CHALMERS
EXAMINATION / TENTAMEN

Course code/kurskod	Course name/kursnamn			
DIT 821	Software Engineering for AI Systems			
Anonymous code Anonym kod	337	Examination date Tentamensdatum	Number of pages Antal blad	Grade Betyg
		2022.1.4	10	VG

* I confirm that I've no mobile or other similar electronic equipment available during the examination.
Jag intygar att jag inte har mobiltelefon eller annan liknande elektronisk utrustning tillgänglig under eximinationen.

Solved task Behandlade uppgifter No/nr		Points per task Poäng på uppgiften	Observe: Areas with bold contour are to completed by the teacher. Anmärkning: Rutor inom bred kontur ifylles av lärare.
1	x	3,0	
2	x	2,5	
3	x	3,5	
4	x	2,5	
5	x	2,8	
6	x	3,0	
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
Bonus poäng	x	1,0	

4

Q1

(a) α is the learning rate, it specifies how the gradient descent algorithm will change. when α is too large, it is hard to coverge and may cause vanish gradient problem, when α is too small, it takes more time to coverge. A reasonable α value will make the algorithm find the global minimum value instead of the local minimum value. 0.5

(b) The cost function : $\frac{1}{2m}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

(a) it is a polynomial equation with the max degree of 2.
in the graph, when w is 60, which is in middle of 20 and 100 the value of J(60) is the minumum, ∴ W=60 1

(b) because the initial weight starts from W=100
it reaches mininum in W=60 value of J(60)
∴ α can be in range of (0.01, 0.1) 1

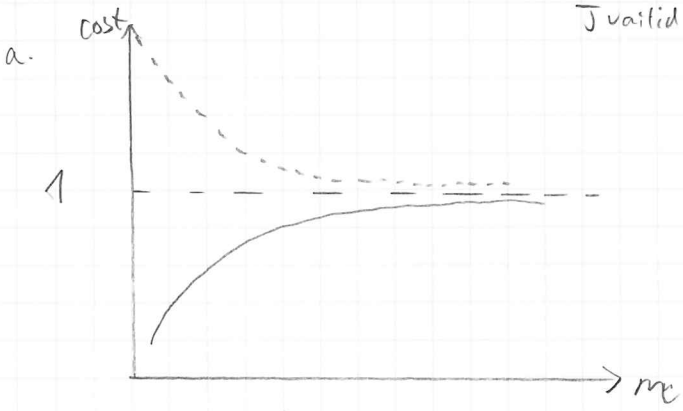
(c) at least 2 times, after the first iteration, make sure the initial weight is still on the right side of W=60 which means after the second or third iteration, the intial weight goes up to the left side of W=60, we will not miss the global mininum value in J(60). 0.5

(c) It can be done with Adam optimizer, which is called adaptive dynamic algorithm method.

(d) when the cost funtion J(w) has converged or the difference between J(w1) and J(w2) is very close to 0, then we can stop the iteration.

Q2

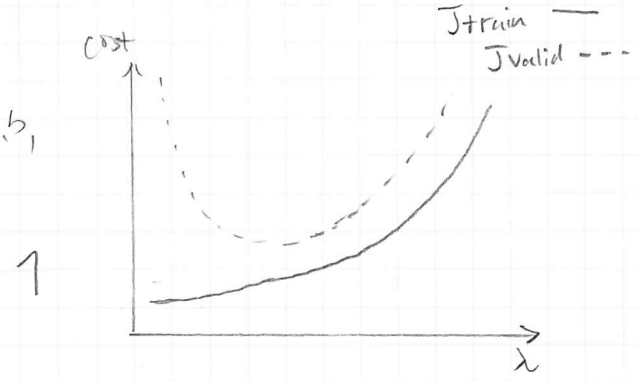
a.



J_{train} ——
J_{vailid} ---

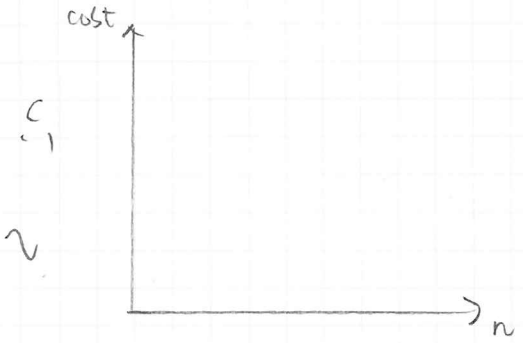
in training data, the value of $J(w)$ would be very small. then increase to a point, then close to one threshold value. In validation, it is opposite

b)



when λ is 0 in the beging, the J will be very small in the training data. while in the validation dataset it is very big.

c)



d) firstly: regularization parameter, then increase the feature number, finally increase dataset size ?

