FINAL EXAM

CT522: Topics in Deep Learning

FINAL EXAM (11th MAY 2022)
DA-IICT, Gandhinagar
Time Duration: **2 hours**
Total Marks: **35**

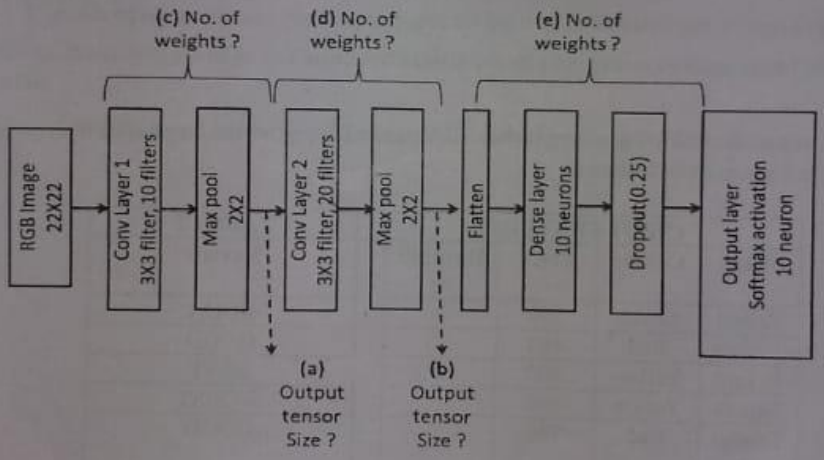
---

1. You need to perform a multivariate non-linear regression. There are 100 input features, and you need to predict three output variables. For that, consider the following ANN architecture. **(2 points)**

   Input layer features and neurons: 100
   Two hidden layers, each one has 10 neurons
   Output layer with 3 neurons

   (a) How many numbers of parameters are in this network?
   (b) What will be your choice of activation function in the hidden layers and the output layer?
   (c) What will be your choice of loss function?

2. For a given convolution neural network architecture



   Answer the **a, b, c, d and e. (3 points)**

3. Refer confusion matrix given below. Compute total accuracy, and precision & recall for the ***Diseased*** class **(3 points)**

		Predicted	
	classes	*Diseased*	*Normal*
Actual	*Diseased*	200	100
	Normal	50	650

4. True or False **(4 points)**

(a) used for regression problem-solving in a deep learning-based approach. F

(b) Data normalization before model training help to reduce the underflow or overflow error in deep learning implementation.The batch normalization layer or technique aims to normalize the input features scale. T (a)

(c) Node Dropout technique aims to reduce the over-fitting of the deep learning model. T

(d) Gated Recurrent Unit (GRU) cell have a lower number of parameters than LSTM cell. F

(e) A bidirectional LSTM recurrent neural network needs a fixed input sequence length. F

(f) The sum total of the output vector from the softmax active function is between 0 and 1. T

(g) Logistic regression can be used for non-linear binary classification problem T

(h) Mean square error (MSE) loss is mainly.

5. Look at the following example dataset. You need to predict the target (Review) variable from given input features.

INPUT FEATURES				TARGET
Shape	**Colour**	**Price**	**Durability**	**Review**
Round	Green	700	4	POOR
Square	Red	460	3.5	GOOD
Round	Yellow	800	2	BEST
Square	Yellow	450	3.7	WORST
Triangle	Red	796	5	GOOD

Shape : {Round, Square, Triangle}
Colour : {Green, Red, Yellow, Pink, Blue}
Price : {Positive Integer}
Durability : {real number}
Review : {WORST, POOR, GOOD, BEST}

(a) How do you encode each input and target feature? **(3 points)**
(b) Design the ANN architecture to solve the above problem. Mentions input, hidden and output layer configuration, activation functions and loss function **(2 points)**

6. What is an output range of derivative of linear, Relu, Tanh and sigmoid activation functions? **(2 points)**

7. Explain the binary cross-entropy loss. When will you use it in an artificial neural network? **(2 points)**

8. Describe the unique characteristics of the popular CNN architectures: **GoogleNet, ResNet**, and **DenseNet**. Please describe it in two to three sentences for each one. **(3 points)**

9. Why does LSTM usually perform better than the standard RNN architecture? When will you use the LSTM? Explain the parameter sharing concept in the LSTM network. **(3 points)**

