Online Test 3 (30%)

- Due 20 Nov at 3:59
- Points 100
- Questions 27
- Available 19 Nov at 12:30 - 20 Nov at 3:59 15 hours and 29 minutes
- Time limit 100 Minutes

Instructions

Due Date: Week 12 Lab Session

Weighting: 30%

About this test:

1. **You must participate in this test in person on campus.**
2. This is an online test worth 30% of the total marks for this unit.
3. It consists of 25 questions on theories and applications. Using KNIME in the quiz is essential to find answers.
4. You may **only attempt this test once**.
5. **Do not navigate away from the test (close the browser tag) before you have completed it and submitted it.**
6. You have **100 minutes** to complete the test; exceeding this will result in a loss of marks and penalties.
7. This is a **close-book test**.
8. Once you begin the quiz, a summary will appear in the upper right corner showing the number of questions (completed and total) and the time remaining.
9. **Please bring your student ID to the test and put it on the desk in front of you so the tutor can verify your identity.**
10. **Please use Chrome to access the quiz. The second options are Firefox and Microsoft Edge. Please do not use Safari in the test.**
11. **The multi-answering questions have a deduction mechanism. Be careful about what you select in the answers.**

This quiz was locked 20 Nov at 3:59.

Attempt history

	Attempt	Time	Score
LATEST	Attempt 1	100 minutes	80 out of 100

Score for this quiz: 80 out of 100
Submitted 19 Nov at 14:11

This attempt took 100 minutes.

Question 1

4 / 4 pts

Which definition of Lift in association rule is TRUE?

(A) $Lift\,(X \Longrightarrow Y) = \frac{Support(X \times Y)}{Support(X) \times Support(Y)}$

(B) $Lift\,(X \Longrightarrow Y) = \frac{Support(X \cup Y)}{Support(X) \times Support(Y)}$

(C) $Lift\,(X \Longrightarrow Y) = \frac{Support(X) \times Support(Y)}{Support(X \cup Y)}$

(D) $Lift\,(X \Longrightarrow Y) = \frac{Support(X) + Support(Y)}{Support(X \cup Y)}$

Please note that the order of the answer can be different than what is listed in the question.

Look at the answers carefully.

- (D)
- (C)
- (A)

Correct!

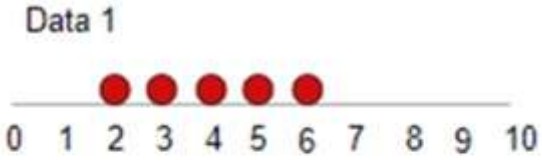
- (B)

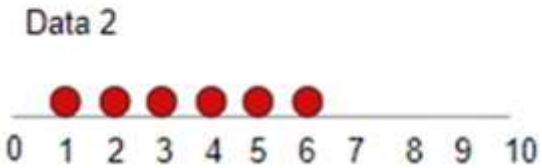
Week 05 - Lecture - P24.

Question 2

4 / 4 pts

Given two datasets as in the figures below:





Which of the following is (are) correct?

Correct!

- [x] The mean value of data 2 is smaller than data 1.
- [ ] There is an outlier in data 2. By removing that outlier, data 2 will become normal.

Correct!

- [x] The median of data 1 is 4.
- [ ] The median of data 2 is 5.

Week 09 - Lecture - P11 - P12.

Question 3

4 / 4 pts

Matching the problem to solve to the most suitable technique.

Correct!

Grouping items by similarity.

Clustering

Correct!

Finding the structure in a temporal process.

Time series analysis

Correct!

Determining the relationship between the outcome and the input variables.

Regression

Correct!

Discovering relationships between actions or items.

Association rules

Correct!

Assigning labels to objects.

Classification

Week02 - Lecture - P4.

Question 4

4 / 4 pts

Which of the description below is NOT the question that can be answered by statistics test?

