

Problem Set 3

Course: ECE418

Nikos Mavros

February 17, 2026

Problem 01

A) Pruning based on L_1 Norm

The L_1 norm of a kernel is calculated as the sum of the absolute values of its elements: $\|K\|_1 = \sum |K_{i,j}|$. This metric is often used as a proxy for the importance of a filter; filters with smaller weights produce smaller activations and contribute less to the network's output.

Calculations:

$$\begin{aligned}\|K_1\|_1 &= |-0.3| + |-0.2| + |-0.4| + |0.5| + |0.7| + |0.6| + |0.1| + |0.4| + |0.3| \\ &= 0.3 + 0.2 + 0.4 + 0.5 + 0.7 + 0.6 + 0.1 + 0.4 + 0.3 = \mathbf{3.5}\end{aligned}$$

$$\begin{aligned}\|K_2\|_1 &= |0.7| + |0.5| + |0.6| + |-0.1| + |-0.3| + |-0.3| + |0.2| + |0.8| + |0.4| \\ &= 0.7 + 0.5 + 0.6 + 0.1 + 0.3 + 0.3 + 0.2 + 0.8 + 0.4 = \mathbf{3.9}\end{aligned}$$

$$\begin{aligned}\|K_3\|_1 &= |0.5| + |0.1| + |0.7| + |0.1| + |0.4| + |0.4| + |-0.5| + |-0.4| + |-0.8| \\ &= 0.5 + 0.1 + 0.7 + 0.1 + 0.4 + 0.4 + 0.5 + 0.4 + 0.8 = \mathbf{3.9}\end{aligned}$$

Decision: We select **Kernel** K_1 for pruning because it has the lowest L_1 norm (3.5), suggesting it is the least significant feature extractor among the three.

B) Pruning based on Similarity Measure

We quantify redundancy using **Cosine Similarity** on the flattened kernel tensors. If two kernels are highly similar (score close to 1) or highly dissimilar/inverse (score close to -1), they capture redundant information. We prune the one with the lower magnitude (L_1 norm) from the redundant pair.

The similarity between kernels A and B is defined as:

$$\text{sim}(A, B) = \frac{A \cdot B}{\|A\|_F \|B\|_F} = \frac{\sum_{i,j} A_{i,j} B_{i,j}}{\sqrt{\sum_{i,j} A_{i,j}^2} \sqrt{\sum_{i,j} B_{i,j}^2}}$$

1. Frobenius Norms ($\|K\|_F$):

- $\|K_1\|_F = \sqrt{(-0.3)^2 + \dots + 0.3^2} = \sqrt{1.65} \approx \mathbf{1.285}$
- $\|K_2\|_F = \sqrt{0.7^2 + \dots + 0.4^2} = \sqrt{2.13} \approx \mathbf{1.459}$
- $\|K_3\|_F = \sqrt{0.5^2 + \dots + (-0.8)^2} = \sqrt{2.13} \approx \mathbf{1.459}$

2. Dot Products ($K_A \cdot K_B$):

- $K_1 \cdot K_2 = (-0.3)(0.7) + (-0.2)(0.5) + \dots + (0.3)(0.4) = \mathbf{-0.53}$
- $K_1 \cdot K_3 = (-0.3)(0.5) + (-0.2)(0.1) + \dots + (0.3)(-0.8) = \mathbf{-0.33}$
- $K_2 \cdot K_3 = (0.7)(0.5) + (0.5)(0.1) + \dots + (0.4)(-0.8) = \mathbf{-0.17}$

3. Similarity Scores:

- $\text{sim}(K_1, K_2) = \frac{-0.53}{1.285 \times 1.459} \approx \mathbf{-0.283}$

- $\text{sim}(K_1, K_3) = \frac{-0.33}{1.285 \times 1.459} \approx -0.176$
- $\text{sim}(K_2, K_3) = \frac{-0.17}{1.459 \times 1.459} \approx -0.080$

Decision: While none of the kernels show extreme redundancy (values close to ± 1), the pair with the highest absolute similarity is K_1 and K_2 ($|\text{sim}| \approx 0.28$). This indicates these two filters are the most correlated (inversely) among the set.

Between this pair, K_1 has the smaller L_1 norm (3.5 vs 3.9). Therefore, we choose to prune **Kernel** K_1 .

