Information

Question	Question title	Marks	Question type
i	Information		Document

Dimensionality Reduction

Question	Question title	Marks	Question type
1	PCA	10	Composite

Classification

Question	Question title	Marks	Question type
2	Decision Tree	16	Composite
3	Naive bayes properties	5	Composite
4	Naive bayes	15	Composite

Deep Learning

Question	Question title	Marks	Question type
5	Neural network	6	Composite
6	Learning curves	9	Matching
7	Activation functions and neural networks	9	Composite
8	RNNs	6	Composite

Clustering

Question	Question title	Marks	Question type
9	Clustering 1	12	Composite
10	Clustering 2	12	Composite

Scratch area

Question	Question title	Marks	Question type
11	Provide your explanation for numerical computations here!	0	Text area

ℹ **Information**

Exam: DAT550 Data Mining
Date: May 8, 2020
Duration: 2 hours
Supporting material: All technical support material is permitted. You are not allowed to get help from other people when working on your exam assignment. We are also reminding you that you, when registering for the semester, you signed that you have read and understood the rules for cheating and plagiarism in the Exam Rules and Regulations at the University of Stavanger. Plagiarism control will be carried out.

The exam consists of 10 problems on a total of 10 pages.
Important contact information during the exam: If something should happen during the exam and you need help, you can call one of the phone numbers below, both if you need technical support, administrative support or have questions for.

1. course responsible. Vinay Setty, tlf. 518 32760, If you spot any errors in the exam or have confusion you can call me and if the error needs to be fixed by everyone I will send an email to everyone.

2. Administrative support: 51 83 17 15 / 51 83 31 33 / 92 81 65 97 / 91 78 67 16 Technical support: 51 83 20 14/51 83 20 30

Your responses should be entered into Inspera Assessment.

There is no negative scoring on any of the questions! Hope you appreciate that :)
For some multiple choice questions there are multiple correct answers so pay attention.

Important Information about Exams in Inspera

If you have been given extra time on the exam, it has been added to your used. You can see the count down on the top of the page. Here you will also find your candidate number. The candidate number should be written on <u>all pages</u>. Remember to write page number on all pages and total page number on the first page. NB! Do not write your student number, name or anything else that can identify you in you file or on your answer sheets, only your candidate number. Inspera will secure your identity and will ensure that the evaluation is anonymous.

You can write your answers using blank, ruled or squared paper. You can download and print a squared sheet with weak squares that are appropriate for scanning from Canvas. You can also write your answers in a program such as Latex if you prefer to. The answers are to be handed in as a pdf-file.

Handing in

The exam will automatically close for uploading when the time is up. Remember that the time given includes the time it will take you to scan and upload your documents (see attachment for tips concerning scanning and uploading). We recommend that you get yourself informed of how to best carry these steps out in due time and before you start working on the assignments.

Best of luck!

We will strictly follow the following grading scale which is a standard grading scheme at UiS:

A: > 90
B: > 80
C: > 60
D: > 50
E: > 40
F: 0

1 **PCA**

The goal of PCA is to interpret the underlying structure of the data in terms of the principal components that are best at predicting the output variable.

Select an alternative

○ True ✔

○ False

The output of PCA is a subset of the original features in lower dimensions

Select an alternative

○ False ✔

○ True

The output of PCA is a new representation of the data that is always of lower dimensionality than the original feature representation.

Select an alternative

○ False ✔

○ True

Subsequent principal components are always orthogonal to each other

Select an alternative

○ False

○ True ✔

Assume we apply PCA to a matrix $X \in R^{n \times m}$ and obtain a set of PCA features, $Z \in R^{m \times n}$. We divide this set into two parts, Z_1 and Z_2. The first part, Z_1, corresponds to the top principal components. The second set, Z_2, corresponds to the remaining principal components. Is it common to expect a point with large feature values in Z_2 and small feature values in Z_1?

Select an alternative

○ False ✔

○ True

2 Decision Tree

It is late Norwegian summer and you decide to go for a mushroom picking with your friends. There are lots of types of mushrooms in the forest, but some of them can be deadly because they may be poisonous. Fortunately, one of your friends has collected some attributes of the poisonous and non-poisonous mushrooms from the past trip as below:

P.S: Don't use this data to classify mushrooms in real life! :D

Sample	Heavy	Smelly	Spotted	Scales	Poisonous
a	0	0	0	0	0
b	0	0	1	0	0
c	1	1	0	1	0
d	1	0	0	1	1
e	0	1	1	0	1
f	0	0	1	1	1
g	0	0	0	1	1
h	1	1	0	0	1

(a) What is the entropy of Poisonous label? [ ] (0.95 - 0.955)

(b) Which of the attributes should you choose as root of the decision tree? Hint: You may have to compute information gain. [▼] (Scales, Smelly, Spotted, Heavy)

(c) What is the information gain of the you got for the attribute you chose the previous questions? [ ] (0.04 - 0.05)

(c) Build a decision tree to classify a mushroom which has all of the attributes as 1

Select one alternative

○ Poisonous (1) ✔

○ Not Poisonous (0)

Maximum marks: 16

3 **Naive bayes properties**

Which of the following are FALSE about Naive bayes?

