning Algorithms: exercise 2 23.03.2023

1. Use Data.txt file and classify points from C_1 and C_2 using classifier **w**=$\frac{1}{\sqrt{2}}$(1,-1) that starts from point **p**=(2,2). Evaluate the performance of the classifier using accuracy, sensitivity and specificity.
2. Calculate also the probability of error for the classifier depicted in task 1.
3. Let us consider two class classifier that has uniform distributions with equal prior probabilities.

$$p(x|\omega_1) = \begin{cases} 1 \; when \; 0 \leq x \leq 1 \\ 0 \; otherwise \end{cases}$$

$$p(x|\omega_2) = \begin{cases} \frac{1}{2} \; when \; \frac{1}{2} \leq x \leq \frac{5}{2} \\ 0 \;\; otherwise \end{cases}$$

What is the probability of an error in classification?

4. Write a function that numerically calculates the probability for an event that falls on a closed interval [*a*,*b*] in the case of one dimensional normal distribution. The parameters for the function is the mean μ, variance δ^2 lower integration limit ***a*** and upper integration limit ***b***. Integral can be approximated for instance, using cumulative sum of probability density function. (cumsum in Matlab). It is sufficient to use “narrow” rectangles in the estimation of differentials.
5. Let us again consider a two class classifier. It uses one dimensional normal distributions N(0,1) for class ω_1 and N(3,2) for class ω_2. Prior probabilities for classes are $P(\omega_1) = 0.3$ and $P(\omega_2) = 0.7$. Estimate the probability of an error using a numerical method. You can reduce your calculations on interval [-1, 6].

Normal distribution of variable *x* that has the mean μ and standard deviation σ has a functional form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Its integral can be numerically estimated using approximation $F(x) = \sum_a^b f(x)dx$, where a is the starting point of integration and b is the stopping point of integration. dx is a small difference between two successive *x* values x_i and x_{i+1}.

6. File data2 contains two columns. First column contains values for variable x and second column contains response values y. Fit polynomials of order 1, 2, 3, 4 and 5 using linear regression to the data. Consider what happens, if you use even higher order polynomials.