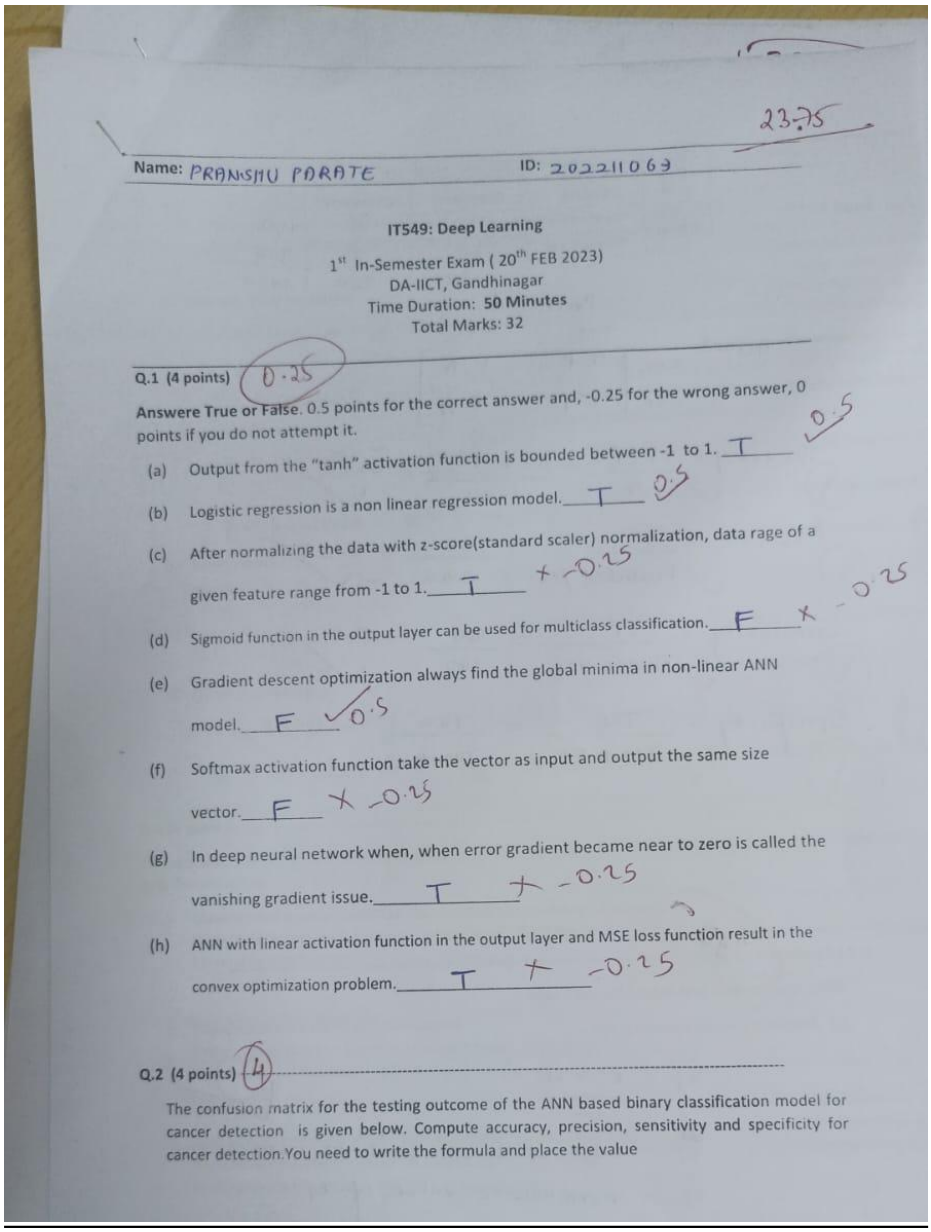
10. Describe the generative adversarial network (GAN) concept. How is the GAN model trained? Mention 2 to 3 applications of the GAN based architectures. **(3 points)**.

