Machine Learning
Exercise Sheet 4

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

November 16th, 2017
Submission until November 23th (noon) by dropping at MACHINE LEARNING postbox
(please indicate in which tutorial are you participating!)
The postboxes is located inside the Samelsonplatz building to the right.

Pick two exercises and solve them! You can solve all of them to have some extra points to fill past bad grades.

Hint: Cramer's rule to find the inverse matrix of a 2x2 matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

And

$$A^{-1} = (1/(ad - bc)) \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Exercise 7: Linear Discriminant Analysis (10 Points)

The HiWi's from Machine Learning 1 have decided to open a new pizza restaurant, where you can design your own pizza. They notice that most clients would only come dressed with a red or black T-Shirt. They took notes from 12 different clients using PTWU (Pizza Topping Weight Units):

$ClientNo.$	$VegetarianToppings$	$MeatToppings$	$ShirtColor$
01	4.9	10	B
02	4.7	15	B
03	4.5	12	B
04	5.8	25	B
05	4.9	10	B
06	4	15	B
07	5	11	B
08	8.2	45	R
09	8.7	50	R
10	6.9	55	R
11	7.2	52	R
12	9	51	R

In order to help clients decide their toppings faster they need you to estimate parameters and a discriminant function for a discriminant analysis in order to guess the shirt color for the client number 13 (V. topping: 6, M. topping: 25). Show all your steps.

Exercise 8: Model Selection (10 Points)

a) Explain in your words the main difference between the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC)!

b) Rafael and Maurício decided to build a model to predict how much of rice and beans they would need to feel less "saudade" from Brazil. They build the following table with data from their experiments:

$Rice(Centigrams)$	$Beans(centigrams)$	$RelievedSaudade(inSaudadeUnits)$
1	2	5
-1	1	7
2	1	2
4	1	-4

The following model has been learned using that data:

$$\beta = (4 \quad -2 \quad 1)$$

. Compute its error as well as its AIC by using the negative RSS as logarithm of the likelihood.

$$log(\mathcal{L}) = -\sum_{i=1}^{N}(y_i - \hat{y}(x_i))^2$$

c) Perform a backward search on the employed variables and compare the three resulting models to the full model. Which one do you choose in the end?

Exercise 8+1: Regularization (10 points)

a) Explain in your own words why regularization is a key aspect in machine learning!

b) Ridge Regression learns model parameters β by minimizing the following objective function:

$$f(\beta, \lambda) = \sum_{i=1}^{N}(y_i - \beta^\top x_i)^2 + \lambda||\beta||_2^2 \qquad \lambda > 0$$

Learn two regression models using the **closed form solution** for $\lambda = 2 \quad and \quad \lambda = 5$ for the following data:

x	y
2	0
4	-2
-1	3
-2	4

Which model performs better for $(x = 5, y = -3)$?

c) Compute the partial derivative $\frac{\partial f}{\partial \lambda}$! Why do we not simply learn λ using gradient descent? Explain what happens if we would.