Machine Learning 2
Exercise Sheet 1

Prof. Dr. Dr. Lars Schmidt-Thieme, Brad Baker
Information Systems and Machine Learning Lab
University of Hildesheim

April 12th, 2017

Submission until Tuesday April 18th, 8:00 AM by learnweb.
Please put your name in all filenames and somewhere visible in the margins of the pdf.
Non-pdf submissions for non-programming exercises will not be graded.

It is highly recommended you read Murphy chapter 9, especially the starting sections. This exercise sheet will focus on topics discussed in this chapter. For further reading on General Linear Models specifically, see Dobson's 2008 book "An Introduction to General Linear Models".

Exercise 1: Exponential Families (10 Points)

a) (3 pts) The form of the exponential family is given in Equation 9.2 in Murphy and on slide 1 of lecture 1. A further generalized form of this equation is given as

$$p(\mathbf{x}|\theta) = h(\mathbf{x}) \exp\left[\eta(\theta)^T \phi(\mathbf{x}) - A(\eta(\theta))\right].$$

(Murphy eq. 9.5, Slide 2)

What condition is satisfied when a distribution belongs to a **natural exponential family**? When the distribution is in **canonical form**? When a distribution is in a **curved exponential family**?

For the Bernoulli, Multinoulli, and Univariate Gaussian distributions, assert whether or not each distribution belongs to a natural exponential family and/or is in canonical form.

b) (4 points) Suppose $y \in \{0, 1, ..., N\}$ is a target variable, $\mathbf{x} \in \mathbb{R}^m$ is a real variable instance, $\mathbf{w} \in \mathbb{R}^m$ are model parameters, N is a known number of instances, and $\pi(x) = \sigma(\mathbf{w}^T\mathbf{x})$ for the logistic function σ. Show that the binomial $Y \sim \text{Bin}(N, \pi)$ distribution belongs to a natural exponential family of canonical form, i.e. show that the probability density $p(y; N, \pi)$ has a representation:

$$p(y; N, \pi) = \binom{N}{y} \pi(\mathbf{x})^y (1 - \pi(\mathbf{x}))^{(N-y)} = \exp(y\mathbf{w}^T\mathbf{x} - A(\mathbf{w}^T\mathbf{x}) + c(y)).$$

Is this given family a valid exponential family for all values of π? Provide an argument why or why not.

c) The **student t distribution** (see Murphy section 2.4.2) has the probability density function

$$p(x; \mu, \sigma^2, \nu) := \left[1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}$$

where μ is the Expected Value (mean) of x, $\sigma > 0$ is a scaling parameter, and ν is the degrees of freedom of a given model. For a specific $\nu \in 1, 2$ or for general ν, show why the student t distribution does not belong to an exponential family (See Murphy section 9.2.2.4 for a hint. You will not get points for simply restating the hint).

exercise 2 Generalized Linear Models (10 points

a) (3 points) Using the binomial distribution given above, derive the Cumulant function $A(\theta)$, the expected value of y, $\mu = E(y|x; w, \sigma^2)$, and the variance of y, $Var(y|x; w, \sigma^2)$ for L2 **regularized** binomial regression.

b) (3 points) Using the negative log likelihood $l(w|x, y)$ for binomial regression, derive the gradient $\nabla_w l(w)$. Write pseudo-code for a **Stochastic Gradient Descent** algorithm to learn binomial regression with the logistic mean function.

c) (4 points)

In section 9.3.2, Murphy discusses the possibility of using different mean functions for binomial general linear models.

For example:

Name	Formula
Logistic	$sigm(\eta)$
Probit	$\phi(\eta)$
Log-log	$\exp(-\exp(-\eta))$
Complimentary log-log	$1 - \exp(-\exp(-\eta))$

where ϕ is the cumulative distribution function (cdf) of the standard normal distribution.

For the Logistic, the Log-Log, and the complimentary Log-Log functions given above, compute the corresponding link functions.

Aranda-Ordaz, in a 1981 paper, presented a general family of link functions for binomial modelling, given as

$$g(x, \alpha) = log\left[\frac{(1-x)^{-\alpha} - 1}{\alpha}\right]$$

, and a method for learning these α parameters to suit data.

For what value of α will the above general form for $g(x)$ reduce to the logit function (the logistic link function)? For which values of α does the form reduce to the complimentary log-log function?

Bonus (5 points on this sheet)

For this exercise it is recommended you use the R programming language, but you are free to use any language of your choice that can be interpreted or compiled in Linux or Windows.

Read Chapter 11.6 in "An Introduction to R" `http://cran.r-project.org/doc/manuals/R-intro.pdf` to see the commands for running Generalized Linear Models in R.

Download the "kalythos.csv" dataset from moodle, and load it as a data frame into R, by using the read.csv command.

A description of the data set is given below (from "An Introduction to R"): "On the Aegean island of Kalythos the male inhabitants suffer from a congenital eye disease, the effects of which become more marked with increasing age. Samples of islander males of various ages were tested for blindness and the results recorded. The data is summarized below:

Age	20	35	45	55	70
Number tested	50	50	50	50	50
Number blind	6	17	26	37	44

a) (2 points)Which probability distribution is well-suited to modelling the age at which the chance of blindness for a male inhabitant is 50%? Why?

b) (3 points) Use the R functions for general linear models to test binomial regression using three different link functions of your choice. Plot the output of each model compared with the results seen in the data, and discuss any differences between the outputs.