Exam Pattern Recognition
Friday, February 5, 2021
11.30-14.30 hours

General Instructions

1. Write your name and student number on every sheet.
2. You are allowed to use a (graphical) calculator.
3. You are allowed to consult 1 A4 sheet of paper with notes (written or printed) on both sides.
4. Always show how you arrived at the result of your calculations.
   Otherwise you cannot get partial credit for incorrect final answers.
5. There are five questions for which you can earn 100 points.
6. Please write your answers to questions 1-3 and 4-5 on separate sheets. This statement is highly ambiguous, but if we tell you that questions 1,2, and 3 will be graded by Ad Feelders, and questions 4 and 5 will be graded by Zerrin Yumak, (and we would like to work in parallel) then you hopefully understand the idea.

Question 1: Linear Regression (20 points)

Chess is a board game for 2 players, where one player plays with the white pieces, and the other player with the black pieces. Each player makes a move in turn. White makes the first move. We analyse the results of 91 games of the TATA steel chess tournament that was held from 15-31 January, 2021. The target variable t is the score of the white player, and has the value 1 if white wins, the value 0 if black wins, and the value 0.5 if the players agree to a draw (tie). The strength of a player is measured by the so-called ELO rating, the higher the rating, the stronger the player. The predictor variable `elo.diff` is the ELO rating of the white player minus the ELO rating of the black player. We estimate a linear regression model using the method of least squares. This yields the following result:

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.5586287  0.0287357  19.440   <2e-16
 elo.diff      0.0005368  0.0003012   1.782   0.0782
---
Multiple R-squared:  0.03445
```

(a) (5 pnts) According to the fitted model, is there an advantage to playing with the white pieces? Explain your answer.

(b) (5 pnts) Is the effect of the predictor variable `elo.diff` on the target variable significant at significance level $\alpha = 0.05$? Explain how you determined the answer.

(c) (5 pnts) A famous Dutch grandmaster once wrote that chess is a game of chance. Given your assessment of how well the model fits the data, would you agree with this statement? Motivate your answer.

(d) (5 pnts) Compute the expected score for games where the white player's ELO rating is a thousand points higher than his opponent's. Does the outcome make sense? Explain.

Question 2: Logistic Regression (20 points)

We analyse a data set with 28×28 pixel images of handwritten digits. Each pixel has a grayscale value beteen 0 (white) and 255 (black). As computer scientists we see no need to go beyond the digits 0 and 1. We extract one feature from each image, called `ink`, which is defined as the sum of the pixel values divided by 1000. We analyse the data with logistic regression, where digit 1 is coded as 1, and digit 0 is coded as 0. We fit the model using maximum likelihood estimation, which gives the following result:

Coefficient	Estimate	Std. Error	z-value	$\Pr(>\|z\|)$
(Intercept)	10.98	0.26	42.22	$< 2 \times 10^{-16}$
`ink`	-0.47	0.01	-41.54	$< 2 \times 10^{-16}$

(a) (5 pnts) Give an interpretation in plain language of the negative sign of the coefficient for `ink`, that is, what does the negative sign of this coefficient mean?

(b) (5 pnts) Explain why it makes sense that we found a negative sign for the coefficient of `ink`.

(c) (5 pnts) Use the fitted model to estimate the probability that an image with `ink` $= 25$ contains the digit 0.

(d) (5 pnts) Use the fitted model to give a rule for the classification of digits.

Question 3: Support Vector Machines (20 points)

We receive the following output from the optimization software for fitting a support vector machine with linear kernel and perfect separation of the training data:

n	$x_{n,1}$	$x_{n,2}$	t_n	a_n
1	1	4	$+1$	0
2	2	5	$+1$	0
3	3	4	$+1$	0
4	7	4	$+1$	1
5	8	3	-1	1
6	9	2	-1	0
7	10	3	-1	0
8	10	4	-1	0

Here $x_{n,1}$ denotes the value of x_1 for the n-th observation, t_n denotes the class label of the n-th observation, and a_n the value of the Lagrange multiplier for the n-th observation. You are given the following formulas:

$$b = t_s - \sum_{n=1}^{N} a_n t_n \mathbf{x}_s^\top \mathbf{x}_n \qquad \text{(for any support vector } \mathbf{x}_s \text{ with label } t_s\text{)}$$

$$\mathbf{w} = \sum_{n=1}^{N} a_n t_n \mathbf{x}_n$$

Answer the following questions about the SVM output given above:

(a) (4 pnts) How many support vectors are there?

(b) (8 pnts) Give the equation of the maximum margin linear decision boundary.

(c) (4 pnts) Compute $y(\mathbf{x}_0)$ for $\mathbf{x}_0 = [8 \;\; 5]^\top$ and predict the class of $\mathbf{x}_0$.

(d) (4 pnts) Compute the value of the margin.

PLEASE WRITE YOUR ANSWERS TO QUESTIONS 4 AND 5 ON A SEPARATE SHEET!

Question 4: Deep Neural Networks (20 points)

(a) (6 pnts) What are the three modes of gradient descent used for training deep neural networks? Write the pseudocode for each one and explain the differences between them.

(b) (6 pnts) Explain the following terms: number of iterations, batch size, mini-batch and epoch.

(c) (8 pnts) Explain how the backpropagation algorithm works using the chain rule based on the figure given below. How does a small change in weight θ_2 affect the final loss $J(\theta)$? How does a small change in weight θ_1 affect the final loss $J(\theta)$? Write down the chain rule formulas with partial derivatives for these two cases.

$$x \xrightarrow{\quad \theta_1 \quad} z_1 \xrightarrow{\quad \theta_2 \quad} \hat{y} \longrightarrow J(\theta)$$

Question 5: Recurrent Neural Networks (20 points)

(a) (8 pnts) Why are recurrent neural networks hard to train and what is the advantage of gated cells (e.g. LSTM and GRU) in comparison to plain recurrent neural networks? Explain intuitively.

(b) (12 pnts) Draw the figure of an LSTM unit and write down the formula to calculate the values inside this unit. Define each of the variables.