0.5

4

Q3.

(a)

The Prediction	Positive (The Actual result)	Negative	Total
Positive	TP = 8	FP = 2	= 10
Negative	FN = 4	TN = 86	= 90

1

(b) $\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} = \frac{92}{100} = 92\%$

$\text{Precision} = \frac{TP}{TP + FP} = \frac{8}{10} = 80\%$

$\text{recall} = \frac{TP}{TP + FN} = \frac{8}{12} = 66\%$

0.8

(c)

$$\begin{cases} TP + FP = 50 \\ \frac{TP}{TP+FP} = 80\% \end{cases} \Rightarrow \begin{cases} TP = 40 \\ FP = 10 \end{cases}$$

$\checkmark$ $\checkmark$

$$\begin{cases} FN + TN = 950 \\ \frac{TP}{TP+FN} = 0.66 \\ TP = 40 \end{cases} \Rightarrow \begin{cases} TP = 40 \\ FN \approx 26 \\ TN \approx 924 \end{cases}$$

? //

1.5

The Prediction	Positive (Actual Result)	Negative
Positive	TP = 40	FP = 10
Negative	FN = 26	TN = 924
Total:	= 66	= 934

Real number of COVID positive : 66.

Real number of COVID Negative: 934

4

Q4.1: Convolutional layer: Extract all features of input data

Max pooling layer: cut the channel into smaller channels, get the new features with smaller dimension. It aims for dimension reduction.

1 ReLU activation layer: mapping the input end of neutron to the output. It adds non-linear combination, otherwise all the output is linear combination of input.

Fully-connected layer: combine all partial smaller features together with dropout, it can avoid overfitting issue.

Q4.2 $a^{(1)}$ is a vector with dimension 3

Q(a) $\theta^{(1)} = \begin{bmatrix} -0.1 & 0.2 \\ 0.3 & -0.4 \\ 0.5 & -0.6 \end{bmatrix}$ $\therefore \theta^{(1)}$ is a 3×2 matrix

$a^{(2)}$ is a vector with dimension 3

$\theta^{(2)} = \begin{bmatrix} -0.7 \\ -0.9 \\ 1.1 \end{bmatrix}$ $\therefore \theta^{(2)}$ is a 3×1 matrix. 1.5

$a^{(3)}$ is a vector with dimension 1

Q(b): $\theta^{(1)} = \begin{bmatrix} -0.1 & 0.2 \\ 0.3 & -0.4 \\ 0.5 & -0.6 \end{bmatrix}$, $\theta^{(2)} = \begin{bmatrix} -0.7 \\ -0.9 \\ 1.1 \end{bmatrix}$

Q(c): sigmoid function is $\frac{1}{1+e^{-x}}$

$a_{21}^{(2)} = sig(-0.1 \times 1 + 0 + 0) = sig(-0.1) = \frac{1}{1+e^{0.1}}$

$a_{22}^{(2)} = sig(0.2 + 0 + 0) = sig(0.2) = \frac{1}{1+e^{-0.2}}$

$h(x) = a^3 = sig(-0.7 - 0.9 \times \frac{1}{1+e^{0.1}} + 1.1 \times \frac{1}{1+e^{-0.2}})$

$= sig(\quad)$ ← Note: use calculator to get the results here

$= \_\_$

Q5

(a) 1. prepare the data, 2 process the data, it includes

2.1 clean the data, which means finding the missing values and replace those with either zero or average value of its column.

2.2 data labelling: assign the data into different categories.

It can be performed by data annotation, tagging, classification and transcription.

2.3 data splitting: split the data into train, validation and test data.

3. Feature Engineering.

4 Train and evaluate model.

5 Deploy the model: deploy the model in the server for end-user

6 Monitor the model: check the model regularily of its accuracy and performance.

(b) Traditional Software: it is based on explaination, which means developpers code and make rules, then get the expected results.

ML-based software: it is based on prediction, which means it handles with a lot of data, developpers do not know the rules until they get the expected results.

0.3

4

Q5

(c) Feature engineering: the process of transforming the raw input data into the most informative features in the machine learning algorithms.