Problem 02

A) Model Generation and RNN Design

We are tasked with modeling an Auto-Regressive (AR) process of order 4 defined by the equation:

$$X_t = 0.5X_{t-1} - 0.25X_{t-2} + 0.1X_{t-3} - 0.2X_{t-4} + U_t$$

where $U_t \sim \mathcal{U}(0, 0.05)$ represents independent, identically distributed uniform noise.

1. Data Generation Logic: To prepare the data for a Recurrent Neural Network, we generate a continuous time series using the AR(4) coefficients. A "burn-in" period of 100 samples is discarded to ensure the sequence stabilizes. The data is then sliding-windowed into input-output pairs (X_{seq}, y_{label}) :

- **Input:** A sequence window of size $T = 10$. This window size is intentionally larger than the lag order (4) to allow the network to autonomously learn which time steps are significant.
- **Target:** The immediate next value X_{t+1} in the sequence.

2. Network Architecture: We designed a Recurrent Neural Network using a **Gated Recurrent Unit (GRU)**. The architecture consists of:

- **Input Layer:** Accepts sequences of dimension 10×1 .
- **Hidden Layer:** A single GRU layer with **32 hidden units**. The GRU was selected over a standard RNN to better capture dependencies without the vanishing gradient problem, while being computationally more efficient than an LSTM.
- **Output Layer:** A fully connected (Linear) layer that maps the final hidden state h_t to a single scalar prediction.

3. Training Strategy: The model is trained using the **Adam optimizer** with an initial learning rate of 0.01 and **Mean Squared Error (MSE)** loss. To achieve the stable convergence seen in the results, we implemented two key optimizations:

- **Averaging:** For each sample size N , the training is repeated over multiple independent trials, and the final Test MSE is averaged. This minimizes the variance caused by random weight initialization.
- **Learning Rate Scheduler:** A scheduler decays the learning rate during training, allowing the model to make large updates early on and fine-tune the weights as it approaches the optimal solution.

B) Experimental Results

The following table summarizes the averaged cost square error for each sample size.

Table 1: Averaged Test MSE vs. Training Sample Size

Training Samples (N)	Averaged Test MSE
100	2.627×10^{-4}
500	2.167×10^{-4}
1000	2.107×10^{-4}
2500	2.105×10^{-4}
5000	2.097×10^{-4}

Observation: As illustrated in Figure 1, the model performance improves sharply as the training set grows from 100 to 500 samples. Beyond 1000 samples, the MSE stabilizes around 2.1×10^{-4} . This value is extremely close to the theoretical variance of the added noise ($\approx 2.08 \times 10^{-4}$), indicating that the GRU has successfully learned the underlying autoregressive rules and has converged to the noise floor of the system.

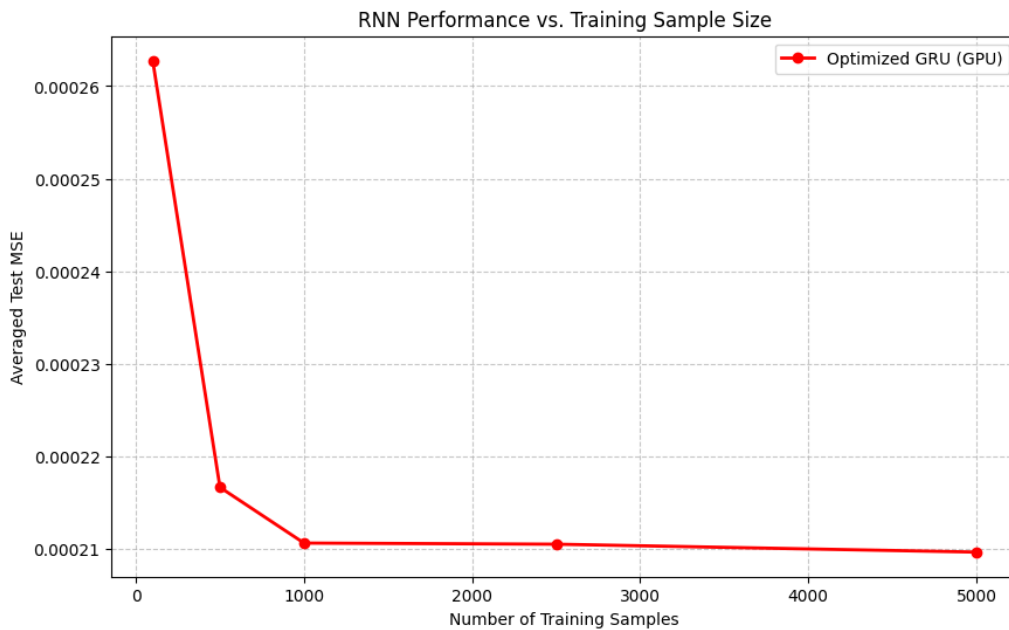


Figure 1: RNN Performance: The error decreases smoothly and converges to the theoretical limit as training data increases.