- Does the model have the desired effect?
- Does the treatment relieve the symptom of the patients?
- Is April this year colder than the same time last year?

Correct!

- Who should I vote for in the next election?

Question 5

4 / 4 pts

Which of the following rule is NOT the rules to be checked in the data cleaning process?

- Consequtive rule
- Null rule
- Unique rule

Correct!

- Association rule

Week03 - Lecture - P21.

Question 6

4 / 4 pts

Some descriptions of the artificial neural network (ANN) are given below.

Select the description(s) that is (are) TRUE.

Correct!

- [x] If the calculation result satisfies the output criteria, the neuron sends out the output signal to the desired destination.
- [ ] A single neuro in ANN takes only single input and performs the designed calculation for generating the output to the next neuro.

Correct!

- [x] Input signals and the bias value are collected and the defined calculation is performed in the neuron.
- [ ] The bias would be set to 1 when a neuro in ANN decides to enable its output.

Week 10 - Lecture - P7.

Question 7

4 / 4 pts

Matching the terms in data transformation from the left to the descriptions on the right.

Correct!
Smoothing

Remove noise from data us

Correct!
Aggregation

Construct data cubes.

Correct!
Discretization

Replace raw values of a nur

Correct!
Normalization

Scale the attribute data to fa

Week 03 - Lecture - P51.

Question 8
4 / 4 pts
Matching the terms from the left to the descriptions on the right.
Correct!
Mean and median

Representing a centre arour

Correct!
Standard deviation

Representing the spread of

Correct!
Quantiles

Representing the relative pc

Correct!

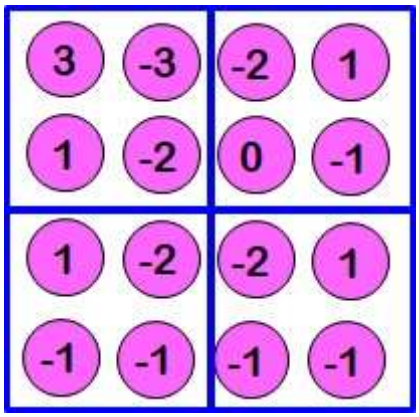
Exploratory data analysis approach

Analysing datasets to summ

Week 04 - Lecture - P10, P11.

Question 9

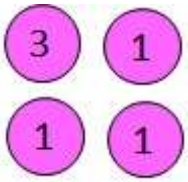
4 / 4 pts



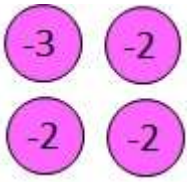
The above figure shows a feature map from a convolutional neural network (CNN).

Which of the below options should be the max-pooling output result?

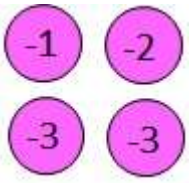
(i)

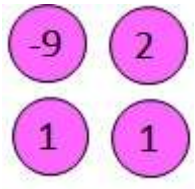


(ii)



(iii)



(iv) 

Please note that the order of the answer can be different than what is listed in the question.

Look at the answers carefully.

- (ii)
- (iii)
- (iv)

Correct!

- (i)

Week 10 - Lecture - P70.

Question 10

4 / 4 pts

Select the description(s) that is (are) TRUE.

Correct!

- [x] Data security is essential for maintaining the confidentiality, integrity, and availability of sensitive information.
- [ ] Data security is optional for maintaining the confidentiality, integrity, and availability of sensitive information, such as personal information, financial data, and intellectual property.
- [ ] Data security is also known as cybersecurity in data science projects.

Correct!

- [x] Failure to implement adequate data security measures can result in serious consequences, such as loss of business reputation, financial loss, legal liability, and damage to individual privacy and rights.

Week 07 - P113.

Question 11

4 / 4 pts

Select the correct formula of the confidence in association rule.