Select one or more alternatives:

☐ Naive bayes can deal with missing attributes

☐ The Naive Bayes algorithm is sensitive to irrelevant attributes.

☐ Naive bayes is an unsupervised classification algorithm ✔

☐ Naiveness in naive bayes comes from the conditional independence assumption.

☐ If there are dependencies between the features Naive bayes is preferred. ✔

Maximum marks: 5

4 Naive bayes

(16%) Consider the training and testing data in the Table below. Classify the test records in Table b using the Naive Bayes classifier trained on the training data in Table a. You only need to compute the probabilities you will need for the classification. For your answer, you need to tell which class has the highest posterior probability – you do not have to compute the final posteriors as long as it is clear which one is bigger.

X	Y	X	Class
1	1	1	+
1	0	0	-
1	1	1	+
1	1	0	+
1	1	1	-
1	1	0	-
0	1	1	-
0	1	1	+
0	0	1	+
0	0	0	+

Using the naive bayes for training from above training data We are interested in classifying two test data entries:
Test1: X=0, Y=1, Z=0
Test2: X=1, Y=0, Z=1

Compute the likelihood of test cases belonging to certain class.

Note: You can omit the final posterior (the denominator) for computing the P(X,Y,Z) in the bayes theorem formula to fill in the following probabilities.

P[C =+|X =0,Y =1,Z =0] = ☐ (0.06 - 0.067)

P[C =-|X =0,Y =1,Z =0] = ☐ (0.03 - 0.038)

P[C =+|X =1,Y =0,Z =1] = ☐ (0.06 - 0.07)

P[C =-|X =1,Y =0,Z =1] = ☐ (0.03 - 0.038)
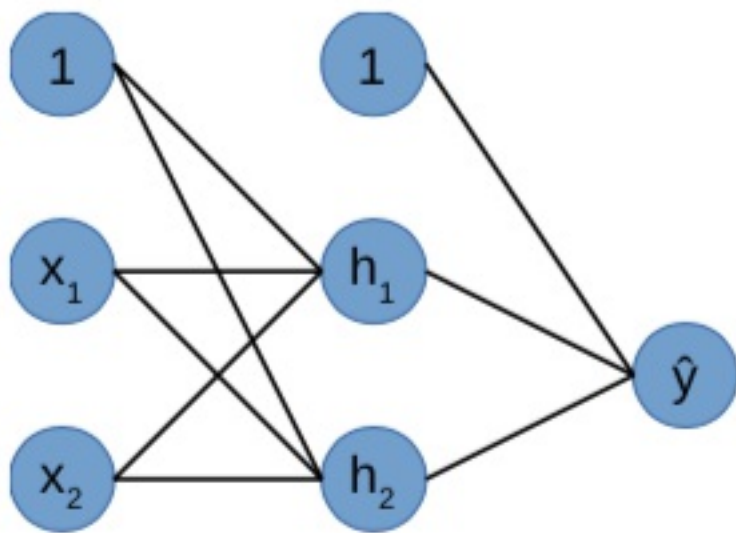
Class for X=0, Y=1, Z=0? ☐▾ (+, -)

Maximum marks: 15

5   Neural network

(5%) Suppose you are given the following neural network (flowing left to right) and param- eters, which uses ReLU activation in every layer and the loss is the squared error L = (y − y')^2.

Weights for the first layer (start from bias) $W_1 = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}$