Problem 03

A) Function of LSTM Gates

The Long Short-Term Memory (LSTM) unit utilizes three distinct gates to regulate the flow of information:

- **Forget Gate (f_t):** This gate decides what information from the previous cell state (C_{t-1}) is no longer relevant and should be discarded. It applies a sigmoid activation to output values between 0 (completely forget) and 1 (completely retain).
- **Input Gate (i_t):** This gate controls the addition of new information. It determines which values from the candidate memory update (\tilde{C}_t)—derived from the current input x_t and previous hidden state h_{t-1} —are significant enough to be stored in the cell state.
- **Output Gate (o_t):** This gate regulates the output of the cell. It filters the current cell state (C_t) to produce the hidden state (h_t) for the current time step, ensuring that only contextually relevant information is passed to the next layer or time step.

B) Positivity of Quantities

We evaluate the range of values for each quantity based on their respective activation functions: the sigmoid function $\sigma(\cdot)$ outputs to $(0, 1)$, while the hyperbolic tangent $\tanh(\cdot)$ outputs to $(-1, 1)$.

Table 2: Analysis of Value Ranges

Variable	Activation Function / Formula	Always Positive?
f_t, i_t, o_t	Sigmoid: $\sigma(Wx + Uh)$	Yes (Range: 0 to 1)
\tilde{C}_t	Tanh: $\tanh(W_c x + U_c h)$	No (Range: -1 to 1)
C_t	Sum of products involving \tilde{C}_t	No (Depends on \tilde{C}_t)
h_t	Product involving $\tanh(C_t)$	No (Depends on \tanh)

C) Gradient Analysis ($f_t = 1, i_t = 0$)

We examine the gradient behavior in the specific case where the forget gate is fully open ($f_t = 1$) and the input gate is fully closed ($i_t = 0$).

1. Simplified Update Rule: Substituting the gate values into the cell state update equation:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$C_t = 1 \cdot C_{t-1} + 0 \cdot \tilde{C}_t = C_{t-1}$$

In this configuration, the cell state remains constant, effectively serving as a perfect memory of the previous state.

2. Local Gradient: The partial derivative of the current cell state with respect to the immediate previous state is:

$$\frac{\partial C_t}{\partial C_{t-1}} = 1$$

3. Gradient Propagation: Using the chain rule to calculate the gradient flow back to an arbitrary previous time step k :

$$\frac{\partial C_t}{\partial C_k} = \prod_{j=k+1}^t \frac{\partial C_j}{\partial C_{j-1}} = \prod_{j=k+1}^t 1 = 1$$

Conclusion: In this scenario, the gradient $\frac{\partial C_t}{\partial C_k}$ is exactly **1**. This demonstrates the core mechanism by which LSTMs mitigate the **vanishing gradient problem**. Unlike standard RNNs, where gradients effectively decay through repeated multiplication of weight matrices (often < 1), the additive nature of the LSTM cell allows the error signal to flow backward through time without attenuation. This capability enables the network to learn dependencies over long sequences.

Problem 04

Given Reference Set: $U = \{A, B, C, D, E, F, G\}$

Fuzzy Sets:

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

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

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

A) $A \cap B$

The intersection is defined by the minimum membership value: $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$.

- $x = A : \min(0, 0.3) = 0$
- $x = B : \min(0.3, 1) = 0.3$
- $x = C : \min(0.7, 0.5) = 0.5$
- $x = D : \min(1, 0.8) = 0.8$
- $x = E : \min(0, 1) = 0$
- $x = F : \min(0.2, 0.5) = 0.2$
- $x = G : \min(0.6, 0.6) = 0.6$

$$A \cap B = \{(A|0), (B|0.3), (C|0.5), (D|0.8), (E|0), (F|0.2), (G|0.6)\}$$

B) $A \cup B$

The union is defined by the maximum membership value: $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$.

