CHALMERS
EXAMINATION / TENTAMEN

Course code/kurskod	Course name/kursnamn			
Dit 821	Software engeneering for AI			
Anonymous code Anonym kod		Examination date Tentamensdatum	Number of pages Antal blad	Grade Betyg
369		2022-10-26	11	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	4.5	
2	X	4	
3	X	5	
4	X	~~3,5~~ 4	
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
Bonus poäng	1		
Total examination points Summa poäng		~~13~~ 16,5	

3

a) The decision boundary equation is:-

$$Y = g(6 + (-5X) + 1 \times X^2$$

$$Y = g(6 - 5X + X^2)$$

4p

b) to solve this question we will choose ~~x~~ the max x that output the maximum propability for $y_1, y_2 y_3$ the labels for the three new labels is

$$(x_1, y_1), (x_2, y_3), (x_3, y_1)$$
$$= (x_1, 0), (x_2, 2), (x_3, 1)$$

4p

c) The certain parameters is the "K" number of clusters it group and classify the data to this ~~not~~ specified number of classes or cluster.

1p

d) ~~K-mean adaptation Continuation~~

~~Repeat~~

~~All~~

The higher K number means we add more clusters.

3

d) Continuation of k-means algorithm :—

Repeat {

for $i=1$ to m

$c^{(i)} :=$ ~~average of points closest to x^i~~
Compute distance and assign the closest x^i.

for $k=1$ to k

μ_k = assign average (mean) of points of all x^i closest to this centroid.

}

7p

e) Sum of the distances between points and their centroids squared divided by the sample size or the number of training set.

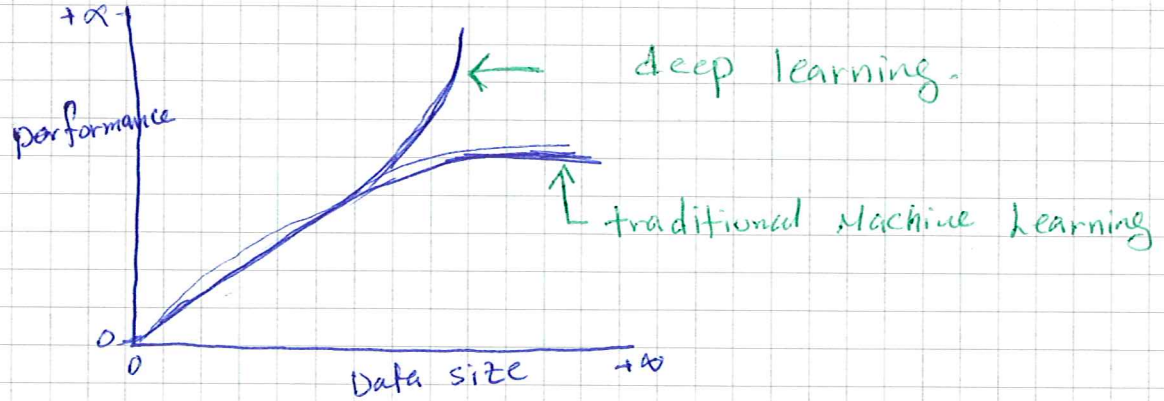
The goal is to find the minumim distances between cluster centroids and points which result in better clustring.

1p

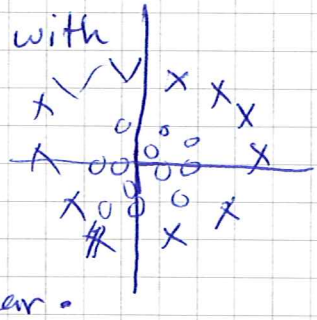
3

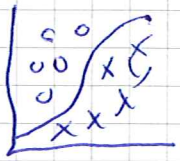
a) 1) in tradational ML we must perform feature engeneering and feature extraction while in Nural Network is not neccessary. Can do automatically through the NN and feature selection.

2) «look at the graph below» in traditional ML perform good when the data size is small but then it will have performance parlled to X-axis in Deep learning The performance increase as the Data increase



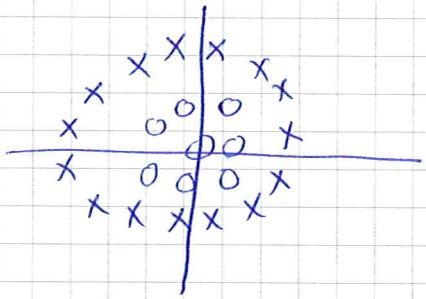
b) Deep learning deals with very complex proplems that traditional can't as in traditional «Logistic Regression deal with one dicison boundry but it can not deal with this proplems → where the decision boundry is non linear.





