ETH Zurich
D-MATH
Dr. C. Buser
Prof. Dr. M. Wüthrich

January 24, 2019

Examination

Data Analytics for Non-Life Insurance Pricing

Please fill in the following table

Last name	
First name	
Programme of study	**MATH** ☐ **SAV** ☐ **Other** ☐
Matriculation number	

Leave blank

Question	Maximum	Points	Check
1	10		
2	10		
3	10		
4	10		
Total	**40**		

Instructions

Duration of exam: 120 min.

Closed book examination: no notes, no books, no calculator, no smartphones, etc., allowed.

Important:

- ◇ Please put your student card (or an identification card for SAV students) on the table.
- ◇ Only pen and paper are allowed on the table. Please do **not** write with a **pencil** or a **red** or **green** pen. Moreover, please do not use **whiteout**.
- ◇ Start by reading all questions and answer the ones which you think are easier first, before proceeding to the ones you expect to be more difficult. Do not spend too much time on one question but try to solve as many questions as possible.
- ◇ Take a new sheet for each question and write your name on every sheet.
- ◇ All results have to be **explained/argued** by indicating intermediate steps in the respective calculations. You can use known formulas from the lecture without derivation.
- ◇ Simplify your results as far as possible.
- ◇ Some of the subquestions can be solved independently of each other.

⋆⋆⋆ Good luck! ⋆⋆⋆

Question 1 (10 **points**)

Assume we have $n = 10$ observations given by $\mathcal{D} = \{(Y_1, \boldsymbol{x}_1, v_1), \ldots, (Y_n, \boldsymbol{x}_n, v_n)\}$. Assume that Y_i are independent and Poisson distributed for $i = 1, \ldots, n$ with

$$Y_i \sim \text{Poisson}\left(\lambda(\boldsymbol{x}_i)v_i\right),$$

for given volumes $v_i = 1$ and with (unknown) regression function $\boldsymbol{x} \mapsto \lambda(\boldsymbol{x}) > 0$.

(a) Assume that the features $\boldsymbol{x}_i$ are (continuous) real-valued for all $i = 1, \ldots, n$. We have collected the following data $\mathcal{D}$:

i	1	2	3	4	5	6	7	8	9	10
Y_i	10	10	10	20	20	20	10	10	10	10
$\boldsymbol{x}_i$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
v_i	1	1	1	1	1	1	1	1	1	1

(i) How many (non-trivial) splits will the standardized binary regression tree perform if we use the Poisson deviance statistics as loss function? How big is the final Poisson deviance statistics loss? Give arguments for your answers.

(ii) Which is the first split that the standardized binary regression tree will make under the Poisson deviance statistics loss function? Give the right guess!

(b) Assume that the features $\boldsymbol{x}_i$ are (unordered) categorical for all $i = 1, \ldots, n$. We have collected the following explicit data $\mathcal{D}$:

i	1	2	3	4	5	6	7	8	9	10
Y_i	10	10	10	20	20	20	10	10	10	10
$\boldsymbol{x}_i$	BE	VS	TG	GR	AI	SG	TI	UR	SZ	LU
v_i	1	1	1	1	1	1	1	1	1	1

(i) How many (non-trivial) splits will the standardized binary regression tree perform if we use the Poisson deviance statistics as loss function? How big is the final Poisson deviance statistics loss? Give arguments for your answers.

(ii) Which is the first split that the standardized binary regression tree will make under the Poisson deviance statistics loss function?

(c) Prove that every standardized binary split considered in items (a) and (b) strictly decreases the Poisson deviance statistics loss.

(d) Comparing the results of items (a) and (b) we obtain the following in-sample losses on $\mathcal{D}$ and out-of-sample losses on test data $\mathcal{T}$.

	in-sample loss	out-of-sample loss
homogeneous model $\lambda(\cdot) \equiv$ constant	14.9630	19.9202
continuous features and 1 split	10.1939	9.4271
categorical features and 1 split	0.0000	15.4518

Discuss the two error measures (in-sample loss and out-of-sample loss) and make a model choice (with justification).

Question 2 (10 points)

Assume we have n observations given by

$$\mathcal{D} = \{Y_1, \ldots, Y_n\}.$$

Assume that Y_i are independent and exponentially distributed for $i = 1, \ldots, n$ with density

$$f_Y(y) \;=\; \theta \exp\{-\theta y\} \;\mathbb{1}_{\{y \geq 0\}},$$

for a given (but unknown) parameter $\theta > 0$.

(a) Calculate the maximum likelihood estimator $\widehat{\theta}^{\mathrm{MLE}}$ for θ, given data $\mathcal{D}$, under the above model assumptions.

(b) Define a Bayesian model for the estimation of the unknown parameter θ using a non-degenerate prior distribution.
Hint: The gamma distribution has density supported on $\mathbb{R}_+$ and given by

$$\pi(\theta) = \frac{c^\gamma}{\Gamma(\gamma)} \theta^{\gamma-1} \exp\{-c\theta\} \;\mathbb{1}_{\{\theta > 0\}}, \tag{1}$$

with given parameters $\gamma, c > 0$. The corresponding mean and variance are given by γ/c and γ/c^2, respectively.

