

|  |  |  |
| --- | --- | --- |
| **Reinforcement Learning** | | |
| Lab Manual | | |
| **Department of Computer Science and Engineering**  **The NorthCap University, Gurugram** | | |
|  | |
|  | |
|  | |
|  | |
|  |  | | |

**Reinforcement Learning**

**Laboratory Manual**

**CSL348**

**Dr. Srishti Sharma**

**Ms. Monika Lamba**



Department of Computer Science and Engineering

The NorthCap University, Gurugram- 122001, India

Session 2022-23

*Published by:*

**Department of Computer Science & Engineering**

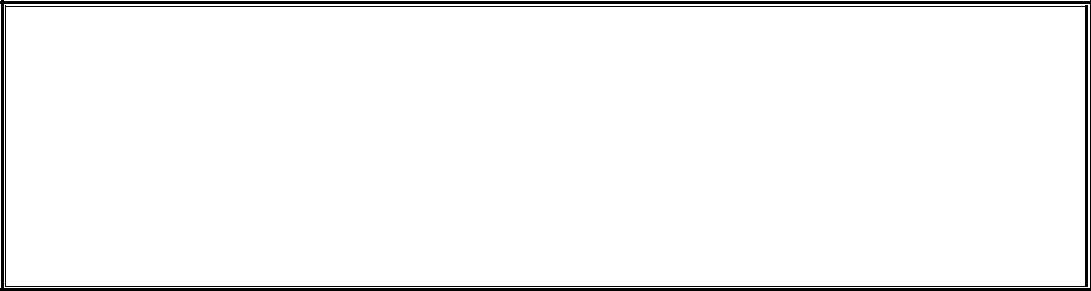
**School of Engineering and Technology**

**The NorthCap University Gurugram**

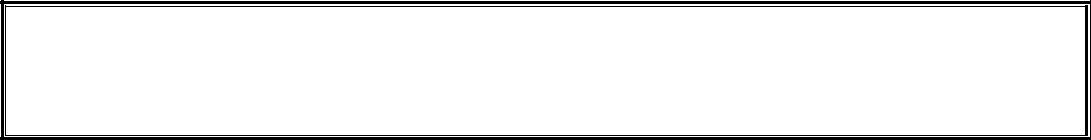
**• Laboratory Manual is for Internal Circulation only**

**© Authors**

*No part of this Practical Record Book may be reproduced, used, stored without prior permission of The NorthCap University*



Copying or facilitating copying of lab work comes under cheating and is considered as use of unfair means. Students indulging in copying or facilitating copying shall be awarded zero marks for that particular experiment. Frequent cases of copying may lead to disciplinary action. Attendance in lab classes is mandatory.



Labs are open up to 7 PM upon request. Students are encouraged to make full use of labs beyond normal lab hours.

**PREFACE**

Applied Computational Statistics Laboratory Manual is designed to meet the course and program requirements of NCU curriculum for B.Tech. fourth semester students of CSE branch. The concept of the lab work is to give brief practical experience for basic lab skills to students. It provides the space and scope for self-study so that students can come up with new and creative ideas.

The Lab manual is written on the basis of “teach yourself pattern” and expected that students who come with proper preparation should be able to perform the experiments without any difficulty. A brief introduction to each experiment with information about self-study material is provided. The laboratory exercises will help students to provide a hands-on each exercise that will help them to understand thoroughly. The students are expected to come thoroughly prepared for the lab. General disciplines, safety guidelines and report writing are also discussed.

The lab manual is a part of curriculum for the The NorthCap University, Gurugram. Teacher’s copy of the experimental results and answer for the questions are available as sample guidelines.

We hope that lab manual would be useful to students of CSE branch and author requests the readers to kindly forward their suggestions / constructive criticism for further improvement of the work book.

Author expresses deep gratitude to Members, Governing Body-NCU for encouragement and motivation.