CHALMERS	Anonymous code	Points for question (to be filled in by teacher)	Consecutive page no. Löpande sid nr 7
Dit 821	Anonym kod 369	Poäng på uppgiften (ifylles av lärare)	Question no. Uppgift nr 8

b) none Linearity (1) allow neural networ to deal with very complex data use the graph" such as:-

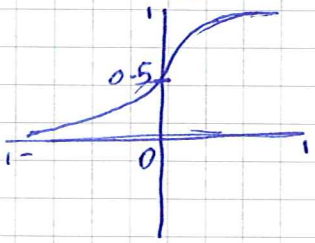


it uses activation function to tackle this proplem where there is no one decision boundry.

*) function is that is used is Sigmoid Activation function

$$g(z) = \frac{1}{1+e^{-g(z)}}$$

$$z = h(x)$$



c) it is ~~m~~ smaller (0.5) because we want to detect only the Important feature and to ~~increase~~ decrease (not sure) model complixity.

and to detekt certain features that exists in certain location in the image.

CHALMERS	Anonymous code	Points for question (to be filled in by teacher)	Consecutive page no. Löpande sid nr 8
Dit 821	Anonym kod 369	Poäng på uppgiften (ifylles av lärare)	Question no. Uppgift nr 3

1

d) 1) deal with spatial variance to reduce distortion in image.

2) reduce model complexity by reducing number of parameters which increase model performance.

3) Capture the most important feature and maximize it to better detection in the model and eliminate non important informations.

1.5

e) Vanishing gradient is when the gradient dis~~-~~ dis appear when we move from one layer in Nural Network to other 

RNN solve it by using sigmoid gate and point wise multiplcation. when sigmoid is equal to 1 it mean allow all the information to pass to nex and when is 0 it means block every thing from flowing to nex by layer.

LSTM

3

a)

a) wrong measures of the data could be that can appear during data collection. and systems breaks as well.

2) Social biased data which lead to biased model to for example gender or race.

1

b) it is ~~Ett~~ important to Categorical data to Numrical Data for the model to be aprable to process. ~~one hot~~ one example is the house data

which one?

Size	Pool Area	Parking area	price
10	0	0	1000
5	5	5	2000
0	0	0	5000
8	8	8	10000

⇐

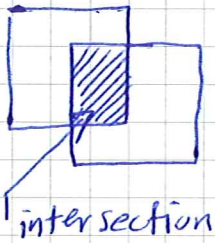
Size	Pool Area	Parking Area	price
	NA	NA	
	medium	medium	
	NA	NA	
	large	large	

on-hot encoding is used to ~~mp~~ map Categorical data into Numric data either by assigning binary values 0 to not available or no and 1 to aviable or yes or by specifing rang for example if X~~#~~ = 0 if pool area is NA and X = 5-7 if pool Area is medium. ✓ ~~5~~

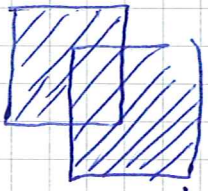
std 1

C) the fomla is to test Inter-Agreement Anotator between tow or more annotators simply by calculating the Area of two intersections of tow boxes and divede it by are of union of two boxes:-

*) The boxes is the specific Area of the Image determined by Annotator and labeled like Car Driving way, house, etc



intersection of two boxes



Union of 2 boxes

the output will be score between 0 and 1 0 for no matching , 1 represent total Agreement.

0.4 represent poor match, 0.7 =good match and 0.9 Excellent Agreement. 1

CHALMERS | Anonymous code / Anonym kod: Dit521 – 369 | Points for question (to be filled in by teacher) / Poäng på uppgiften (ifylles av lärare): 1 | Consecutive page no. / Löpande sid nr: 11 | Question no. / Uppgift nr: 4

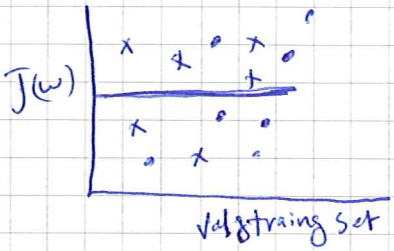
d) it's not always apicable ~~As if Cow~~ in most cases we could generate metrics from the requirment for example the house prediction the price should fall between minum value and maximum value for price.