11. What do you mean by hyper-parameters in ANN or Deep learning? Give two examples of it. **(2 points)**

12. Brief on **any one topic of your choice** mentioned below with respect to deep learning **(3 points)**

Data normalization **or** *Model regularization* **or** *Stochastic gradient descent technique*

1st in-sem



23.75

Name: PRANSHU PORATE ID: 202211069

IT549: Deep Learning

1st In-Semester Exam (20th FEB 2023)
DA-IICT, Gandhinagar
Time Duration: **50 Minutes**
Total Marks: 32

Q.1 (4 points) 0.25

Answere True or False. 0.5 points for the correct answer and, -0.25 for the wrong answer, 0 points if you do not attempt it.

(a) Output from the "tanh" activation function is bounded between -1 to 1. T ✓ 0.5

(b) Logistic regression is a non linear regression model. T 0.5

(c) After normalizing the data with z-score(standard scaler) normalization, data rage of a given feature range from -1 to 1. T × -0.25

(d) Sigmoid function in the output layer can be used for multiclass classification. F × -0.25

(e) Gradient descent optimization always find the global minima in non-linear ANN model. F ✓ 0.5

(f) Softmax activation function take the vector as input and output the same size vector. F × -0.25

(g) In deep neural network when, when error gradient became near to zero is called the vanishing gradient issue. T × -0.25

(h) ANN with linear activation function in the output layer and MSE loss function result in the convex optimization problem. T × -0.25

Q.2 (4 points) 4

The confusion matrix for the testing outcome of the ANN based binary classification model for cancer detection is given below. Compute accuracy, precision, sensitivity and specificity for cancer detection.You need to write the formula and place the value

23.75

		Prediction		
	Classes	Cancer(+v)	Normal(-ve)	
Actual	Cancer(+v)	80	20	100
	Normal(-ve)	200	700	900
		280	720	1000

Actual \ Predicted	+ve	-ve
+ve	TP	FN
-ve	FP	TN

(1) Accuracy = $\frac{TP+TN}{N} = \frac{80+700}{1000} = 0.78$

(1) Precision = $\frac{TP}{\text{Predicted yes}} = \frac{80}{280} = 0.28$

(1) Sensitivity = $\frac{TP}{\text{Actual yes}} = \frac{80}{100} = 0.8$

(1) Specificity = $\frac{TN}{TN+FP} = \frac{700}{700+200} = 0.77$

Q.3 (6 points) 4.50

Write the formulas and denote the each term.

(a) Z-score(standard nomralization)

$$\hat{x}_i = \frac{x_i - \mu}{\sigma}$$ (1)

x_i - value which we want to normalized

$\hat{x}_i$ - normalized value of x_i

μ - mean

σ - Standard Deviation

23.75

Name:

(b) Relu function

$$g(z) = \begin{cases} z, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

1 $g(z)$ - function, $z - [0, \infty)$

(c) Multiple linear regession

$$y = w_0 + \sum_{m=1}^{N} w_m x_m$$ 1

y - dependent variable, x - independent variable, w - weight

(d) Logistic regression

$$y = \sigma(z) = \frac{1}{1+e^{-z}}$$ 1/2

y - function $z - (-\infty, \infty)$

(e) Mean square error

$$MSE = \frac{1}{n}(y_n - \hat{y}_n)^2$$ 1/2

y_n - Actual Value $\hat{y}_n$ - Predicted Value

(f) Binary cross entropy

$$BCE = -(t_k \log y_k + (1-t_k)\log(1-y_k))$$ 1

t_k - Target Value y_k - Predicted Value

Q.4 (6 points) 6

Fill the blank; MLP network with following scenario (you can specify some pre-processing of output data if you think so...)

a. Classification problem with ten classes; One hot encoding of output;
Preferable activation function at output layer: Softmax
Preferable cost function: Cross Entropy 2

b. Classification problem with two classes
Preferable Activation function at output layer: Sigmoid
Preferable cost function: Binary cross entropy 2

c. Regression problem with one output variable, output ranges unbounded.
Preferable Activation function at output layer: Linear
Preferable cost function: Mean Square Error 2

23.75

Name: T.