- $x = A : \max(0, 0.3) = 0.3$
- $x = B : \max(0.3, 1) = 1$
- $x = C : \max(0.7, 0.5) = 0.7$
- $x = D : \max(1, 0.8) = 1$
- $x = E : \max(0, 1) = 1$
- $x = F : \max(0.2, 0.5) = 0.5$
- $x = G : \max(0.6, 0.6) = 0.6$

$$A \cup B = \{(A|0.3), (B|1), (C|0.7), (D|1), (E|1), (F|0.5), (G|0.6)\}$$

C) $A \cap B^c$

First, we define the complement B^c as $1 - \mu_B(x)$, then take the minimum with A .

- $x = A : \min(0, 1 - 0.3) = 0$
- $x = B : \min(0.3, 1 - 1) = 0$
- $x = C : \min(0.7, 1 - 0.5) = 0.5$
- $x = D : \min(1, 1 - 0.8) = 0.2$
- $x = E : \min(0, 1 - 1) = 0$
- $x = F : \min(0.2, 1 - 0.5) = 0.2$
- $x = G : \min(0.6, 1 - 0.6) = 0.4$

$$A \cap B^c = \{(A|0), (B|0), (C|0.5), (D|0.2), (E|0), (F|0.2), (G|0.4)\}$$

D) $(A \cup B^c) \cap C$

First, calculate $A \cup B^c = \max(\mu_A, 1 - \mu_B)$, then intersect with C .

- $x = A : \max(0, 0.7) \cap 1 \implies \min(0.7, 1) = 0.7$
- $x = B : \max(0.3, 0) \cap 0.5 \implies \min(0.3, 0.5) = 0.3$
- $x = C : \max(0.7, 0.5) \cap 0.5 \implies \min(0.7, 0.5) = 0.5$
- $x = D : \max(1, 0.2) \cap 0.2 \implies \min(1, 0.2) = 0.2$
- $x = E : \max(0, 0) \cap 0 \implies \min(0, 0) = 0$
- $x = F : \max(0.2, 0.5) \cap 0.2 \implies \min(0.5, 0.2) = 0.2$
- $x = G : \max(0.6, 0.4) \cap 0.9 \implies \min(0.6, 0.9) = 0.6$

$$(A \cup B^c) \cap C = \{(A|0.7), (B|0.3), (C|0.5), (D|0.2), (E|0), (F|0.2), (G|0.6)\}$$

E) $(A \cap B)^c \cup C^c$

Using De Morgan's Law, this is equivalent to finding the complement of the intersection of all three sets, or calculating directly: $\max(1 - \min(A, B), 1 - C)$.

- $x = A : \max(1 - 0, 1 - 1) = 1$
- $x = B : \max(1 - 0.3, 1 - 0.5) = 0.7$
- $x = C : \max(1 - 0.5, 1 - 0.5) = 0.5$
- $x = D : \max(1 - 0.8, 1 - 0.2) = 0.8$
- $x = E : \max(1 - 0, 1 - 0) = 1$
- $x = F : \max(1 - 0.2, 1 - 0.2) = 0.8$
- $x = G : \max(1 - 0.6, 1 - 0.9) = 0.4$

$$(A \cap B)^c \cup C^c = \{(A|1), (B|0.7), (C|0.5), (D|0.8), (E|1), (F|0.8), (G|0.4)\}$$

F) $(A \cap A^c) \cup A$

Calculate $\mu_{A \cap A^c}(x) = \min(\mu_A, 1 - \mu_A)$, then take the maximum with μ_A . Note that mathematically, this expression simplifies back to set A .

- $x = A : \max(\min(0, 1), 0) = 0$
- $x = B : \max(\min(0.3, 0.7), 0.3) = 0.3$
- $x = C : \max(\min(0.7, 0.3), 0.7) = 0.7$
- $x = D : \max(\min(1, 0), 1) = 1$
- $x = E : \max(\min(0, 1), 0) = 0$
- $x = F : \max(\min(0.2, 0.8), 0.2) = 0.2$
- $x = G : \max(\min(0.6, 0.4), 0.6) = 0.6$

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

Problem 05

Given Fuzzy Subsets:

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

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

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

A) Algebraic Sum: $A \hat{+} B \hat{+} C$

The membership function for the algebraic sum of three sets is derived as:

$$\mu_{A \hat{+} B \hat{+} C}(x) = \mu_A + \mu_B + \mu_C - (\mu_A \mu_B + \mu_A \mu_C + \mu_B \mu_C) + \mu_A \mu_B \mu_C$$

This formula represents the probabilistic union (inclusion-exclusion principle). Note: If any membership value is 1, the algebraic sum is 1.

