# Lab04: Artificial Neural Networks and Support Vector Machines

**Handed out:** Wednesday, Nov. 20, 2019

**Return date:** Monday, Dec. 9, 2019, the latest

**Grades:** Lab04 counts 12 % towards your final grade.

**Objectives:** Explore properties of simple feed-forward neural networks and support vector machines.

**Format of answer:** Your answers (statistical figures and verbal description) should be submitted as ***hardcopy***. Add a running title with the following information: Lab04, your name and page numbers. You may use this document as template. Copy the requested statistical figures into your document. Trial and error answers will lead to a deduction of points. Label each answer properly with the bold task and sub-task headings. You are expected to hand in professionally formatted answers: use a fixed pitch font, like **Courier New**, for any  code the use mathematical type-setting when equations are required. Copy and paste figures into your document. Make sure that each figure has a proper ***caption*** describing its content.

## Part 1: Neural Networks [4 points]

You will us the  code provided to you in the script **Task1&2.r** and answer the following questions

**Task 1:** Comparison of logistic regression with a one-layer one-neuron network. [2 point]

**Q1**: Why can logistic regression be used technically for a normalized dependent variable?   
Notes: You can ignore the warning message here and that the specified model performs poorly.

**Q2**: Which options in the **neuralnet( )** function call makes the neural network comparable to the logistic regression model?

**Q3**: Are the network weights comparable to the logistic regression coefficients?

**Q4**: Why are the intercept and the bias coefficients allowed to differ?

**Task 2:** Use of cross-validation to avoid model overfitting and identify the proper neural network specification for a small dataset. [2 points]  
Notes: This task is best performed with Microsoft’s Open  version on a computer with multiple processor cores because neural networks make heavy use of matrix multiplications. The run time can vary between 30 minutes to several hours depending on the  version and computer capabilities.

**Q1**: How is cross-validation algorithm implements in the  code.

**Q2**: What is the maximum number of neurons before the model overfits the Boston median home value data.

## Part 2: Support Vector Machines [5 point]

In this part you will answer two applied exercises in James et al., 2013. *An Introduction to Statistical Learning with Application in R*. Use 25% of the observations as test dataset. Please follow the sequence of tasks/questions in the exercises.

**Task 3:** Answer the task/questions of exercise 5 on pages 369-370. [4 point]

**Task 4:**. Answer the task/questions of exercise 8 on pages 371-372. [4 point].   
With regards to the interpretation of the **cost** parameter please see the first paragraph of the section *Support Vector Classifier* on page 359.