When dealing with numeric data, it can be performed by normalization, Standalization, log transformation etc.

when dealingwith categorical data, it can be done using one-hot encoding and hash encoding

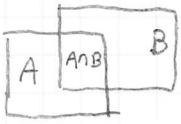
when dealing with textual data, it can be done using Bag of words, it is a simplified expression model in natural language and information retrieval. In this model, textual data such as sentenses can be represented in a bag containing these words. Bag of words doesn't consider the grammar or the order of words

0.5 Finally, principal Component Analysis (PCA) can be done to reduce dimention of data.

(d) Inter-annotator agreement measures how well two or more annotators can make the same annotation decision on one category of object.

0.5 It can be measured by calculating Intersection of Union (IoU)

For example: in the box, Interesection means $A \cap B$.



while union means $A \cup B$

By calculating the value of $\frac{A \cap B}{A \cup B}$ can measure, how well the inter-annotator agreement

Q5,

a) The first deployment pattern is expose REST API.

pros: it is based on HTTP protocol, it is lightweight, it can be easily tested on browser it can be easily performed among different enterprises, because it follows HTTP protocol.

cons: Since it is based on HTTP protocol, session can not be maintained, and all the calls are asychronous.

The second pattern is embedded pattern

pros: the embedded pattern means it integrates feature selection process with model training process. The model training process is automatically performed when doing the feature selection process.

cons. it takes more time and time-consuming compared to the expose REST API

Q6:

(a) Logistic Regression: measures the relationship between one categorical dependent variable y and one or more independent variables x by estimating the probability of one thing happens using logistic function. The cost function is $\begin{cases} -\log(h_w(x)), & y=1 \\ -\log(1-h_w(x)), & y=0 \end{cases}$.

Linear Regression: is a regression analysis method between one dependent variable y that is continuous and one or more independent variables x. The hypothesis function is $h_w(x) = b_0 + b_1 x$ and cost function is $\frac{1}{2m}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

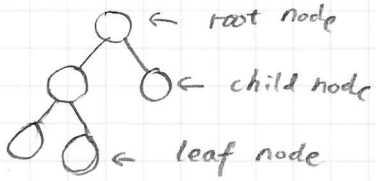
(b) feature scaling: make sure all input data are in a similar scale either between (0, 1) or (-1, 1). It can be performed in

1° Mean normalization: $x = \frac{x - \bar{x}}{x_{max} - x_{min}}$, $\bar{x}$ is mean value, x_{max} is max value, x_{min} is min value.

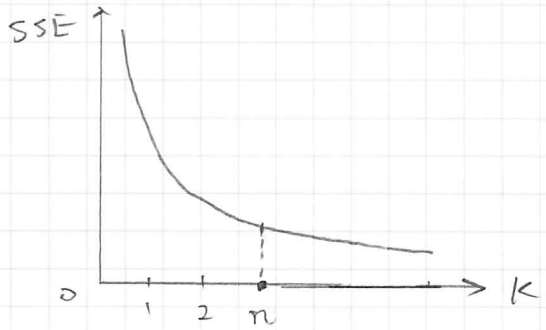
2° Z score normalization: $x = \frac{x - \bar{x}}{std}$, $\bar{x}$ is mean value, std is standard deviation value

Q 6

(c) Decision tree is to solve regression or classification tasks.
It starts from the root node, test certain features of samples, assign the samples to their child node according to the test results. Each child node is correspond to different values.
In this way, all samples can be tested and distributed recursively until they reach the leaf node.



(d) It can be decided by using the elbow method.
in the graph, the x axis represents the number of K.
the y axis represents the SSE = sum of squared error.



when the SSE value decreases slowly at one point, the corresponding K number is the optimal number of clusters.

Q 6

e, By using CNN, it follows more layers than Fully connected Neural Network. CNN model contains (1) convolutional layer which is to extract all features (2) ReLU layer: adding non-linear characteries and avoid all outputs are all linear combination of inputs. 3, pooling layer: cut the image into several small regions, take the average value or maximum value in each region. It helps to reduce dimension. 4, flatten layer: flatten the multi-dimensional tensors into one single dimension, which increase efficiency (5) Fully connected layer: combine all partial features into a global feature. Thus, CNN reduces dimensions and more efficient.

f, (1). human, social and environment. The AI system should help human being, society and environment and make all well-being life better.

2, human-centered value: The AI system should base on development of human.

3, data protection and privacy. The AI system should keep users privacy and protect personal data.

4, Fairness. The AI system should be fair and open to society.

5, Transparency. The AI system should be transparent to society.

6, Contestability The AI system should allow competition among different companies.

7, Accountability. The AI system should be accountable to their actions.