Calculations per element:

- **A:** $0 + 0.3 + 1 - (0 + 0 + 0.3) + 0 = \mathbf{1.0}$
- **B:** $0.3 + 1 + 0.5 - (0.3 + 0.15 + 0.5) + 0.15 = 1.8 - 0.95 + 0.15 = \mathbf{1.0}$
- **C:** $0.7 + 0.5 + 0.5 - (0.35 + 0.35 + 0.25) + 0.175 = 1.7 - 0.95 + 0.175 = \mathbf{0.925}$
- **D:** Since $\mu_A(D) = 1$, the sum is **1.0**.
- **E:** Since $\mu_B(E) = 1$, the sum is **1.0**.
- **F:** $0.2 + 0.5 + 0.2 - (0.1 + 0.04 + 0.1) + 0.02 = 0.9 - 0.24 + 0.02 = \mathbf{0.68}$

Problem 06

The power set of fuzzy subsets, denoted as M^E , represents the set of all possible membership functions $\mu : E \rightarrow M$. The total number of such fuzzy subsets is given by the cardinality $|M|^{|E|}$.

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

Here, $|E| = 2$ and $|M| = 4$. The total number of fuzzy subsets is $4^2 = 16$. We denote each subset as a set of pairs $\{(x_1, \mu(x_1)), (x_2, \mu(x_2))\}$.

The Power Set $\mathcal{P}(E)$:

$\{(x_1, 0), (x_2, 0)\}, \{(x_1, 0), (x_2, 1/3)\}, \{(x_1, 0), (x_2, 2/3)\}, \{(x_1, 0), (x_2, 1)\},$
 $\{(x_1, 1/3), (x_2, 0)\}, \{(x_1, 1/3), (x_2, 1/3)\}, \{(x_1, 1/3), (x_2, 2/3)\}, \{(x_1, 1/3), (x_2, 1)\},$
 $\{(x_1, 2/3), (x_2, 0)\}, \{(x_1, 2/3), (x_2, 1/3)\}, \{(x_1, 2/3), (x_2, 2/3)\}, \{(x_1, 2/3), (x_2, 1)\},$
 $\{(x_1, 1), (x_2, 0)\}, \{(x_1, 1), (x_2, 1/3)\}, \{(x_1, 1), (x_2, 2/3)\}, \{(x_1, 1), (x_2, 1)\}$

B) Universe $E = \{x_1, x_2, x_3\}$, Membership $M = \{a, b, c\}$ (where $a < b < c$)

Here, $|E| = 3$ and $|M| = 3$. The total number of fuzzy subsets is $3^3 = 27$. We denote each subset as $\{(x_1, \mu_1), (x_2, \mu_2), (x_3, \mu_3)\}$.

The Power Set $\mathcal{P}(E)$:

$\{(x_1, a), (x_2, a), (x_3, a)\}, \{(x_1, a), (x_2, a), (x_3, b)\}, \{(x_1, a), (x_2, a), (x_3, c)\},$
 $\{(x_1, a), (x_2, b), (x_3, a)\}, \{(x_1, a), (x_2, b), (x_3, b)\}, \{(x_1, a), (x_2, b), (x_3, c)\},$
 $\{(x_1, a), (x_2, c), (x_3, a)\}, \{(x_1, a), (x_2, c), (x_3, b)\}, \{(x_1, a), (x_2, c), (x_3, c)\},$

 $\{(x_1, b), (x_2, a), (x_3, a)\}, \{(x_1, b), (x_2, a), (x_3, b)\}, \{(x_1, b), (x_2, a), (x_3, c)\},$
 $\{(x_1, b), (x_2, b), (x_3, a)\}, \{(x_1, b), (x_2, b), (x_3, b)\}, \{(x_1, b), (x_2, b), (x_3, c)\},$
 $\{(x_1, b), (x_2, c), (x_3, a)\}, \{(x_1, b), (x_2, c), (x_3, b)\}, \{(x_1, b), (x_2, c), (x_3, c)\},$

