In the name of God

Quiz – Data Science Class
Instructor: Dr. Arian

1) Describe the difference between OLS (ordinary least squares) and regularized linear regression.

2) What are the main advantages and disadvantages of regularized linear regression. Explain where regularized linear regression is a benefitial method for us and where using it may not be efficient.

3) The dataset “Firm costs.csv” is attached to CW. Manually define a function which estimates the coefficients of the regularized linear regression by minimizing the function described in lecture7-Regularization.

4) Consider a feed-forward multi-layer neural network whose units have the sigmoid output function $g(a)=$ $[1+\exp (-a)]^{-1}$, where a denotes the activation value of a unit. Assume the network is trained using the back-propagation learning algorithm on a training set T made up of $n=100$ examples belonging to two classes A and B, using the error function $E=\frac{1}{2} \sum_{i=1}^{n}\left(y_{i}-t_{i}\right)^{2}$, where dove y_{i} denotes the network output for the i-th example and t_{i} the corresponding desired (target) output, defined as 0 for class A and 1 for class B. Assume further that the back-propagation algorithm finds values of the connection weights that provide the global minimum of the above error function. If all the training examples have $i d e n t i c a l$ attribute values, but 80 of them are of class B and the remaining 20 ones are of class A, what class label would the trained network assign to every such example?

5) Consider a squared loss function of the following form form for an Artificial Neural Network

$$E = \frac{1}{2} \iint \{y(\mathbf{x}, \mathbf{w}) - t\}^2 p(\mathbf{x}, t)\, \mathrm{d}\mathbf{x}\, \mathrm{d}t$$

where $y(x, w)$ is a parametric function. Assume that (you don’t need to prove this!) the function $y(x, w)$ that minimizes this error is given by the conditional expectation of t given x. Use this result to show that the second derivative of E with respect to two elements ω_r and ω_s of the vector w is given by

$$\frac{\partial^2 E}{\partial w_r \partial w_s} = \int \frac{\partial y}{\partial w_r} \frac{\partial y}{\partial w_s} p(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$