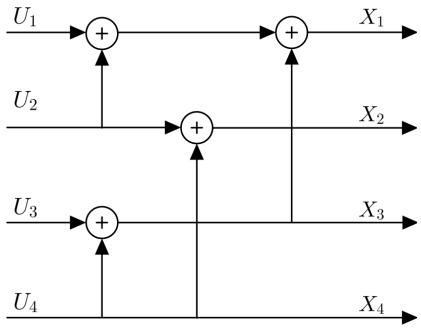
Final Exam - Moed B

Total time for the exam: 2 hours!

Please copy the following sentence and sign it: " I am respecting the rules of the exam: Signature:_____ "

1) **Decoding a Compressed Message (50 Points):** In this problem, you will learn how to decode a compressed message using polar codes. Consider the encoding scheme with a block size of $N = 4$ shown in the figure below, where $U_i \sim \text{Bern}(0.1)$ for $i = 1, \ldots, 4$.



Items (a) to (c) are related and should be solved in sequence.

a) **(8)** Suppose we want to compress the message $(U_1, \ldots, U_4)$ with a rate of $R = 0.75$. Specify which bits $\{X_i\}_{i=1}^4$ should be used to represent the compressed message. Justify your answer.

b) **(8 points)** Given the message $(U_1, U_2, U_3, U_4) = (1, 0, 1, 1)$, and based on your answer to the previous item, write the compressed message.

c) **(12 points)** We would like to decode the original message from the compressed message. Apply the successive cancellation (SC) decoding algorithm, and determine if decoding succeeded or not.

Items (d) to (f) are independent and not related to the previous items.

d) **(6 points)** Can the encoder predict which errors the SC decoder might make? Explain your answer in no more than two lines.

e) **(6 points)** What additional information could the encoder send to the decoder to help correct errors?

f) **(10 points)** Consider that we aim to achieve lossless compression (without errors during decoding) by allowing a variable block length. Using the SC algorithm, propose a method to ensure this while maintaining the compression rate as low as possible.
Remark: Variable block length refers to a compression scheme where the length of the blocks of data being encoded or compressed can vary and depend on the transmitted data.

2) **Combining Logistic Regression with MINE (50 Points):**
In this question, we aim to predict an individual's life expectancy based on their health information. The features of the health information are as follows: Smoking habit: $x_1 \in [0, 1]$ (proportion of smoking), BMI (Body Mass Index): $x_2 \in \mathbb{R}^+$, Hours worked per day: $x_3 \in \{0, \ldots, 24\}$, Regular exercise: $x_4 \in \{0, 1\}$ (1 indicates the individual exercises regularly, 0 otherwise). Let $\bar{X} = (x_1, x_2, x_3, x_4)$ represent the vector of these four features. The target variable $Y \in \{0, 1, \ldots, 120\}$ represents the possible lifespan (in years) of the individual. We are interested in calculating the probability $P(Y = n \mid \bar{X})$, which is the probability that the individual will live for exactly n years. For example, $P(Y = 5 \mid \bar{X})$ is the probability that the individual, based on their health features $\bar{X}$, will live exactly 5 years. In this question, we will compute $P(Y = n \mid \bar{X})$ in two different ways.

Part 1: Probability Analysis Using Machine Learning
The goal is to build a model that outputs a vector of size 121 (for ages 0 to 120), representing the

probability of death for each year. In other words, the model's output is:

$$(\hat{P}(Y = 0 \mid \bar{X}), \hat{P}(Y = 1 \mid \bar{X}), \ldots, \hat{P}(Y = 120 \mid \bar{X}))$$

where

$$\hat{P}(Y = i \mid \bar{X}) = \frac{e^{\theta_i^T \cdot \bar{X}}}{\sum_{j=0}^{120} e^{\theta_j^T \cdot \bar{X}}}$$

and $\theta_i \in \mathbb{R}^4$ are learnable weights.

a) **(6 points)** Is this problem an example of regression or classification? Explain your answer in two sentences.

b) **(14 points)** Describe how such a model can be built. Explain the role of each block in the model and provide a diagram illustrating the model's layers. Limit your answer to half a page.

c) **(10)** What should the cost function be for this model? Justify your choice in no more than three sentences.

Part 2: Probability Analysis Using Information Theory

In this section, we explore another method to predict $P(Y \mid \bar{X})$ using MINE (Mutual Information Neural Estimator).

d) **(4 points)** Let Q_z be a random variable uniformly distributed on $\{0, \ldots, 120\}$. Express the Kullback-Leibler (KL) divergence $D_{KL}(P_Z \| Q_Z)$ in terms of the entropy of $Z \sim P_Z$.

e) **(6 points)** We have sampled a large number of i.i.d pairs $(\bar{X}_1, Y_1), \ldots, (\bar{X}_M, Y_M)$. Using the Donsker-Varadhan representation, propose a method to estimate $P(Y)$. Provide a detailed algorithm.

f) **(10 points)** Is it possible to estimate $P(Y \mid \bar{X})$ using MINE? If yes, provide the algorithm. If no, explain why not.

Good Luck!