# Resultaatexport

| | |
|---|---|
| **Aangemaakt door** | Ad Feelders |
| **Aanmaakdatum** | 16-02-2024 16:31 |
| **Hoofdstukken** | Details van het resultaat, Antwoorden |
| **Lege hoofdstukken** | Toevoegen |

# Inhoudsopgave

# Details van het resultaat

| | |
|---|---|
| **Kandidaat:** | Maria Dukmak |
| **Code/Kenmerk:** | 0550280 |
| **Kandidaatgroepen:** | #AlleKandidaten |
| **Cursus:** | Pattern Recognition and Deep Learning (INFOMPRDL) |
| **Toetsmatrijsnaam:** | [20240202] INFOMPRDL - Pattern Recognition and Deep Learning - 2 - USP |
| **Toetsmatrijssoort:** | Toets |
| **Status:** | Afgesloten |
| **Startdatum:** | 02-02-2024 17:00 |
| **Einddatum:** | 02-02-2024 18:26 |
| **Aantal vragen:** | 8 |

# Antwoorden

## Vraag 1

**Beantwoord op: 2024-02-02 17:09:45** **Duur: 6 min. en 10 sec.** **Weegfactor: 1** **Score: 10 van 14 pt.**

# TRUE OR FALSE?

For each of the statements below, indicate whether it is true or false.

In generative classification models, one models the joint probability distribution of the features and the class label.

*Score: 2 van 2 pt.*

- ◉ TRUE ✓ 2 pt.
- ○ FALSE 0 pt.

In linear discriminant analysis, it is assumed that the covariance matrix of the features is the same within each class.

*Score: 2 van 2 pt.*

- ◉ TRUE ✓ 2 pt.
- ○ FALSE 0 pt.

Regularization tends to reduce bias and variance.

*Score: 2 van 2 pt.*

- ○ TRUE 0 pt.
- ◉ FALSE ✓ 2 pt.

A support vector machine with non-linear kernel can always achieve perfect separation of the training data.

*Score: 2 van 2 pt.*

- ○ TRUE 0 pt.
- ◉ FALSE ✓ 2 pt.

In a linear regression problem with a single binary feature $x \in \{0,1\}$, the least squares estimates of $w_0$ and $w_1$ in $y(x) = w_0 + w_1 x$, are $w_0 = \mu_0$ , and $w_1 = \mu_1$ , where $\mu_0$ is the mean t value for the training examples with

x=0 and $\mu_1$ is the mean t value for the training examples with x=1.

*Score: 0 van 2 pt.*

| | | |
|---|---|---|
| ◉ TRUE | ✗ | 0 pt. |
| ○ FALSE | ✓ | 2 pt. |

Application of linear regression to binary classification problems may lead to negative predicted probabilities.

*Score: 2 van 2 pt.*

| | | |
|---|---|---|
| ◉ TRUE | ✓ | 2 pt. |
| ○ FALSE | | 0 pt. |

In k-nearest neighbor regression, as the value of k increases, the regression function becomes more complex (less smooth).

*Score: 0 van 2 pt.*

| | | |
|---|---|---|
| ◉ TRUE | ✗ | 0 pt. |
| ○ FALSE | ✓ | 2 pt. |

## Vraag 2

**Beantwoord op: 2024-02-02 17:24:33** **Duur: 18 min. en 3 sec.** **Weegfactor: 1** **Score: 10 van 15 pt.**

# Linear Regression

A computer science department keeps track of the  starting salaries of its graduates. We are interested in the effect of the grade point average, gender, and the deep learning course, on the student's starting salary. We have data of the 50 most recent graduates (N=50). For each graduate, the following data has been recorded: Salary (in euro's), GPA = grade point average, DL=1 if the student took the course in deep learning (and 0 otherwise), and Gender=1 if the student is female (and 0 otherwise). We fit a linear regression model using the method of least squares, and obtain the following results:

| | Estimate | p-value |
|---|---|---|
| (Intercept) | 24241.7 | 0.000 |
| GPA | 1657.8 | 0.000 |
| DL | 5023.5 | 0.000 |
| Gender | -204.8 | 0.628 |

The sum of squared errors (SSE) is 96,226,078 and the (unbiased) estimate of the variance of Salary is 7,502,130.

Answer the following questions:

(a) What salary does the fitted model predict for a male student with a GPA of 3.5, who took the deep learning course? (don't round your answer)

Predicted Salary: 35.067,5 5 pt. 35.067,5 5 pt.