Q.5 (5 points) (3)

One university has an extensive dataset of students' information. Example observations and data range of features are provided below. University wants to develop **a non-linear ANN-based model** to predict student programming skill from input features **(5 points)**

.75

Input Features						TARGET
AGE	GENDER	Math Skill	GRADES	Problem-solving skill		Programing skill
20	M	0.7	9	0.7		GOOD
22	F	0.4	5	0.6		AVERAGE
23	M	0.8	6	0.5		POOR
28	M	0.2	6	0.92		EXCELLENT
25	F	0.6	7	0.96		VERY POOR

AGE : {18-30, Positive Integer}
GENDER: {M, F}
Math Skill: {real value between 0 and 1}
Programming Skill: {VERY POOR, POOR, AVERAGE, GOOD, and EXCELLENT}
Problem solving skill: {real value between 0 and 1}
GRADES : {0 to 10, positive integer}

(a) How many neurons in the input layer will you choose? 5 (1)

(b) There is only one hidden layer. Which activation function will you use? ReLU (1/2)

(c) Which activation function will you use in the output layer? Softmax

(d) How many neurons will you choose in the output layer? 5 (1.5)

(e) Which loss function will you use? Cross Entropy

Q.6 (4 points) (3)

Fill in the gap.

(a) A logistic regression model with 13-dimensional input features has 13 number of learnable (1/2)
parameters and multiple linear regression has 14 number of learnable parameters.

(b) How many parameter in ANN modlel that contains 10 neuron in input layers, 10 neuron in hidden layer and one neuron in output layer. 110 (10 X 10 + 10 X 1) (1/2)

23.75

ne: Jagrut

(c) Gradient value of ReLu activation function for positive input is 1 and for negative input value is 0. (1)

(d) Maximum gradient value of tanh function is 1 and sigmoid function is 0.25. (1)

Q.7 (3 points) 3

Given a following matrix; compute the performance measures e.g. **precision, recall, and specificity** for '***Neutral***' class from following confusion matrix.

		Predicted			
	classes	***Like***	***Neutral***	***Dislike***	total
Actual	***Like***	210	25	25	250
	Neutral	50	400	50	500
	Dislike	40	50	150	250
	total	300	475	225	1000

	+ve	+ve	-ve	-ve
ve	TN	FP	TN	TN
ve	FN	TP	FN	FN
ve	TN	FP	TN	TN
ve	TN	FP	TN	TN

Neutral : N

$TP_N = 400$

$FP_N = 25 + 50 = 75$

$FN_N = 50 + 50 = 100$

$TN_N = 210 + 25 + 40 + 150 = 425$

$$\text{Precision} = \frac{TP}{\text{Predicted yes}} = \frac{400}{475} = 0.84 \quad (1)$$

$$\text{Recall} = \frac{TP}{\text{Actual yes}} = \frac{400}{500} = 0.8 \quad (1)$$

$$\text{Specifiuty} = \frac{TN}{TN+FP} = \frac{425}{425+75} = 0.85 \quad (1)$$

1st insem (2)

IT641: Deep Learning

1st In-Semester Exam (14th SEPTEMBER **2023**)
DA-IICT, Gandhinagar
Time Duration: **60 Minutes**
Marks: **24 marks**

---

(1) True or False (**7 points**)

(a) Exploding gradients issue is likely to happen in the deep neural network if the sigmoid activation function is used in the hidden layer. (F)
(b) Logistic regression approximates linear decision boundary for binary classification problems. (T)
(c) Use of the Node dropout technique may increase the training error of the ANN model, compared to if not using it. (T)
(d) Multiple linear regression models have a close-form solution to compute weight value from training data. (T)
(e) L2 regularization reduces the effective capacity of the non-linear ANN model. (T)
(f) ANN with linear activation function in the output layer and MSE loss function result in the convex optimization problem. (F)
(g) The sum total of the output of the Softmax function is one. (T)
(h) The maximum gradient value of the "linear activation" function is 1, and the minimum gradient value is -1 (F)
(i) Using L1 regularization reduces the number of parameters in the ANN model. (F)
(j) We cannot take more neurons in the hidden layer than neurons in the input layer. (F)
(k) The sigmoid function in the output layer can be used for multiclass classification. (T)
(l) The binary cross-entropy function is used for classification applications but not for regression applications. (F)
(m) Sigmoid function output is bounded between -1 to 1. (F)
(n) Mean Absolute Error (MAE) may be used as a loss function. (T)

(2) The confusion matrix for the testing outcome of the logistic regression model for fraud detection is given below. Compute true positive, false positive and false negative, precision, sensitivity and specificity. Note that, fraud(+ve) is the class of interest (**3 points**)

		Prediction	
	Classes	***fraud(+ve)***	***Normal(-ve)***
Actual	***fraud(+ve)***	300	100
	Normal(-ve)	100	10000

