Assignment (50 Marks)

1. (Marks: 10)
   (Feature Masking as regularization) consider linear regression model by minimizing the squared loss function $\sum_{n=1}^{N} (y_n - w^T x_n)^2$. Suppose we decide to mask out or "drop" each feature x_{nd} of each input x_n ϵR^D, independently, with probability (1-p) (equivalently, retaining the feature with probability p). Masking or dropping out basically means that we will set the feature x_{nd} to 0 with probability (1-p). Essentially, it would be equivalent to replacing each input x_n by $\overline{x_n} = x_n \circ m_n$, where ◦denotes element wise product and m_n denotes the $D \times 1$ binary mask vector with $m_{nd} \sim Bernoulli(p)$ (m_{nd} = 1 means the feature x_{nd} was retained; m_{nd} = 0 means the feature x_{nd} was masked/zeroed).

   Let us now define a new loss function using these masked inputs as follows: $\sum_{n=1}^{N} (y_n - w^T \overline{x_n})^2$. Show that minimizing the expected value of this new loss function (where the expectation is used since the mask vectors m_n are random) is equivalent to minimizing a **regularized** loss function. Clearly write down the expression of this regularized loss function.

2. Explore, define, and explain the following terms: (11 marks)
   1. Least squares problem
   2. Generalization
   3. Cross-validation, Hold-out validation, Leave-one out validation
   4. KKT conditions
   5. Feature Correlation
   6. Bias and Variance tradeoff
   7. Prior Probability, Conditional Probability, and Posterior Probability

3. Explain the issues with the current AI models and the problem of unity of perception. (4 Marks)

4. Consider the following set of training examples: (Marks: 2)

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of a2 relative to these training examples? Provide the equation for calculating the information gain.

5. How do you handle collinearity in a linear regression model? (Hint: read about assumptions of linear regression) (Marks: 2)

6. Given a dataset for utility fraud detection, you built a classifier model that achieved a performance score of 98.5%. Is this a good model? If yes, justify your answer. If not, what can you do to improve it? (Marks: 2)

7. Why would you Prune a decision tree? (Marks: 2)

8. There is a dictionary containing information about employees in a company. (Marks: 8)
employee_data = { 'EmployeeID': [101, 102, 103, 104, 105],

'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],

'Department': ['HR', 'Finance', 'IT', 'HR', 'Finance'],

'Salary': [60000, 70000, 80000, 65000, 72000],

'JoiningDate': ['2020-01-15', '2019-03-23', '2021-06-01', '2020-07-30', '2018-11-20']

Your tasks are as follows:

1. Create a pandas DataFrame from the **employee_data** dictionary.
2. Display the **first 3 rows** of the DataFrame.
3. Calculate the **average salary of employees** in each department.
4. Find the employee with the **highest salary**.
5. Determine the **number of employees** who joined after **January 1, 2020**.
6. Add a new column called **YearsInCompany** that shows the number of years each employee has been in the company (as of the current date).
7. Filter and display the data of employees who are in the **'Finance' departmen**t.

9. (Marks: 9)

Derive the corresponding equations and solutions for the primal and dual problems in binary SVM for the cases below:

a) hard margin b) soft margin