(b) What percentage (between 0 and 100; round to whole numbers) of the variation in Salary is explained by the regression model?

Explained variation: 7 0 pt. 74 (+/- 1) 5 pts. %

(c) What feature (derived from the given features) would you add to the model if you want to determine whether the added value of deep learning is different for men and women?

New feature: gender * DL 5 pt. DL*Gender 5 pt. DL * Gender 5 pt. Gender*DL 5 pt. Gender * DL 5 pt.

## Vraag 3

**Beantwoord op: 2024-02-02 18:14:48** **Duur: 12 min. en 35 sec.** **Weegfactor: 1** **Score: 12,33 van 15 pt.**

# Logistic Regression

A bank is analyzing a data set of accepted loan applicants. For each accepted applicant the age (in years), debt ratio (fixed monthly expenses divided by monthly income), and whether or not the applicant defaulted on the loan are recorded. The target variable has the value 1 if the applicant defaulted, and 0 otherwise. We fit a logistic regression model with age and debt ratio as features, using the method of maximum likelihood. This produces the following results:

| Feature | Estimate | p-value |
|---|---|---|
| (Intercept) | 1.386 | 0.01362 |
| age | -0.034 | 0.00192 |
| debt ratio | 0.760 | 0.05934 |

According to the fitted model, the probability that a 50 year old applicant with a debt ratio of 0.75 defaults on the loan is (round your final answer to 3 decimal places):

*Score: 5 van 5 pt.*

- ○ 0.256 — 0 pt.
- ◉ 0.564 — ✓ 5 pt.
- ○ 0.436 — 0 pt.
- ○ 0.244 — 0 pt.

Antwoorden

Which of the following statements is correct? (Note: one or more statements may be correct!)

According to the fitted model:

*Scoringsmethodiek: productregel* | *Basisscore: 10 pt.* | *Score: 7,33 van 10 pt.*

- [x] All else equal, younger people have a higher probability of default than older people. ✓
- [ ] All else equal, men have a higher probability of default than women.
- [x] For every one-unit increase of the debt ratio, the odds of default against no default are more than doubled. ✓
- [ ] At significance level α = 0.05, the debt ratio has a significant influence on the probability of default.
- [x] If we want to maximize accuracy, we should predict class 1 (default) if
  0.034 x age - 0.76 x debt ratio > 1.386,
  and class 0 (no default) otherwise. ✗

## Vraag 4

**Beantwoord op: 2024-02-02 18:23:28** | **Duur: 13 min. en 15 sec.** | **Weegfactor: 1** | **Score: 9 van 15 pt.**

# Support Vector Machines

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 | 3 | 4 | +1 | 0 |
| 2 | 2 | 2 | +1 | 2 |
| 3 | 4 | 4 | +1 | 2 |
| 4 | 1 | 4 | +1 | 0 |
| 5 | 2 | 1 | -1 | 1 |
| 6 | 4 | 3 | -1 | 3 |
| 7 | 4 | 1 | -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, etc.

You are given the following formulas:

$$b = t_s - \sum_{n=1}^{N} a_n t_n \mathbf{x}_s^\top \mathbf{x}_n$$

for any support vector $x_s$ with label $t_s$

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

Answer the following questions about the maximum margin decision boundary:

(a) The value of the bias term is:

| | | |
|---|---|---|
| 1 4 pt. | | 1 4 pt. |

(b) The coefficient of $x_1$ is:

| | | |
|---|---|---|
| 2 0 pt. | | -2 3 pt. |

(c) The coefficient of $x_2$ is:

| | | |
|---|---|---|
| -2 0 pt. | | 2 3 pt. |

(d) The value of the margin is (round to 2 decimal positions):

| | | |
|---|---|---|
| 0,35 5 pt. | 0.35 (+/- 0.01) 5 pts. | 0.71 (+/- 0.01) 5 pts. |

## Vraag 5

**Beantwoord op: 2024-02-02 17:01:07** **Duur: 1 min. en 50 sec.** **Weegfactor: 1** **Score: 5 van 5 pt.**

# Deep Learning I

Which one of the following statements is NOT correct?