TP=300 , FP = 100, FN=100
Precisioin=(300/[300+100])
Sensitivity = (300/[300+100])
Specificity = (10000/[10000+100])

14/09/23

Name: K Sathu Srinatha

Roll No: 202218021

MscDs Deep Learning Mid Sem 1

20/21

① ⓐ T ✗ Q1 4.5
ⓑ T ✓ Q2 3
ⓒ T ✓ Q3 2
4.5 ⓓ T ✓ Q4 3
ⓔ F ✗ Q5 3
ⓕ F ✓ Q6 3
ⓖ T ✓ Q7 2
ⓗ T ✗ 20.5
ⓘ T ✗
ⓙ F ✓
ⓚ T ✓
ⓛ T ✗
ⓜ F ✓
ⓝ T ✓

②

C		Pred	
		Fra(+)	Nvs(-)
Actual	Fra(+)	300 (TP)	100 (FN)
	Nvs(-)	100 (FP)	10000 (TN)

3 TP = 300, FP = 100, FN = 100,

Precision = $\frac{TP}{TP+FP} = \frac{300}{400} = 0.75$ ✓

Sensitivity = $\frac{TP}{TP+FN} = \frac{300}{300+100} = 0.75$ ✓

Specificity = $\frac{TN}{TN+FP} = \frac{10000}{10000+100} = \frac{10000}{10100} =$

Deep Learning Atte | IR_A5 - Colaborator | IT641 2023 DEEP L | Lab Assignment 1 | Assignment 6 - Go

bjector=1

SEM_EXAM_solution.pdf

Open with

below. Compute true positive, false positive and false negative, precision, sensitivity and specificity. Note that, fraud(+ve) is the class of interest (3

		Prediction	
	Classes	*fraud(+ve)*	*Normal(-ve)*
Actual	*fraud(+ve)*	300	100
	Normal(-ve)	100	10000

TP=300 , FP = 100, FN=100
Precisioin=(300/[300+100])
Sensitivity = (300/[300+100])
Specificity = (10000/[10000+100])

(3) The confusion matrix for the testing outcome of the multiclass classification with three classes {C1, C2, C3} is given below.

Compute the ***precision*** and **recall** for class ***C2***. **(2 points)**

		Prediction		
	Classes	*C1*	*C2*	*C3*
Actual	*C1*	1000	100	500
	C2	2000	200	400
	C3	3000	300	300

Precision = (200/[100+200+300])
Recall = (200/[2000+200+400])

(4) You need to train a non-linear deep ANN model (with. 50 hidden layers) for data with 100 classes and 5000 features. **(3 points)**

(a) How many neurons in the input layer will you choose? 5000
(b) Which activation function will you use in hidden layers? Tanh or relu or sigmoid
(c) Which activation function will you use in the output layer? Softmax or sigmoid
(d) How many neurons will you choose in the output layer? 100
(e) Which loss function will you use? CE or BCE

(5) One university has an extensive dataset of students' information. Example observations and data range of features are provided below. University wants to develop **a non-linear ANN-based model** to predict programming skill from input features. (I advise you to attempt this question; you may answer the part of the question) **(3 points)**

Page 2 / 3

Input Features | TARGET

③

		Pred C1	Pred C2	Pred C3
Act	C1	1000 (TN)	100 (FP)	500 (TN)
	C2	2000 (FN)	200 (TP)	400 (FN)
	C3	3000 (TN)	300 (FP)	300 (TN)

2

$$\text{Precision}^{(C2)} = \frac{TP}{TP+FP} = \frac{200}{200+100+300} = \frac{200}{600} = 0.33$$

$$\text{Recall}^{(C2)} = \frac{TP}{TP+FN} = \frac{200}{200+2000+400} = \frac{200}{2600} = \frac{1}{13} = 0.0$$

④ @ 5000
ⓑ Relu
ⓒ Softmax
3 ⓓ 100
ⓔ Cross Entropy

⑤ We could solve it normally using Softmax, Cross Entropy but to improve efficiency we will convert Output into numeric by label encoding, to do regression.