Weights for second layer (start from bias) $W_2 = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$
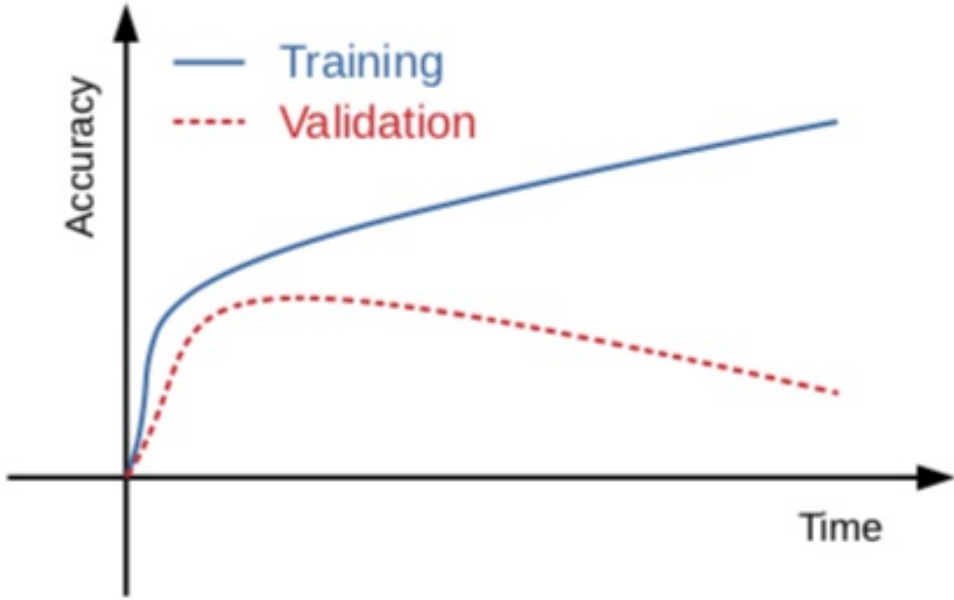

What is the loss value (squared loss) for the input x = (2 4) and y = 26 ⬚ (784)

Maximum marks: 6

6 Learning curves

For each of the following pairs of training and validation curves, select the appropriate reason for this behaviour.

Please match the values:

Plot 3 — Too small training data

Plot 1 — Underfitting

Plot 2 — Overfitting

Maximum marks: 9

7   Activation functions and neural networks

Select all statements which are true.

Select one or more alternatives

- ☑ Increasing the training set size generally does not hurt an algorithm's performance, and it may help significantly. ✓

- ☑ The tanh activation usually works better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data better for the next layer. ✓

- ☑ Increasing the size of a neural network generally does not hurt an algorithm's performance, and it may help significantly. ✓

- ☐ In logistic regression, the weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, logistic regression will fail to learn a useful decision boundary as it will fail to 'break symmetry'.

- ☐ The ReLU activation usually works better than sigmoid activation function for hidden units because the sigmoid activations are sparse.

Maximum marks: 9

8   RNNs

a. The network learns where to "pay attention" by learning the values e < t,t' >, which are computed using a small neural network: We can't replace s< t − 1 > with s as an input to this neural network. This is because s depends on α < t, t' > which in turn depends on e < t,t' > so at the time we need to evalute this network, we haven't computed s yet.

Select an alternative

- ◯ True ✓

- ◯ False

b. You have to fill a blank in a sentence of a long passage: "...............Sam liked teddy as a leader.............". There are many options for the same such as "bear" or "roosevelt" etc. You'd like to build a model to do it for you. Use of bi-RNN is not necessary for this purpose.

Select an alternative

- ◯ False ✓

- ◯ True

Maximum marks: 6

9 Clustering 1

Answer the following questions appropriately.

What is the minimum numbers of variables or features required to perform clustering? [____] (1)

K means and K-medioids are example of which type of clustering method? [____ ▾] (Partition, Hierarchical, Probabilistic, Spectral)

Which of the following are considered as unsupervised learning?
Select one or more alternatives

- ☐ Clustering   ✔
- ☐ Dimensionality reduction   ✔
- ☐ Decision trees
- ☐ Logistic regression
- ☐ None of these

Maximum marks: 12

10 Clustering 2

Consider the following statements pertaining to K-Means algorithm

a. Once an example has been assigned to a particular centroid, it will never be reassigned to another centroid.

Select an alternative

- ○ True
- ○ False   ✔

A good way to initialize K-means is to select K (distinct) examples from the training set and set them as cluster centroids.

Select an alternative

- ○ False
- ○ True   ✔

On every iteration of K-means, the cost funtion $J(c^{(1)},...,c^{(m)},\mu_1,...,\mu_k)$ (the distortion function) either stays the same or decreases; in particular, it should never increase.

Select an alternative

- ○ True   ✔
- ○ False

K-Means will always give the same results regardless of the initialization of the centroids.

Select an alternative

○ True

○ False   ✔

For some datasets, the "right" or "correct" value of K (the number of clusters) can be ambiguous, and hard even for a human expert looking carefully at the data to decide.

Select an alternative

○ True   ✔

○ False

If we are worried about K-means getting stuck in bad local optima, one way to reduce this problem is if we try using multiple random initializations.

Select an alternative

○ True   ✔

○ False

Maximum marks: 12

11   Provide your explanation for numerical computations here!

You can use this space to provide explanation for your answers!

Fill in your answer here

Maximum marks: 0