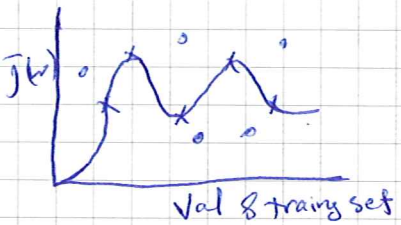
but in some cases we can not specify metrics & for evalution from the requirement like algorithm that predict human behaviour. also to ~~meauser~~ meauser how bias the model is.

V

3

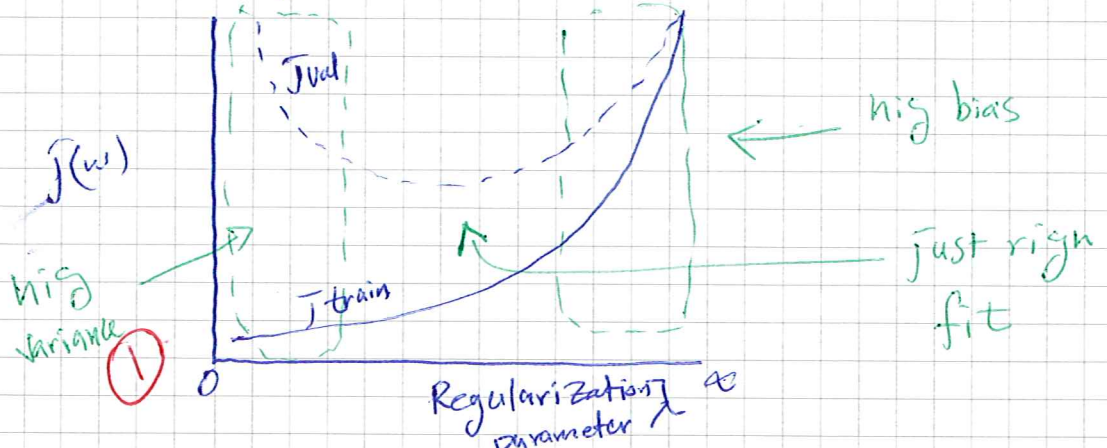
a) under fitting is when the model is too simple it can't generalize well in the trainging data and val.
over fitting is when the model is too complex and it fits perfectly the traing data but it produce hig error in validation set.





x traing set
o validation set

b) Regualrization help avoid overfitting by minimizing the affect of the high order features x^4, x^5 ...
So the model will not just fitt training data but it well geralize well in the validation set.



$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h(x)-y)^2 + \sum w_i^2\right]$$

Regularization term · λ

it well penalize the model with ~~small~~ value related to model complexity. the more complex model the bigger lamda is needed to better fitting the data

what? how

3

c) - gradient decent equation

$$J(w) = w - \alpha/m \sum_{i=1}^{m} (h(x_i) - y_i) \cdot x^i$$

- Normal equation

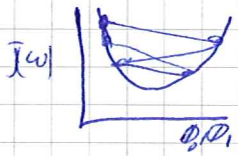
$$w = (X^T X)^{-1} X^T y$$

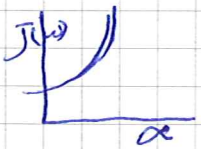
- in gradient decent we must specify number of iterations while in normal equation not required.
- in gradient decent we must choose alpha while in normal equation not required
- gradient decent work well with big Data while normal equation does not wor will with big Data $(X^T X)$ is computationally expensive and the inverse in $(X^T X)^{-1}$ is more expensive and some time is not applicable.

d) I would prefer gradient decent as it's big Data and the normal equation is is not work well in big Data $m = 50 \quad n = 200,000$
- if we look at the equation $(X^T X)^{-1}$, the X transpose is computationally expensive and the inverse may not be applicable for big Data.

CHALMERS	Anonymous code	Points for question (to be filled in by teacher)	Consecutive page no. Löpande sid nr 3
DIT 821	Anonym kod 369	Poäng på uppgiften (ifylles av lärare)	Question no. Uppgift nr 1

e) a) false if we tak specify α is large value the gradient decent may not converge and we might over shoot the optime points that result in in increasing of the cost.





α is large

b) false the

the reason is that each one has diffrant affect of the cost function wo may ~~con~~ decrease faster or slower than w1, since we might have local optimal.