UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL
& COMPUTER ENGINEERING

Νευρο-Ασαφής Υπολογιστική (Neuro-Fuzzy Computing)
Fall Semester 2025-2026
Δημήτριος Κατσαρός (prof. Dimitrios Katsaros)

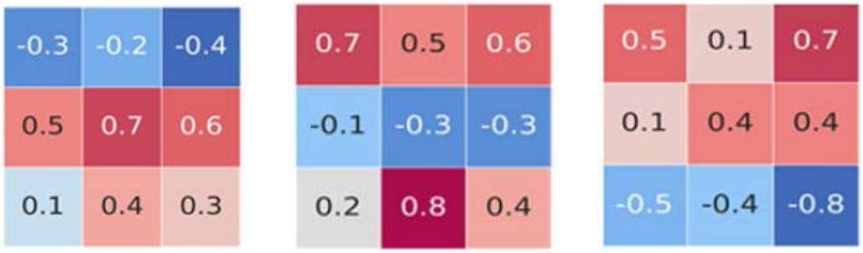
Problem-set 3rd : **Teamwork** (2-persons)

Announcement: Thursday, January 09, 2025
Delivery deadline: Thursday, February 12, 2026

SECTION 1: Model Reduction on CNNs

Problem-01

You are given the following three kernels (filters) in a convolutional layer of a CNN.



Suppose you are asked to prune one of these filters in order to reduce the model size of the CNN, e.g., for purposes of running a TinyML application.

A. Which one would you prune based on the L_1 norm?
B. Which one would you prune based on some similarity measure among the three kernels? Define your similarity measure. For instance, you may get inspiration from matrix similarity concepts (each kernel is a matrix/tensor after all).

For both [A] and [B] show you analytic calculations.

SECTION 2: Recurrent neural networks

Problem-02

You are asked to generate an Auto Regressive (AR) model and then create an RNN (such as LSTM, or GRU) that predicts it. Generate samples of an Auto Regressive model of the form:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + a_4 X_{t-4} + U_t$$

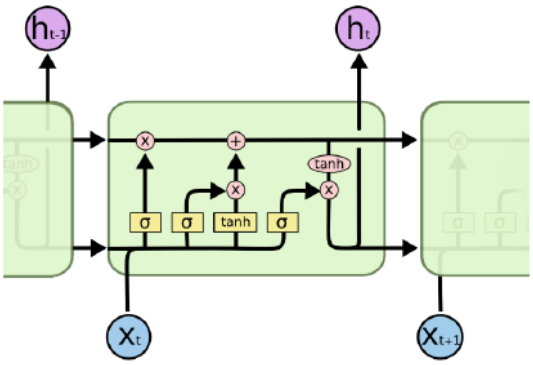
where $a_1 = 0.5$, $a_2 = -0.25$, $a_3 = 0.1$, $a_4 = -0.2$ and U_t is independent-identically distributed Uniform in the interval (0, 0.05).

A. Now design an RNN of your choice that predicts the sequence.
B. Apply the training algorithm on new samples and calculate the averaged cost square error cost function. Investigate different number of training samples. (You are not

allowed to use your model knowledge when you design the RNN.) Show the accuracy vs. training samples.

Problem-03

Recall the elements of a module in an LSTM and the corresponding computations, where ⊗ stands for point-wise multiplication:



$$f_t = \sigma(W_f h_{t-1} + U_f x_t)$$
$$i_t = \sigma(W_i h_{t-1} + U_i x_t)$$
$$o_t = \sigma(W_o h_{t-1} + U_o x_t)$$
$$\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t)$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$
$$h_t = o_t \odot \tanh(C_t)$$

A. What do the gates f_t, i_t and o_t do?

B. Which of the quantities next to the figure are always positive?

Let's now try to understand how this architecture approaches the vanishing gradients problem. To calculate the gradient $\partial L/\partial\theta$, where θ stands for the parameters (W_f, W_o, W_i, W_c), we now have to consider the cell state C_t instead of h_t (see course slides). Like h_t in normal RNNs, C_t will also depend on the previous cell states C_{t-1}, …, C_0, so we get a formula of the form[1]:

$$\frac{\partial L}{\partial W} = \sum_{t=0}^{T} \sum_{k=1}^{t} \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial C_k} \frac{\partial C_k}{\partial W}.$$

C. We know that $\dfrac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^{t} \dfrac{\partial C_t}{\partial C_{t-1}}$.

Let $f_t=1$ and $i_t=0$ such that $C_t=C_{t-1}$ for all t. What is the gradient $\partial C_t/C_k$ in this case?

SECTION 4: Fuzzy subsets theory

Problem-04

Consider the following reference set: {A, B, C, D, E, F, G}, and the fuzzy subsets

$\mathbb{A}$= {(A|0), (B|0.3), (C|0.7), (D|1), (E|0), (F|0.2), (G|0.6)},

$\mathbb{B}$= {(A|0.3), (B|1), (C|0.5), (D|0.8), (E|1), (F|0.5), (G|0.6)},

$\mathbb{C}$= {(A|1), (B|0.5), (C|0.5), (D|0.2), (E|0), (F|0.2), (G|0.9)}.