ⓐ 5
3 ⓑ linear
ⓒ linear (Relu can also be used)
ⓓ 1
ⓔ Mean Square Error

⑥

ⓐ 11
ⓑ 1, 0
3 ⓒ Softmax, Cross entropy

(4) You need to train a non-linear deep ANN model (with. 50 hidden layers) for data with 100 classes and 5000 features. (**3 points**)

(a) How many neurons in the input layer will you choose? 5000
(b) Which activation function will you use in hidden layers? Tanh or relu or sigmoid
(c) Which activation function will you use in the output layer? Softmax or sigmoid
(d) How many neurons will you choose in the output layer? 100
(e) Which loss function will you use? CE or BCE

(5) One university has an extensive dataset of students' information. Example observations and data range of features are provided below. University wants to develop **a non-linear ANN-based model** to predict programming skill from input features. (I advise you to attempt this question; you may answer the part of the question) (**3 points**)

	Input Features					TARGET
AGE	GENDER	Math Skill	GRADES	Problem-solving skill		Programming skill
20	M	0.7	9	0.7		GOOD
22	F	0.4	5	0.6		AVERAGE
23	M	0.8	6	0.5		POOR
28	M	0.2	6	0.92		EXCELLENT
25	F	0.6	7	0.96		VERY POOR

AGE : {18-30, Positive Integer}
GENDER: {M, F}
Math Skill: {real value between 0 and 1}
Programming Skill: {VERY POOR, POOR, AVERAGE, GOOD, and EXCELLENT}
Problem-solving skill: {real value between 0 and 1}
GRADES : {0 to 10, positive integer}

(f) How many neurons in the input layer will you choose? 5 or 6
(g) There is only one hidden layer. Which activation function will you use? Relu / vanishing gradient
(h) Which activation function will you use in the output layer? **LINEAR or SOFTMAX or SIGMOID**
(i) How many neurons will you choose in the output layer? **ONE** or **FIVE**
(j) Which loss function will you use? **MSE** or **CE** or **BCE**

(6) Fill in the blanks (**3 points**)

(a) A multiple linear regression model with 10-dimensional input features has ___10 or 11___ number of learnable parameters.
(b) The gradient value of ReLu activation function for positive input is ___1___, and for negative input value is__0____.
(c) Modelling multiple classification problems with the ANN model, your choice of activation function is ____Softmax or Sigmoid____ in the output layer, and the loss function is ____CE or BCE____.

7

a (10×10+10) + (10×1+1) = 110+11 = 121

b We use nodes equal on hidden layer so we have 8 learnable parameters

c Same as first. (10×10+10) + (10×1+1) = 110+11 = 121

hidden → input → output

7+8+1 = 17

_solution.pdf

Open with

22	F	0.4	5	0.6		AVERAGE
23	M	0.8		0.5		POOR
28	M	0.2		0.92		EXCELLENT
25	F	0.6	7	0.96		VERY POOR

AGE : {18-30, Positive Integer}
GENDER: {M, F}
Math Skill: {real value between 0 and 1}
Programming Skill: {VERY POOR, POOR, AVERAGE, GOOD, and EXCELLENT}
Problem-solving skill: {real value between 0 and 1}
GRADES : {0 to 10, positive integer}

(f) How many neurons in the input layer will you choose? 5 or 6
(g) There is only one hidden layer. Which activation function will you use? Relu / vanishing gradient
(h) Which activation function will you use in the output layer? **LINEAR** or SOFTMAX or SIGMOID
(i) How many neurons will you choose in the output layer? **ONE** or FIVE
(j) Which loss function will you use? **MSE** or **CE** or BCE

(6) Fill in the blanks (**3 points**)
(a) A multiple linear regression model with 10-dimensional input features has ___10 or 11___ number of learnable parameters.
(b) The gradient value of ReLu activation function for positive input is ___1___, and for negative input value is__0____.
(c) Modelling multiple classification problems with the ANN model, your choice of activation function is ____Softmax or Sigmoid____ in the output layer, and the loss function is ____CE or BCE____.

(7) Consider the ANN model with ten neurons in the input layer, ten in the hidden layer, and one in the output layer. (**3 points**)
(a) How many learnable parameters are in the above ANN model? 121
(b) How many learnable parameters are in the above ANN model if we use the Nodedrop out regularization with p=0.2 on the hidden layer? 121
(c) How many learnable parameters are in the above ANN model if we use the tanh activation function instead of ReLu in a hidden layer? 121