**Authors**

**The NorthCap University**

**Gurugram, India**

|  |  |  |  |
| --- | --- | --- | --- |
| **CONTENTS** | | | |
| S.No. | **Details** | Page No. | |
|  |  |  | |
| 1 | **Introduction** | vi | |
| 2 | **Lab Requirement** | vii |
| 3 | **General Instructions** | viii |
| 4 | **List of Experiments** | x |
| 5 | **List of Flip Experiments** | xi |
| 6 | **List of Projects** | xii |
| 7 | **Rubrics** | xiv |
| 8 | **Annexure 1 (Format of Lab Report)** | xv |
| 9 | **Annexure 2 (Format of Project Report)** | xlviii |

1. **INTRODUCTION**



That ‘learning is a continuous process’ cannot be over emphasized. The theoretical knowledge gained during lecture sessions need to be strengthened through practical experimentation. Thus, practical makes an integral part of a learning process. ­­­­­­­­­­­­­­­­­­­­­

**COURSE OBJECTIVES:**

1. **Understand the basics of descriptive and inferential statistics and be able to apply appropriate descriptive statistical and exploratory methods to analyze datasets.**
2. **Recognize the concept & need of probability in real world. Students will understand the basics of probability, sample space, events, statistics and apply them to real life problems to determine marginal, conditional and joint probabilities.**
3. **Understand the probability mass function and distinguish between the different discrete distributions through application on real-world examples.**
4. **Understand the probability density function and distinguish between the different continuous distributions through application on real-world examples.**
5. **Identify the need for statistical hypothesis testing. Apply the appropriate hypothesis test, interpret the results and devise appropriate strategies.**
6. **Translate real world problems into probability models using Bayesian statistics.**
7. **LAB REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Requirements** | **Details** |
| **1** | **Software Requirements** | Python 3.x, Numpy, Pandas, Matplotlib, Seaborn, statistics, sci-kit learn |
| **2** | **Operating System** | Windows 7 onwards or Linux (32 or 64 bit) |
| **3** | **Hardware Requirements** | 4 GB RAM (Recommended)  2.60 GHz (Recommended) |
| **4** | **Required Bandwidth** | NA |

1. **GENERAL INSTRUCTIONS** 
   1. **General discipline in the lab**
   * Students must turn up in time and contact concerned faculty for the experiment they are supposed to perform.
   * Students will not be allowed to enter late in the lab.
   * Students will not leave the class till the period is over.
   * Students should come prepared for their experiment.
   * Experimental results should be entered in the lab report format and certified/signed by concerned faculty/ lab Instructor.
   * Students must get the connection of the hardware setup verified before switching on the power supply.
   * Students should maintain silence while performing the experiments. If any necessity arises for discussion amongst them, they should discuss with a very low pitch without disturbing the adjacent groups.
   * Violating the above code of conduct may attract disciplinary action.
   * Damaging lab equipment or removing any component from the lab may invite penalties and strict disciplinary action.
   1. **Attendance**

* Attendance in the lab class is compulsory.
* Students should not attend a different lab group/section other than the one assigned at the beginning of the session.
* On account of illness or some family problems, if a student misses his/her lab classes, he/she may be assigned a different group to make up the losses in consultation with the concerned faculty / lab instructor. Or he/she may work in the lab during spare/extra hours to complete the experiment. No attendance will be granted for such case**.**
  1. **Preparation and Performance**
* Students should come to the lab thoroughly prepared on the experiments they are assigned to perform on that day. Brief introduction to each experiment with information about self -study reference is provided on LMS.
* Students must bring the lab report during each practical class with written records of the last experiments performed complete in all respect.
* Each student is required to write a complete report of the experiment he has performed and bring to lab class for evaluation in the next working lab. Sufficient space in work book is provided for independent writing of theory, observation, calculation and conclusion.
* Students should follow the Zero tolerance policy for copying / plagiarism. Zero marks will be awarded if found copied. If caught further, it will lead to disciplinary action.
* Refer **Annexure 1** for Lab Report Format