(A) $Confidence\,(X \longrightarrow Y) = \frac{Support(X \cap Y)}{Support(X)}$

(B) $Confidence\,(X \longrightarrow Y) = \frac{Support(X \cup Y)}{Support(Y)}$

(C) $Confidence\,(X \longrightarrow Y) = \frac{Support(X \times Y)}{Support(Y)}$

(D) $Confidence\,(X \longrightarrow Y) = \frac{Support(X \cup Y)}{Support(X)}$

Please note that the order of the answer can be different than what is listed in the question.

Look at the answers carefully.

Correct!

- (D)
- (A)
- (C)
- (B)

Week 05 - Lecture - P22.

Question 12

4 / 4 pts

Select the description(s) that is (are) TRUE.

- [ ] T-test is used to compare samples based on their standard deviations.

Correct!

- [x] If the t-test result indicates that the samples are not significantly different from each other, it implies that the samples are highly overlapping on each other when plotting the values on figures.
- [ ] In t-test, we usually determine whether the test result is significantly different by looking at the R-value.

Correct!

- [x] Hypothesis test is a common approach to draw inferences on whether the two populations are different from each other.

Week 04 - Lecture - P36.

Question 13

4 / 4 pts

Match the contents to their correct groups.

Correct!

True positive rate (TPR)

Sensitivity

Correct!

True negative rate (TNR)

Specificity

Correct!
Precision

Calculated by the true positi

Correct!
Area under the curve (AUC)

The plasma concentration-ti

Week 08 - Lecture - P40.

Question 14
4 / 4 pts
Which of the following descriptions is generally used in a statistics test?
Correct!

- Assume the body weight of the participant is the same before and after the diet control.
- Assume the body weight of the participant is heaver after the diet control.
- Assume the body weight of the participant is lighter after the diet control.
- Assume the body weight of the participant is different before and after the diet control.

Question 15
4 / 4 pts

Which definition of Accuracy in the predictive model evaluation is TRUE?

(A) $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$

(B) $Accuracy = \frac{TP+TN+FP+FN}{TP+TN} \times 100$

(C) $Accuracy = \frac{TP+TN}{FP+FN} \times 100$

(D) $Accuracy = \frac{TP \times TN}{TP+TN+FP+FN} \times 100$

TP, TN, FP, and FN stand for the true positive, the true negative, the false positive, and the false negative, respectively.

Please note that the order of the answer can be different than what is listed in the question.

Look at the answers carefully.

(B)

Correct!

(A)

(C)

(D)

Week 08 - Lecture - P44.

Question 16

0 / 4 pts

2-D Matrix:

1	0	1	0
1	1	0	1
0	0	1	1
1	0	0	0

Mask:

-2	0	2
-3	4	1
1	0	-1

Convolution result:

-1	0	0	0	1	0
0	3	-1	4	-3	1
3	5	0	-3	3	-2
1	2	0	5	1	**?**
1	4	-1	2	-2	-2
2	0	-2	0	0	0

The above matrices are a 2-D matrix, a 3x3 mask, and their convolution result.

Which of the values should the question mark represent?

- 3
- 5

You Answered

- -3

Correct answer

- -5

-1	0	0	0	1	0
0	3	-1	4	-3	1
3	5	0	-3	3	-2
1	2	0	5	1	-5
1	4	-1	2	-2	-2
2	0	-2	0	0	0

Question 17

4 / 4 pts

Select the description(s) that is (are) TRUE.

Correct!

- [x] Rejecting a hypothesis does not always equate a failure. Instead, a failure usually refers to the inability to strike the balance between two possible analytics extremes.
- [ ] Rejecting a hypothesis means the project is failed. The possible option is to restart the project from the beginning again with other assumptions.
- [ ] The success or failure criteria are determined by whether the data and the chosen analytics models are able to accept or reject the initial hypotheses formulated in data analytics lifecycle phase 2.

Correct!

- [x] The success or failure criteria are determined by whether the data and the chosen analytics models are able to accept or reject the initial hypotheses formulated in data analytics lifecycle phase 1.

