CS 214: Artificial Intelligence Lab

Common Instructions: You have to submit a code without any syntax errors. Don't copy from others; write your own piece of code with comments as much as possible. Try to Run with the test cases before submitting the code.

Assignment 5

NOTE:

- Test your code and submit it on Moodle (`https://moodle.iitdh.ac.in/user/index.php?id=992`).
- Deadline for submission is **Thursday 7 March by (11:59 PM)**
- We will run a plagiarism check for all the submissions; if found, a penalty will be applied

Lab Instructions:

1.
   - Generate 50 linearly spaced samples between 0 to 10 and add noise to each of these samples from a Gaussian distribution scaled by a factor of 0.5 (effectively variance = 0.25 mean 0). Assuming that the true weights of each of these samples are 2.5 and the bias is 1, generate a Y vector of outputs

```
np.random.seed(0)
num_samples = 50
X = np.linspace(0, 10, num_samples)
true_weights = np.array([2.5])
noise = 0.5 * np.random.randn(num_samples)
# ading some noise so the x to y relationship isn't perfectly linear
y = X * true_weights + 1 + noise
# true weight is w = [2.5] and bias term is 1 (effectively a 2-dim w = [1, 2.5])

# Add a column of ones to X for the bias term
X_bias = np.c_[np.ones(X.shape[0]), X]
```

   - Construct a ridge regression based solution mapping the X and Y terms using the closed form equation. (lambda = 0.01)
   - Construct a ridge regression based solution using the gradient descent based solution. (lambda = 0.01)
   - Plot the generated data against the line generated from the closed form solution and the gradient descent based solution in separate plots.
     For the gradient descent plot also plot the initial line.
   - Sample 80 values from a uniform distribution scaled by a factor of 5 for X. Apply sin to these values and add noise to them sampled from a gaussian distribution with scaling factor 0.5
   - Implement Ridge regression with Polynomial feature transformation for non linear case and plot the true result and the result from ridge regression.

2. Consider that you are collecting medical data for a study. You have collected data about a set of patients, all of whom suffered from the same illness. During their course of treatment, each patient responded to one of 5 medications: Drug A, Drug B, Drug C, Drug X, and Y (data is provided in the drug200.csv file).
   - Your job is to build a model to find out which drug might be appropriate for a future patient with the same illness. The features of this dataset are the Age, Sex, Blood Pressure, and Cholesterol of the patients, and the target is the drug that each patient responded to.

- It is a sample of a multi-class classifier, and you can use the training part of the dataset to build a decision tree and then use it to predict the class of an unknown patient or to prescribe a drug to a new patient.
- Convert all the values in the "drug" attribute to 0(drugA), 1(drugB), 2(drugC), 3(drugX), and 4(drugY). Implement the standard decision tree algorithm discussed in class, using information gain to choose which attribute to split at each point.
- Can you use logistic regression to perform the same task? Justify with the implementation.