Calculate the following:

A. $\mathbb{A} \cap \mathbb{B}$

B. $\mathbb{A} \cup \mathbb{B}$

C. $\mathbb{A} \cap \mathbb{B}^c$

D. $(\mathbb{A} \cup \mathbb{B}^c) \cap \mathbb{C}$

E. $(\mathbb{A} \cap \mathbb{B})^c \cup \mathbb{C}^c$

F. $(\mathbb{A} \cap \mathbb{A}^c) \cup \mathbb{A}$.

[1] The real formula is a little more complicated since C_t also depends on f_t, i_t and \tilde{C}_t, which in turn all depend on W, but we choose to neglect this.

Problem-05

Considering the three fuzzy subsets of Problem-04, calculate:

A. $\mathbb{A} \hat{+} \mathbb{B} \hat{+} \mathbb{C}$ (algebraic sum)

B. $\mathbb{A} . (\mathbb{B} \hat{+} \mathbb{C})$

and prove:

C. $\mathbb{A} . \mathbb{A} \subset \mathbb{A}$ and $\mathbb{A} \hat{+} \mathbb{A} \supset \mathbb{A}$.

D. $\mathbb{A} . \mathbb{B} \hat{+} \mathbb{A} . \mathbb{C} \supset \mathbb{A} . (\mathbb{B} \hat{+} \mathbb{C})$

Problem-06

Give the power set of fuzzy subsets for the following cases:

A. $E=\{x_1, x_2\}$, $M=\{0, 1/3, 2/3, 1\}$

B. $E=\{x_1, x_2, x_3\}$, $M=\{a, b, c\}$, $a < b < c$.

Problem-07

Prove the fuzzy DeMorgan laws:

A. $X \cap Y = (X^c \cup Y^c)^c$

B. $X \cup Y = (X^c \cap Y^c)^c$

Problem-08

Dimitris and Fany go to park if it is a *beautiful day* and it is not too *hot*, or if it isn't *raining*. Assuming that:

- It is a *beautiful day* with 0.6 degree
- It is *hot* with 0.4 degree
- It is *raining* with 0.8 degree

with which degree Dimitris and Fany will go to park?

Problem-09

Let P(x) and Q(x) be fuzzy truth functions, each of which can only give truth values of 0, 0.5 and 1. That is, for all x, P(x) is in the set {0, 0.5, 1} and Q(x) is in the set {0, 0.5, 1}.

Recall the Kleene-Dienes definition of implication: "a → b" is equivalent to

"(not a) or b"

Compute the truth table for the fuzzy statement "(P(x) and (P(x) → Q(x))) → Q(x)".

How does this compare for the same truth table in crisp logic (where P(x) and Q(x) can only be "true" or "false")?

Problem-10

Assume that the truth function of A(x) is the following:

A(x) = 1 for x <= 2

A(x) = 1 - (x-2)/3 for 2 < x < 5

A(x) = 0 for x >= 5

and B(x) has the following:

B(x) = 0 for x <= 3

B(x) = (x-3)/4 for 3 < x < 7

B(x) = 1 for x >= 7.

Find for which values of x, the following statement has the maximum truth value:

"not(A(x) OR B(x))".

Problem-11

"very" is used as an adjective to reduce vagueness on fuzzy set membership. The interpretation is that if the statement "A is true" has truth value equal to x, then the statement "A is very true" has truth value x^2, because the "very true" is more demanding.

A. True or false (explain your answer): Let "S" be a fuzzy set. Then "Very S" is a fuzzy subset of "S".

"more or less" is used as an adjective to increase vagueness - the interpretation is that if "A is true" has truth value x, then "A is more or less true" has truth value $\sqrt{x}$.

B. True or false (explain your answer): Let "S" be a fuzzy set. Then "S" is a fuzzy subset of "more or less S".

C. Using the definitions just given, is it true that "not very S" is a subset of "more or less S", or vice versa, or is it impossible to say?

D. Is "not more or less S" a subset of "very S", or vice versa, or is it impossible to say?

Useful information:

The deadline is strict. It is possible to get an extension (up to 4 days), but you need to get the approval of the instructor, and this is going to cost a 10% penalty to the final grade of this Problem-Set. Delivery of the solutions' pdf (typeset or very very easy to read handwritten) is done by email to dkatsar@uth.gr. The subject of the message should be **strictly**: CE418-Problem set 03: AEM1-AEM2

Symbol interpretation:

It requires "algorithmic" and/or mathematical thinking.

It requires the development of code (in any language/platform you wish). The final deliverable should contain: a) The solution of the exercise, and b) The implementation source code.