 $\{(x_1, c), (x_2, a), (x_3, a)\}, \{(x_1, c), (x_2, a), (x_3, b)\}, \{(x_1, c), (x_2, a), (x_3, c)\},$
 $\{(x_1, c), (x_2, b), (x_3, a)\}, \{(x_1, c), (x_2, b), (x_3, b)\}, \{(x_1, c), (x_2, b), (x_3, c)\},$
 $\{(x_1, c), (x_2, c), (x_3, a)\}, \{(x_1, c), (x_2, c), (x_3, b)\}, \{(x_1, c), (x_2, c), (x_3, c)\}$

Problem 07

To prove the fuzzy De Morgan laws, we utilize the standard Zadeh membership definitions for fuzzy sets X and Y with membership functions $\mu_X(u)$ and $\mu_Y(u)$ in the interval $[0, 1]$.

- Intersection: $\mu_{X \cap Y} = \min(\mu_X, \mu_Y)$
- Union: $\mu_{X \cup Y} = \max(\mu_X, \mu_Y)$
- Complement: $\mu_{X^c} = 1 - \mu_X$

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

We start by expanding the membership function of the Right Hand Side (RHS).

$$\begin{aligned}\mu_{(X^c \cup Y^c)^c} &= 1 - \mu_{(X^c \cup Y^c)} \\ &= 1 - \max(\mu_{X^c}, \mu_{Y^c}) \\ &= 1 - \max(1 - \mu_X, 1 - \mu_Y)\end{aligned}$$

We apply the arithmetic identity $1 - \max(a, b) = \min(1 - a, 1 - b)$:

$$1 - \max(1 - \mu_X, 1 - \mu_Y) = \min(1 - (1 - \mu_X), 1 - (1 - \mu_Y))$$

Simplifying the terms inside the minimum function:

$$\min(\mu_X, \mu_Y) = \mu_{X \cap Y}$$

Since the membership function of the RHS simplifies to exactly the membership function of the LHS ($X \cap Y$), the identity is proven.

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

Similarly, we expand the membership function of the RHS.

$$\begin{aligned}\mu_{(X^c \cap Y^c)^c} &= 1 - \mu_{(X^c \cap Y^c)} \\ &= 1 - \min(\mu_{X^c}, \mu_{Y^c}) \\ &= 1 - \min(1 - \mu_X, 1 - \mu_Y)\end{aligned}$$

We apply the arithmetic identity $1 - \min(a, b) = \max(1 - a, 1 - b)$:

$$1 - \min(1 - \mu_X, 1 - \mu_Y) = \max(1 - (1 - \mu_X), 1 - (1 - \mu_Y))$$

Simplifying the terms inside the maximum function:

$$\max(\mu_X, \mu_Y) = \mu_{X \cup Y}$$

Thus, the RHS is equivalent to the LHS, and the law is proven.

Problem 08

The decision rule for going to the park is described by the compound statement:

*“(Beautiful day **AND** Not too hot) **OR** (Isn’t raining)”*

1. Define Fuzzy Variables: We assign the given membership degrees to the fuzzy variables:

- B : It is a beautiful day ($\mu_B = 0.6$)
- H : It is hot ($\mu_H = 0.4$)

- R : It is raining ($\mu_R = 0.8$)

2. Mathematical Formulation: We translate the linguistic connectives into fuzzy logic operators:

- **NOT** ($\neg A$): Complement $\rightarrow 1 - \mu_A$
- **AND** ($A \cap B$): Intersection $\rightarrow \min(\mu_A, \mu_B)$
- **OR** ($A \cup B$): Union $\rightarrow \max(\mu_A, \mu_B)$

The membership function for the decision to go to the park (μ_{Park}) is:

$$\mu_{Park} = \max(\min(\mu_B, 1 - \mu_H), 1 - \mu_R)$$

3. Step-by-Step Calculation:

1. Calculate Complements (Negations):

- Not Hot (H^c): $1 - 0.4 = \mathbf{0.6}$
- Not Raining (R^c): $1 - 0.8 = \mathbf{0.2}$

2. Evaluate the Conjunction (AND): "*Beautiful day AND Not hot*"

$$\mu_{B \cap H^c} = \min(\mu_B, \mu_{H^c}) = \min(0.6, 0.6) = \mathbf{0.6}$$

3. Evaluate the Disjunction (OR): "*(Beautiful day AND Not hot) OR (Not Raining)*"

$$\mu_{Park} = \max(\mu_{B \cap H^c}, \mu_{R^c}) = \max(0.6, 0.2) = \mathbf{0.6}$$

Conclusion: The degree to which Dimitris and Fany will go to the park is **0.6**.

