# Lab04: Artificial Neural Networks and Support Vector Machines

**Handed out:** Thursday, April 22, 2021

**Return date:** Thursday, May 6, 2021, ***by noon*** into the **Lab04Submit** link.

**Grades:** Lab04 counts 19 % 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: Support Vector Machines [15 points]

**Task 1:** You will answer an applied exercise in James et al., 2013. *An Introduction to Statistical Learning with Application in R*. Please follow the sequence of tasks/questions in the exercises*. Answer the questions (a) to (i) of exercise 5 on pages 369-370.* Show your code. [5 point]

**Task 2:**. For the following tasks continue working with the **credit.csv** data set to predict the default probabilities. Split the data into a stratified training data set with 70% of the observations and a test data set with the remaining 30% of the observations. *Use a radial kernel support vector classifier. Identify with cross-evaluation the “optimal” cost parameter. Evaluate your optimal model with the confusion matrix for the test dataset and the ROC curve including the AUC*. Show your code. [5 points]

**Task 3:** Read up on section 14.3.2 of Boehmke et al. and the support vector regression code in [Chapter 14: Support Vector Machines (koalaverse.github.io)](https://koalaverse.github.io/homlr/notebooks/14-svm.nb.html) underneath Figure 14.10. *Estimate a 20% test sample of the* ***Ames*** *dataset the predicted* ***Sale\_Price****. Scatterplot the observed and predicted sales price for the test sample against each other.* Show your code. [5 points]

## Part 2: Neural Networks [4 points]

You will us the  code provided to you in the script **Task4&5.R** and answer the following questions

**Task 4:** Comparison of logistic regression with a one-layer one-neuron network. [2 points]

[a] *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.

[b] *Which options in the* ***neuralnet( )*** *function call makes the neural network comparable to the logistic regression model?*

[c] *Are the network weights comparable to the logistic regression coefficients?*

[d] *Why are the intercept and the bias coefficients allowed to differ?*

**Task 5:** 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 tensor operations. Alternatively, one could rewrite the loops with foreach to allow for parallel execution of the loops. The run time can vary between 20-50 minutes depending on the  version and computer capabilities.

[a] *Describe how is the cross-validation algorithm implemented in the  code.*

[a] *What is the maximal number of neurons before the model overfits the Boston median home value data.*