CS301 - Data Science Final Spring 2023

PROBLEM SET 1 - TAKE AT HOME (35 Points)

DO NOT ORGET TO UPLOAD THE TAKE AT HOME PROBLEM SOLUTION IN CANVAS. THE DEADLINE IS THE END OF YOUR IN-PERSON FINAL EXAM

PROBLEM SET 2 (10 Points)

PS 2.A (5 points)

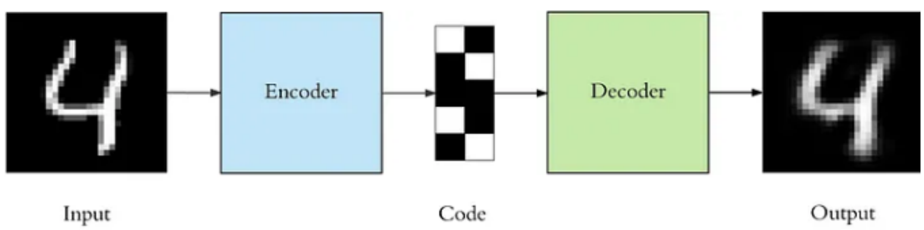
What is the probability of $x = 0.1$ when $x \sim N(0, 1)$? Explain your answer.

PS 2.B (5 points)

What is the likelihood of $x = 0.1$ when $x \sim N(0, 1)$? Explain your answer.

PROBLEM SET 3 (20 Points)

PS 3.A (15 points)



Design a neural network that will receive an input image $x \in R^{28 \times 28}$ and will *reconstruct it* at the output.

Your design should include all dimensions of the network including the number of layers, the number of neurons per layer, the activation functions, the loss function, the optimizer. You can draw the network and include *all* information on the drawing or provide it under the drawing.

PS 3.B (5 points)

What can we use such network for ? List all possible applications that you can think of.

PROBLEM SET 4 (20 Points)

A penalty factor of a weight decay regularizer for a densely connected layer is given by the equation:

$$L_R = \lambda \times \sum_{i,j} W_{i,j}^2$$

PS 4.A (10 points)

Draw the block diagram that will allow the regularizer to adjust the loss of a neural network with two dense hidden layers and tasked to do K-class classification. Ensure that you quote all needed dimensions in the tensors of the diagram.

PS 4.B (10 points)

Forward propagate the network to the point where you can get to the equation of the *total* loss.

PROBLEM SET 5 (15 Points)

A random variable y is distributed according to the Bernoulli distribution with parameter θ. You met the Bernoulli distribution in the context of the coin toss experiment and in binary classification. The probability density function of y is given by:

$$p(y|\theta) = \theta^y (1-\theta)^{1-y}$$

where $y \in \{0, 1\}$ and $\theta \in [0, 1]$.

The mean of y is $E(x) = \theta$ and its variance $\sigma^2 = E(x^2) - (E(x))^2 = \theta(1-\theta)$.

You create a set D of m i.i.d. samples $\{y_1, y_2, ..., y_m\}$ from $p(y|\theta)$ e.g.

$$D = \{1, 0, 1, 1, 0, 0, 0, 0\}$$

This random stream of values goes through an amplifier that creates at its output a stream of z values $z = 1000y$.

PS 5.A (5 points)

Write down the set Z that was trivially created by amplification.

PS 5.B (10 points)

Your colleague claims that the variance of z is a better metric than entropy to capture the uncertainty of the random variable z.

Show why they are wrong.

HINT: To answer this question you need to show the variance and the entropy of y and z or equivalently you can calculate the KL divergence between $p(y|\theta)$ and $p(z|\theta)$.

The KL diverfence is given by:

$$D_{KL}(P||Q) = \sum_i P(i) \log_2(P(i)/Q(i))$$