1. **LIST OF EXPERIMENTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Title of the Experiment** | **Software**  **used** | **Unit**  **Covered** | **CO**  **Covered** | **Time**  **Required** |
|  | Implement Probability using Python | Python | 1 | CO1 | 2 Hours |
|  | Write a Python Program to compute Karl Pearson and Spearman’s Rank Correlation Coefficient | Python  (Jupyter) | 1 | CO1 | 2 hours |
|  | Write a Python program to solve the Multi-Armed Bandit problem using the Upper Confidence Bound Algorithm. Compare the reward obtained with random sampling. | Python  (Jupyter) | 2 | CO2 | 2 hours |
|  | Write a Python program to solve Multi-Armed Bandit problem using Thompson sampling. | Python  (Jupyter) | 2 | CO2 | 2 hours |
|  | Write a program to implement Q-Learning in Python. | Python  (Jupyter) | 3 | CO3 | 2 hours |
|  | Write python program to implement Markov Decision Process. | Python  (Jupyter) | 3 | CO3 | 2 hours |
|  | Write a python program to implement policy iteration in Dynamic programming. | Python  (Jupyter) | 4 | CO4 | 2 hours |
|  | Write a python program to implement value iteration in Dynamic programming. | Python  (Jupyter) | 4 | CO4 | 2 hours |
|  | Write a Python Program to implement  Monte Carlo method | Python  (Jupyter) | 5 | CO5 | 2 hours |
|  | Write a Python Program to implement  TD and SARSA in Reinforcement Learning | Python  (Jupyter) | 5 | CO5 | 2 hours |
|  | Implement function approximation methods | Python  (Jupyter) | 6 | CO6 | 2 hours |
|  | Implement function approximation methods. | Python  (Jupyter) | 6 | CO6 | 2 hours |
| **Value Added Experiments** | | | | | |
|  | Use RL algorithms to solve CartPole Balancing | Python  (Jupyter) | 5 | CO2,3,4,5,6 | 2 hours |
|  | Create Deep Reinforcement Learning Algorithms to play Atari games. | Python  (Jupyter) | 5 | CO2,3,4,5,6 | 2 hours |
|  | Implement Q-Learning and Markov Algorithms with Python and OpenAI | Python  (Jupyter) | 5 | CO3 | 2 hours |

1. **LIST OF FLIP EXPERIMENTS**

|  |  |  |
| --- | --- | --- |
| **Exp. No.** | **Title of the Experiment** | **Mapped CO** |
|  | Apply advanced deep RL algorithms to games  such as Minecraft | CO 1, 2,3,4,5,6 |
|  | Deploy RL algorithms using OpenAI Universe | CO1,2,3,4,5,6 |
|  | Implement basic actor-critic algorithms for  continuous control | CO1,2,3,4,5, 6 |

1. **LIST OF PROJECTS**

|  |  |  |
| --- | --- | --- |
| **Sr No.** | **Project Title** | **Mapped CO** |
|  | Traffic Light Control | CO 1,2,3,4,5,6 |
|  | Robotics | CO1,2,3,4,5,6 |
|  | News Recommendation System. | CO1,2,3,4,5,6 |

1. **RUBRICS (Only for Lab components)**

|  |  |
| --- | --- |
| **Marks Distribution** | |
| **Continuous Evaluation (25 Marks)** | **Project Evaluations (20 Marks)** |
| Each experiment shall be evaluated for 5 marks and at the end of the semester proportional marks shall be awarded out of total 25. | Project shall be evaluated for 20 marks and at the end of the semester viva will be conducted related to the project. |
| **Viva and Reporting (25 Marks)**  Following is the breakup of 25 marks for each  **10 Marks**: Observation & conduct of experiment. Teacher may ask questions about experiment in mid-term viva.  **10 Marks:** Observation & conduct of experiment.  **5 Marks:** For report writing |

**Annexure 1**

**CSL348**

Lab Practical Report



Faculty name **Ms. Monika Lamba** Student name: Mukul

Roll No.: 20CSU355

Semester: CSE-V-AB

Group:

Department of Computer Science and Engineering

The NorthCap University, Gurugram- 122001, India

Session 2022-2023

**INDEX**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No.** | **Experiment** | **Page No.** | **Date of Experiment** | **Date of Submission** | **Marks** | **Signature** |
| **1** | Implement Probability using Python. |  |  |  |  |  |
| **2** | Write a Python Program to compute Karl Pearson and Spearman’s Rank Correlation Coefficient. |  |  |  |  |  |
| **3** | Write a Python program to solve the Multi-Armed Bandit problem using the Upper Confidence Bound Algorithm. Compare the reward obtained with random sampling. |  |  |  |  |  |
| **4** | Write a Python program to solve Multi-Armed Bandit problem using Thompson sampling. |  |  |  |  |  |
| **5** | Write a program to implement Q-Learning in Python. |  |  |  |  |  |
| **6** |  |  |  |  |  |  |
| **7** |  |  |  |  |  |  |
| **8** |  |  |  |  |  |  |
| **9** |  |  |  |  |  |  |
| **10** |  |  |  |  |  |  |
| **11** |  |  |  |  |  |  |
| **12** |  |  |  |  |  |  |
| **13** |  |  |  |  |  |  |
| **14** |  |  |  |  |  |  |
| **15** |  |  |  |  |  |  |
| **16** |  |  |  |  |  |  |

**EXPERIMENT NO. 1**

|  |
| --- |
| **Student Name and Roll Number:Mukul 20CSU355** |
| **Semester /Section: CSE-V-AB** |
| **Link to Code:** |
| **Date:** |
| **Faculty Signature:** |
| **Marks/Grade:** |

|  |
| --- |
| **Objective(s):**   * Familiarization with probability |
| **Outcome:** Revision of the concepts of probability and probability distributions and implementing the same using Python. |
| **Problem Statement:** Implement Probability using Python |
| **Background Study:**  Python has libraries like Statistics and SciPy. Statistics which contain functions for several descriptive and inferential statistics tasks which can be of help to the students. |
| **Question Bank:**   1. **What is difference between discrete and continuous probability distributions?**   Ans: A discrete distribution is one in which the data can only take on certain values, for example integers. A continuous distribution is one in which data can take on any value within a specified range (which may be infinite).   1. **Enlist some discrete probability distributions.**   Ans: The most common discrete probability distributions include binomial, Poisson, Bernoulli, and multinomial.   1. **Enlist some continuous probability distributions.**   Ans: Beta distribution,  Normal Distribution.  Cauchy distribution,  Exponential distribution,  Gamma distribution,  Logistic distribution,  Weibull distribution. |

**Student Work Area**

**Algorithm/Flowchart/Code/Sample Outputs**

**Graphical user interface, application, Word

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated**

**EXPERIMENT NO. 2**

|  |
| --- |
| **Student Name and Roll Number: Mukul 20CSU355** |
| **Semester /Section: CSE-V-AB** |
| **Link to Code:** |
| **Date:** |
| **Faculty Signature:** |
| **Marks/Grade:** |

|  |
| --- |
| **Objective(s):**  Compute correlation for two given series |
| **Outcome:** Understanding the meaning of correlation |
| **Problem Statement:** Compute Karl Pearson’s and Spearman’s Rank Correlation |
| **Background Study:** In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data. Although in the broadest sense, "correlation" may indicate any type of association, in statistics it normally refers to the degree to which a pair of variables are linearly related. |
| **Question Bank:**   1. **Differentiate between correlation and causation.**   Ans: A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable. Causation indicates that one event is the result of the occurrence of the other event, i.e., there is a causal relationship between the two events.   1. **How to compute Spearman’s rank correlation coefficient for repeated ranks.**   Ans: For example, if two individuals are placed in the 8th place, they are given the rank [8+9] / 2 = 8.5 each, which is common rank to be assigned and the next will be 10; and if three ranked equal at the 8th place, they are given the rank [8 + 9 +10] /3 = 9 which is the common rank to be assigned to each.   1. **Elucidate on the graphical method for estimating correlation.**   Ans: If the dots of the two series are advancing in a definite direction like a current, this condition indicates that the data series are correlated.  Positive and Negative Correlation  (ii) When the arrays of dots advance from left to right in upward direction, the correlation is positive.  (iii) When the scatter diagram advances from left to right in downward direction, the correlation’ is negative.  (iv) When the dots are not in definite arrays and are scattered haphazardly, this condition indicates that there is no correlation between the data series.  Perfect Negative Correlation, No Correlation and Perfect Positive Correlation  (v) When the dots appear to be situated on a line which advances upward at 45° angle from the O-X axis, this condition indicates perfect positive correlation among the data series.  (vi) If the dots appear to be situated on a line which moves from left to right in downward direction at 45° angle from 0-X axis, this condition is indicative of perfect negative correlation. |