Week 11 - Lecture - P8.

Question 18

4 / 4 pts

Given a dataset containing 7 points: **{2, 5, 6, 7, 10, 12, 13}**.

Which of the following is (are) correct?

Correct!

- [x] The sample sdandard deviation of this dataset is 3.976

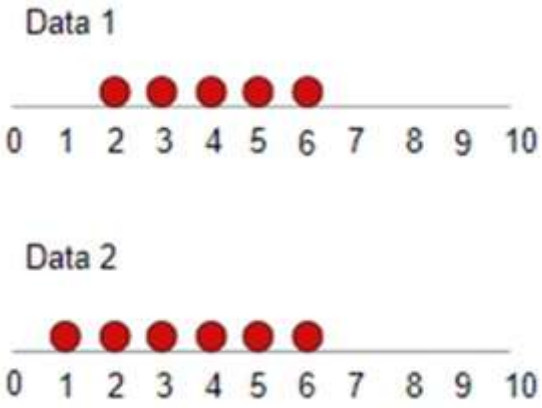
Correct!

- [x] The population sdandard deviation of this dataset is 3.681
- [ ] The mean value of this dataset is 7.
- [ ] The population sdandard deviation of this dataset is 3.976

Week 09 - Lecture - P13.

Question 19
4 / 4 pts

Given two datasets as in the figures below:



Which of the following is (are) correct?

- [ ] There is an outlier in data 2. By removing that outlier, data 2 will become normal.
- [ ] The median of data 2 is 5.

Correct!

- [x] The mean value of data 2 is smaller than data 1.

Correct!

- [x] The median of data 1 is 4.

Week 09 - Lecture - P11 - P12.

Given the smartphone dataset below, try to process the data with KNIME or any preferred tool to answer the questions.

Dataset: **phone_dataset.csv (https://swinburne.instructure.com/courses/62961/files/33245888?wrap=1)** **(https://swinburne.instructure.com/courses/62961/files/33245888/download?download_frd=1)**

The dataset is pre-shuffled and thus you are required to use the "Take from top" method to extract 90% of the data into the training set and the remaining go into the test set.

Question 20

0 / 4 pts

Build a Naive Bayes classifier to predict the "price_range".

Since the "price_range" has been quantified into integer ranges, you can treat them as categorical data.

Use all available attributes as the input and use default values for the classifier parameters.

Which of the following description(s) is (are) TRUE?

Correct answer

- [ ] The accuracy is 76.5% on the test result.
- [ ] The price_range 3 class has the highest recall value because it has the lowest false positive count.

You Answered

- [x] There is a relatively higher chance that the predicted price_range 1 item is actually in other price_range classes.

Correct answer

- [ ] The trainned model performs the worst in price_range 1 group because it has the lowest precision and recall values.

Question 21

2 / 4 pts

Build a Naive Bayes classifier to predict the "price_range".

Since the "price_range" has been quantified into integer ranges, you can treat them as categorical data.

Use all available attributes as the input and use default values for the classifier parameters.

Which of the following description(s) is (are) TRUE?

- [ ] The price_range 1 class has the lowest precision because it has the lowest false negative count.

Correct!

- [x] The price_range 3 class has the highest recall value because the model has less false negative results generated from this class.

Correct!

- [x] The trainned model performs relatively better when encountering data belongs to price_range 0 or 3 classes than price_range 1 or 2 classes.

You Answered

- [x] The high false positive count in the price_range 1 class causes the low precision value of this class.

Question 22

0 / 4 pts

By analysing the correlations between the input attributes, we decide to exclude "fc," "px_height," and "sc_w" from the input.

Create a Multi-layer Perceptron (MLP) model with the following settings:

- Maximum iteration: 1,800
- Number of hidden layers: 2
- Number of neurons per layer: 18
- Ignore missing values
- Use seed for random initialisation with value 4

Our target is to create the MLP model for predicting the "price_range".