Problem 09

Definitions: Let $P(x)$ and $Q(x)$ be fuzzy truth functions with values in the set $\{0, 0.5, 1\}$. We use the **Kleene-Dienes** definition of implication:

$$a \rightarrow b \equiv \neg a \vee b \equiv \max(1 - a, b)$$

The logical expression to evaluate is the Modus Ponens form:

$$\Phi = (P(x) \wedge (P(x) \rightarrow Q(x))) \rightarrow Q(x)$$

where conjunction (\wedge) is $\min(a, b)$ and disjunction (\vee) is $\max(a, b)$.

Fuzzy Truth Table

We calculate the intermediate values step-by-step:

1. **Imp:** $P \rightarrow Q = \max(1 - P, Q)$
2. **Premise:** $P \wedge (P \rightarrow Q) = \min(P, \text{Imp})$
3. **Result:** $\text{Premise} \rightarrow Q = \max(1 - \text{Premise}, Q)$

P(x)	Q(x)	Imp ($P \rightarrow Q$)	Premise ($P \wedge \text{Imp}$)	Result (Φ)
0	0	1	0	1
0	0.5	1	0	1
0	1	1	0	1
0.5	0	0.5	0.5	0.5
0.5	0.5	0.5	0.5	0.5
0.5	1	1	0.5	1
1	0	0	0	1
1	0.5	0.5	0.5	0.5
1	1	1	1	1

Comparison with Crisp Logic

1. Crisp Logic (Classical Logic): In crisp logic, variables can only take the values $\{0, 1\}$ (False or True). If we look exclusively at the rows in the table where P and Q are either 0 or 1 (rows 1, 3, 7, and 9), the **Result** is always **1**. This means that in classical logic, the statement $(P \wedge (P \rightarrow Q)) \rightarrow Q$ is a **tautology**—it is always true regardless of the truth values of the variables.

2. Fuzzy Logic (Kleene-Dienes): In this fuzzy system, the statement is **not a tautology**. As seen in the table, there are specific cases (e.g., when $P = 0.5$ and $Q = 0$, or $P = 1$ and $Q = 0.5$) where the result is **0.5**. The introduction of the "uncertain" truth value (0.5) allows the validity of the implication chain to degrade. Specifically, if the premise is partially true, the conclusion is not guaranteed to be fully true.

Problem 10

Given Fuzzy Membership Functions:

$$A(x) = \begin{cases} 1 & x \leq 2 \\ 1 - \frac{x-2}{3} & 2 < x < 5 \\ 0 & x \geq 5 \end{cases} \quad B(x) = \begin{cases} 0 & x \leq 3 \\ \frac{x-3}{4} & 3 < x < 7 \\ 1 & x \geq 7 \end{cases}$$

1. Objective Function

We want to maximize the truth value of the statement:

$$\Phi(x) = \neg(A(x) \vee B(x))$$

Using standard fuzzy logic definitions ($\neg z = 1 - z$ and $a \vee b = \max(a, b)$), the objective function becomes:

$$f(x) = 1 - \max(A(x), B(x))$$

To **maximize** the result $f(x)$, we must **minimize** the term $\max(A(x), B(x))$.

2. Analysis of Intersection

We observe the behavior of the functions in the interval where they overlap:

- $A(x)$ is a **decreasing** function (going from 1 to 0 on $(2, 5)$).
- $B(x)$ is an **increasing** function (going from 0 to 1 on $(3, 7)$).

The minimum value of $\max(A(x), B(x))$ for two intersecting monotonic functions (one increasing, one decreasing) occurs exactly at their **intersection point**.

The intersection must occur in the interval $x \in (3, 5)$, where both functions are active (neither is 0 or 1).

