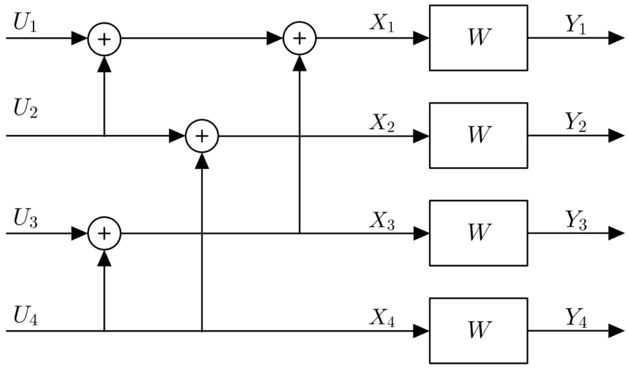
Introduction to Information Theory and Machine Learning July 28th, 2024.
(Prof. Permuter Haim, Mr. Bashar Huleihel, and Mr. Eyal Yakir)

Final Exam - Moed A
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) **Polar code design (50 Points):** In this problem, we explore how the selection of frozen bits affects the decoding process in the construction of polar codes. Consider a binary erasure channel (BEC) W, and the encoding scheme of block size $N = 4$ in the figure below.



For items (a) to (c), assume that the bits U_1 and U_2 are frozen to 0, while U_3 and U_4 are used for message bits.

a) **(4 points)** Calculate the code rate.
b) **(4 points)** Given the message bits $(U_3, U_4) = (1, 1)$, perform the encoding to determine the codeword (X_1, X_2, X_3, X_4).
c) **(10 points)** Apply successive cancellation decoding to the received vector $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 1)$. Determine if decoding is successful, and if so, identify the decoded message.

For items (d) to (f), consider a different scenario where U_2 and U_3 are frozen to 0, while U_1 and U_4 are message bits.

d) **(10 points)** Apply successive cancellation decoding for received vector $(Y_1, Y_2, Y_3, Y_4) = (1, ?, ?, 0)$, and show that the decoding might fail when decoding U_1.
e) **(5 points)** Explain the reason for the decoding failure.
f) **(10 points)** Perform optimal maximum likelihood decoding for the same received vector $(1, ?, ?, 0)$. To do this:
   - First, compute the codeword (X_1, X_2, X_3, X_4) for all four possible input messages (U_1, U_4).
   - Then, identify the input message(s) (U_1, U_4) that could result in the received vector $(1, ?, ?, 0)$. If more than one message corresponds to the received vector, declare a failure in decoding.

   Determine if the decoding is successful, and if so, specify the decoded message.
g) **(7 points)** For a code rate $R = 0.5$, determine which bits should be set as frozen to achieve the best performance in terms of bit error rate when using the successive cancellation decoding algorithm. Justify your selection.

2) **Combine linear regression with MINE (50 Points):** In this question, we will create and analyze a model for weather prediction. The task is as follows: given sequences of size N denoted as $X^N = (x_1, x_2, \ldots, x_N)$ where x_i is the temperature on the i-th day we would like to prediect the next day x_{N+1}. To solve this problem, the following model was suggested:

$$x_{i+1} = \sum_{j=0}^{D} w_j \cos\left(\frac{2\pi j}{365} \cdot i\right) + \sum_{j=0}^{L} \alpha_j x_{i-j} + b_{i+1}, \tag{1}$$

where D, L are constants, and b_i is an i.i.d random variable.

a) **Model Analysis**
   i) **(7 points)** In class, we learned that a linear regression problem can be written as

$$Y = \theta^T \cdot X. \tag{2}$$

Can you build a linear estimator for $\hat{x}_{i+1}$ based on previous samples using linear regression? If yes, provide Y, θ, and X. If not, explain why.

ii) **(4 points)** The model in Eq (1) assume some assumptions, explain what are the assumptions.

iii) **(13 points)** Write the gradient descent formula for updating the weights using the MSE cost function.

iv) **(9 points)** The researcher decided to change the cost function to the following:

$$\text{cost} = \sum_{x \in \text{data}} (y - \hat{y}(x))^2 + \lambda \sum_{i=1}^{M} \theta_i^2 - 2\theta_i.$$

Where $\theta = (\theta_1, \theta_2 \ldots, \theta_M)$ is the vector defined in Eq (2). How would we expect the model and model performance to change under the following λ:

- $\lambda \to \infty$
- λ is a positive finite number

Remember, when first initializing the model, the weights are assigned at random.

b) **Probability Analysis** Let $z_{i+1} = x_{i+1} - \hat{x}_{i+1}$ be the error process with $P_z : [-50, 50] \to [0, 1]$ with i.i.d distribution. Let $Q_z \sim U[-50, 50]$ represent a uniformly distributed random variable.

i) **(5 points)** Express the KL divergence $D_{KL}(P_z, Q_z)$ in terms of the entropy of $Z \sim P_z$.

We sampled a large number of samples $x_1, \ldots, x_M$. In the class we learned the Donsker-Varadhan representation

$$D_{KL}(P \parallel Q) = \sup_{T:\Omega \to \mathbb{R}} \left(\mathbb{E}_P[T] - \log(\mathbb{E}_Q[e^T]) \right). \tag{3}$$

ii) **(6 points)** Offer a method to estimate the entropy of the process z_i.

iii) **(6 points)** Offer a method to estimate the probability density of the process z_i .

\- 2

Good Luck!