(c) Calculate the posterior estimator $\widehat{\theta}^{\mathrm{post}}$ for θ, given data $\mathcal{D}$, under the Bayesian model assumptions using π given in (1) as prior density.

(d) Give a credibility theory interpretation of the posterior estimator $\widehat{\theta}^{\mathrm{post}}$ derived in item (c). Which is the parameter driving prior uncertainty if we assume that the prior mean $\theta_0 = \gamma/c$ is a given constant? Give an argument for your answer.

(e) Derive the (conditional) mean square error of prediction of $\widehat{\theta}^{\mathrm{post}}$ derived in item (c). What happens with this error if $n \to \infty$? Give an argument for your answer.

Question 3 (10 **points**)

Assume we have n claims given by

$$\mathcal{D} = \{(Y_1, \boldsymbol{x}_1), \dots, (Y_n, \boldsymbol{x}_n)\}.$$

Assume that $\boldsymbol{x}_i \in \mathcal{X} = \mathbb{R}$, and that Y_i are independent and log-normally distributed for $i = 1, \dots, n$ with density supported in $\mathbb{R}_+$ and given by

$$Y_i \ \sim \ f(y|\boldsymbol{x}_i) = \frac{1}{\sqrt{2\pi}} \frac{1}{y} \exp\left\{ -\frac{1}{2} \left(\log y - \mu(\boldsymbol{x}_i) \right)^2 \right\}, \qquad \text{for } y \geq 0,$$

for a given (but unknown) regression function $\mu : \mathcal{X} \to \mathbb{R}$.

(a) Calculate the deviance statistics for this problem using a general regression function $\mu : \mathcal{X} \to \mathbb{R}$.

(b) Set up a single hidden layer neural network with 10 hidden neurons for this regression problem, using the hyperbolic tangent activation function given by $\phi(x) = (e^x - e^{-x})/(e^x + e^{-x})$ for $x \in \mathbb{R}$. Calculate the number of parameters of this model.

(c) Calculate one step of the gradient descent optimization algorithm explicitly for the deviance statistics loss function derived in item (a) and the single hidden layer neural network defined in item (b). Why is the hyperbolic tangent an attractive activation function in the application of the gradient descent algorithm?

(d) Choose the neural network defined in item (b) but replace the hyperbolic tangent activation function by the step function activation $\phi(x) = \mathbb{1}_{\{x \geq 0\}}$ for $x \in \mathbb{R}$.

  (i) Can the number of parameters of this regression function be reduced compared to the one in item (b) without affecting the regression function itself? If yes, how many parameters are sufficient? Justify your answer.

  (ii) How many different output values $\mu(\boldsymbol{x})$ can this regression model at most produce? Justify your answer.

(e) Assume that the feature space $\mathcal{X}$ is categorical having 11 different labels, i.e. the feature space is given by $\mathcal{X} = \{a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11\}$. We use dummy coding for these categorical feature components, and then we set up a single hidden layer neural network having one hidden neuron and hyperbolic tangent activation function.

  (i) Calculate the number of parameters that this regression model receives.

  (ii) How many different output values $\mu(\boldsymbol{x})$ can this regression model at most produce? Justify your answer.

Question 4 (10 points)

Assume we have n independent and identically distributed claims count observations given by the data

$$\mathcal{D} = \{(N_1, \boldsymbol{x}_1), \ldots, (N_n, \boldsymbol{x}_n)\}.$$

(a) What is the advantage of a log-linear regression function structure in terms of model interpretation of the coefficients considered? Give a short proof of your statement for the following regression function

$$\log \lambda(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_d x_d. \tag{2}$$

(b) You have the feature component `age` as a continuous variable in the regression function together with other continuous variables and categorical factors:

$$\log \lambda(x) = \beta_0 + \beta_1 \texttt{age} + \beta_2 x_2 + \ldots + \beta_d x_d.$$

You expected the variable `age` to be highly significant, but it is not. What could be the problem and how can you solve it? Give 2 possible solutions.

(c) One of your explanatory variables (feature component x_1) shows the following problem. The variable is significant in a model without the other explanatory variables:

$$\log \lambda(x) = \beta_0 + \beta_1 x_1.$$

But in a model with all explanatory variables given by (2) the significance of feature component x_1 vanishes while many other variables are significant. What is a possible reason? Would you use the variable x_1 for your final tariff? Give an argument for your decision.

(d) (i) Assume you have a categorical explanatory variable with many levels (lots of them with only few observations). How would you use this variable in your generalized linear model? Give 2 possibilities with their advantages and disadvantages.

(ii) Assume you have several continuous explanatory variables and you do not know the best functional form to include them in the generalized linear model. What can you do? Give 2 possibilities with their advantages and disadvantages.