Create a KNIME workflow to find answers.

Which of the following description(s) is (are) correct?

Correct answer

- [ ] The trained MLP puts half of the price_range 2 class data into the false negative cases and thus the recall is 0.5

Correct answer

- [ ] The test accuracy is only 59%.

You Answered

- [x] The trainned model performs relatively better when encountering data belongs to price_range 1 or 2 classes than price_range 0 or 3 classes.
- [ ] The low false positive count enables the price_range 0 class has a relatively high recall value.

Given the medical record dataset below, try to process the data with KNIME or any preferred tool to answer the questions.

Dataset: **Drugged_Data.csv (https://swinburne.instructure.com/courses/62961/files/33246095?wrap=1)** **(https://swinburne.instructure.com/courses/62961/files/33246095/download?download_frd=1)** **(https://swinburne.instructure.com/courses/40743/files/18907493?wrap=1)**

This dataset is collected for studying the anti-anxiety medicine effects on memory recall when primed with happy or sad memories.

Data Dictionary:

- first_name: First name of the patient.

- last_name: Last name of the patient.
- age: The age of the patient.
- Happy_Sad_group: H means happy while S means sad.
- Dosage: Records the different volumes of drugs being applied.
- Drug: A - Alprazolam (Xanax, Long-term) [1mg/3mg/5mg]; T - Triazolam (Halcion, Short-term) [0.25mg/0.5mg/0.75mg]; S- Sugar Tablet (Placebo) [1 tab/2tabs/3tabs].
- Mem_Score_Before: The quantified memory score before the dosage.
- Mem_Score_After: The quantified memory score after the dosage.
- Diff: The difference between Mem_Score_After and Mem_Score_Before.

*Dosages follow a 1:1 ratio to ensure validity

*Happy or Sad memories were primed 10 minutes prior to testing

*Participants tested every day for 1 week to mimic addiction

In some analysis, you may need to refer to the Chi-square table attached below.

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Question 23

0 / 4 pts

Select the right statistical test method to apply and answer the question below.

Which of the description(s) below is (are) correct?

You Answered

- [x] Having the highest dosage of different drugs significantly affects the mood (happy or sad).

Correct!

- [x] Having different dosage of Triazolam doesn't affect the mood at all.

Correct answer

- [ ] Having different drugs doesn't affect the mood (happy or sad).
- [ ] Whether age above 50 has a significant impact on the mood.

Question 24

2 / 4 pts

If we only look at the happy participants (Happy_Sad_group = "H") receiving Alprazolam, which of the following description(s) is (are) correct?

Correct!

- [x] The memory scores of participants for both the low dosage (1ml) and high dosage (5ml) groups before receiving the dosage basically have no differences.

Correct answer

- [ ] Having a low dosage (1mg) and high dosage (5mg) results in significant different memory scores.
- [ ] The age of participants receiving the low dosage (1ml) is significantly younger than the high dosage (5ml) group.
- [ ] The memory scores of participants for both the low dosage (1ml) and high dosage (5ml) groups before receiving the dosage basically have significant differences.

Question 25

4 / 4 pts

If we only look at the sad participants (Happy_Sad_group = "S") receiving Triazolam, which of the following description(s) is (are) correct?

- [ ] The memory scores of participants for both the low dosage (0.25 ml) and high dosage (0.75 ml) groups before receiving the dosage basically have significant differences.
- [ ] If we assume the variances between groups are equal, the memory score after dosage would be significant between the low dosage (0.25 ml) and the high dosage (0.75 ml) groups.

Correct!

- [x] The average age between the low dosage (0.25 ml) group and the high dosage (0.75 ml) group has a 6.3455 difference but the difference is not significant.

Correct!

- [x] The memory scores of participants for both the low dosage (0.25 ml) and high dosage (0.75 ml) groups before receiving the dosage basically have no differences.

Quiz score: 80 out of 100