*Score: 5 van 5 pt.*

| | | |
|---|---|---|
| ○ | In dropout, at test time all neurons are tuned on and the output is scaled. | 0 pt. |
| ○ | The sigmoid activation function has two saturated regions that might kill the gradient. | 0 pt. |
| ◉ | The Rectified Linear Unit (ReLU) activation function will produce zero centered output. ✓ | 5 pt. |
| ○ | Deep learning uses automatic differentiation in reverse mode for computing the gradients. | 0 pt. |

## Vraag 6

Beantwoord op: 2024-02-02 17:41:26 | Duur: 5 min. en 0 sec. | Weegfactor: 1 | Score: 0 van 10 pt.

# Deep Learning II

Suppose a sigmoid activation function is applied elementwise on a 150 dimensional input vector, which are all negative numbers in the range [-3,-2].

What is the size of Jacobian matrix associated with the input-output of this operator? How many elements of this Jacobian matrix are strictly positive, ( i.e. greater than zero)?

Hint: $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$.

*Score: 0 van 10 pt.*

- (●) The Jacobian will be of size 150, with 150 strictly positive numbers. ✗ 0 pt.
- ( ) The Jacobian will be of size 150, with all negative numbers. 0 pt.
- ( ) The Jacobian will be of size 150 x 150, with 150 strictly positive numbers on the main diagonal. ✓ 10 pt.
- ( ) The Jacobian will be of size 150 x 150, with 150 x 150 = 22,500 strictly positive numbers. 0 pt.

## Vraag 7

Beantwoord op: 2024-02-02 18:25:57 | Duur: 8 min. en 15 sec. | Weegfactor: 1 | Score: 0 van 10 pt.

# Deep Learning III

Suppose the input to a ReLU layer as well as the gradient from the layer above it (next layer) are given as follows.

$$\text{Input} = \begin{bmatrix} 2 & -1 & -3 & 12 \\ -1 & 6 & 5 & -9 \\ -3 & 5 & 7 & 1 \\ 4 & -9 & 6 & 8 \end{bmatrix} \quad \text{Gradient from above} = \begin{bmatrix} 5 & 3 & 2 & 11 \\ 8 & -6 & -5 & 4 \\ 4 & 2 & -3 & -6 \\ 9 & 4 & -5 & 12 \end{bmatrix}$$

What would be the backpropagation for this ReLU layer?

*Score: 0 van 10 pt.*

- ◉ $\begin{bmatrix} 2 & 0 & 0 & 12 \\ 0 & 6 & 5 & 0 \\ 0 & 5 & 7 & 1 \\ 4 & 0 & 6 & 8 \end{bmatrix}$ ✗ 0 pt.
- ○ $\begin{bmatrix} 5 & 0 & 0 & 11 \\ 0 & -6 & -5 & 0 \\ 0 & 2 & -3 & -6 \\ 9 & 0 & -5 & 12 \end{bmatrix}$ ✓ 10 pt.
- ○ $\begin{bmatrix} 5 & 3 & 2 & 11 \\ 8 & 0 & 0 & 4 \\ 4 & 2 & 0 & 0 \\ 9 & 4 & 0 & 12 \end{bmatrix}$ 0 pt.
- ○ $\begin{bmatrix} 2 & -1 & -3 & 12 \\ -1 & 0 & 0 & -9 \\ -3 & 5 & 0 & 1 \\ 4 & -9 & 0 & 8 \end{bmatrix}$ 0 pt.

## Vraag 8

**Beantwoord op: 2024-02-02 18:26:51** **Duur: 15 min. en 6 sec.** **Weegfactor: 1** **Score: 16 van 16 pt.**

# Deep Learning IV

Consider a batch of training data composed of 80 RGB color images with 32x32 spatial dimension. We have applied a 2d-convolutional layer with 20 filters of size 5x5 with ReLU activation function followed by average pooling layer for all these 80 images.

What is the total number of parameters in these two layers?

*Score: 8 van 8 pt.*

- ○ 520 0 pt.
- ○ 41,600 0 pt.
- ◉ 1,520 ✓ 8 pt.
- ○ 121,600 0 pt.

If we add a second convolutional layer (2d) with the same filter size to the output of the first conv layer and we apply the max pooling layer, does the total number of parameters of these two new layers (conv + max pooling) change with respect to the first two layers (convolutional + average pooling) ?

*Score: 8 van 8 pt.*

| | | |
|---|---|---|
| ○ | No, it remains the same. | 0 pt. |
| ◉ | Yes, and it will increase. ✓ | 8 pt. |
| ○ | Yes, and it will decrease. | 0 pt. |