3. Calculation

We set the equations equal to each other in the interval $(3, 5)$:

$$1 - \frac{x-2}{3} = \frac{x-3}{4}$$

Multiply by 12 to clear the denominators:

$$\begin{aligned} 12 - 4(x-2) &= 3(x-3) \\ 12 - 4x + 8 &= 3x - 9 \\ 20 - 4x &= 3x - 9 \\ 29 &= 7x \\ x &= \frac{29}{7} \approx \mathbf{4.14} \end{aligned}$$

4. Verification and Maximum Truth Value

First, we find the value of the intersection (the "OR" part):

$$\max(A, B) = B\left(\frac{29}{7}\right) = \frac{\frac{29}{7} - 3}{4} = \frac{\frac{29-21}{7}}{4} = \frac{8}{28} = \frac{2}{7} \approx 0.286$$

Now, we calculate the final truth value for the "NOT" statement:

$$f(x) = 1 - \frac{2}{7} = \frac{5}{7} \approx \mathbf{0.714}$$

Conclusion: The statement has its maximum truth value of **0.714** at $\mathbf{x \approx 4.14}$.

Problem 11

In fuzzy set theory, "linguistic hedges" modify the shape of membership functions. Let S be a fuzzy set with membership function $\mu_S(x) = x$, where $x \in [0, 1]$. We use the definitions provided in the problem statement:

- **Very S:** $\mu_{Very}(x) = x^2$ (Concentration)
- **More or Less S:** $\mu_{MoL}(x) = \sqrt{x}$ (Dilation)
- **Not A:** $\mu_{\neg A}(x) = 1 - \mu_A(x)$

Note on Subset Definition: A fuzzy set A is a subset of B ($A \subseteq B$) if and only if $\mu_A(x) \leq \mu_B(x)$ for all x .

A) "Very S" is a fuzzy subset of "S"

Answer: True

Proof: We compare $\mu_{Very}(x) = x^2$ and $\mu_S(x) = x$. Since the truth values are normalized in the interval $x \in [0, 1]$, the square of a number is always less than or equal to the number itself:

$$x^2 \leq x \quad \forall x \in [0, 1]$$

Therefore, $\mu_{Very}(x) \leq \mu_S(x)$, which satisfies the definition of a subset.

$$\text{Very } S \subseteq S$$

B) "S" is a fuzzy subset of "More or Less S"

Answer: True

Proof: We compare $\mu_S(x) = x$ and $\mu_{MoL}(x) = \sqrt{x}$. For any $x \in [0, 1]$, the square root is greater than or equal to the number itself:

$$x \leq \sqrt{x} \quad \forall x \in [0, 1]$$

Therefore, $\mu_S(x) \leq \mu_{MoL}(x)$, which satisfies the definition of a subset.

$$S \subseteq \text{More or Less } S$$

C) "Not Very S" vs "More or Less S"

Answer: Neither is a subset of the other (Impossible to say)

Analysis:

- $\mu_{\neg Very}(x) = 1 - x^2$
- $\mu_{MoL}(x) = \sqrt{x}$

We test the relationship at the boundaries of the universe $[0, 1]$:

- At $x = 0$: $\mu_{\neg Very}(0) = 1$ and $\mu_{MoL}(0) = 0$. Here, **LHS $\not\subseteq$ RHS**.
- At $x = 1$: $\mu_{\neg Very}(1) = 0$ and $\mu_{MoL}(1) = 1$. Here, **RHS $\not\subseteq$ LHS**.

Since the inequality direction changes depending on the value of x , neither set is contained entirely within the other.

D) "Not More or Less S" vs "Very S"

Answer: Neither is a subset of the other (Impossible to say)

Analysis:

$$- \mu_{\neg MoL}(x) = 1 - \sqrt{x}$$

$$- \mu_{Very}(x) = x^2$$

We test the relationship at the boundaries:

$$- \text{At } x = 0: \mu_{\neg MoL}(0) = 1 \text{ and } \mu_{Very}(0) = 0. \text{ Here, } \mathbf{LHS} \not\subseteq \mathbf{RHS}.$$

$$- \text{At } x = 1: \mu_{\neg MoL}(1) = 0 \text{ and } \mu_{Very}(1) = 1. \text{ Here, } \mathbf{RHS} \not\subseteq \mathbf{LHS}.$$

Similar to part C, the curves intersect within the interval, meaning neither condition ($A \subseteq B$ or $B \subseteq A$) holds for all x .