**Student Work Area**

**Algorithm/Flowchart/Code/Sample Outputs**

**Graphical user interface, text, application, email

Description automatically generated**

**Chart, scatter chart

Description automatically generated**

**Chart, scatter chart

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Chart

Description automatically generated**

**A picture containing timeline

Description automatically generated**

**A picture containing timeline

Description automatically generated**

**Chart

Description automatically generated**

**A picture containing chart

Description automatically generated**

**EXPERIMENT NO. 3**

|  |
| --- |
| **Student Name and Roll Number: Mukul 20CSU355** |
| **Semester /Section: CSE-V-AB** |
| **Link to Code:** |
| **Date:** |
| **Faculty Signature:** |
| **Marks/Grade:** |

|  |
| --- |
| **Objective(s):**  Solve the Multi-Armed Bandit Problem |
| **Outcome:** Understanding and comparing bandit strategies. |
| **Problem Statement:** Solve the muti-armed bandit problem using the Upper Confidence Bound Algorithm. Compare the reward obtained with random sampling. |
| **Background Study:** In probability theory and machine learning, the **multi-armed bandit problem** (sometimes called the ***K* or *N*-armed bandit problem** is a problem in which a fixed limited set of resources must be allocated between competing (alternative) choices in a way that maximizes their expected gain, when each choice's properties are only partially known at the time of allocation, and may become better understood as time passes or by allocating resources to the choice. This is a classic reinforcement learning problem that exemplifies the exploration–exploitation tradeoff dilemma. |
| **Question Bank:**   1. **Differentiate between exploration and exploitation.**   Ans: Exploration involves activities such as search, variation, risk taking, experimentation, discovery, and innovation. Exploitation involves activities such as refinement, efficiency, selection, implementation, and execution.   1. **Differentiate between greedy and epsilon greedy strategies for solving multi-armed bandit problem.**   Ans: Greedy-Action selection always exploits current knowledge to maximize immediate reward.  Greedy-epsilon tries to converge Qt () to q\*(a) by exploring with the probability of e, and exploiting with probability of 1-e.   1. **Explain the Upper Confidence Bound Algorithm for solving multi-armed bandit problem.**   Ans: Upper-Confidence Bound action selection uses uncertainty in the action-value estimates for balancing exploration and exploitation. Since there is inherent uncertainty in the accuracy of the action-value estimates when we use a sampled set of rewards thus UCB uses uncertainty in the estimates to drive exploration. |

**Student Work Area**

**Algorithm/Flowchart/Code/Sample Outputs**

**A picture containing graphical user interface

Description automatically generated**

**Text

Description automatically generated**

**Text, application

Description automatically generated**

**Chart, histogram

Description automatically generated**

**EXPERIMENT NO. 4**

|  |
| --- |
| **Student Name and Roll Number: Mukul 20CSU355** |
| **Semester /Section: CSE-V-AB** |
| **Link to Code:** |
| **Date:** |
| **Faculty Signature:** |
| **Marks/Grade:** |

|  |
| --- |
| **Objective(s):** Solve the Multi-Armed Bandit Problem. |
| **Outcome:** Understand Thompson sampling as a solution to the Multi-Armed Bandit Problem. |
| **Problem Statement:** Write a python program to solve the Multi-Armed Bandit Problem using Thompson Sampling. |
| **Background Study:** Thompson sampling, named after William R. Thompson, is a heuristic for choosing actions that addresses the exploration-exploitation dilemma in the multi-armed bandit problem. It consists of choosing the action that maximizes the expected reward with respect to a randomly drawn belief. |
| **Question Bank:**   1. **What are beta distributions and why are they used for Thompson sampling?**   Ans: The beta distribution is a family of continuous probability distributions defined on the interval [0, 1] parameterized by two positive shape parameters, denoted by alpha and beta, that appear as exponents of the random variable and control the shape of the distribution.  **Thompson Sampling** generates a model of the reward probabilities. When, as in this case, the available rewards are binary (win or lose, yes or no, charge or no charge) then the Beta distribution is ideal to model this type of probability.   1. **Compare and contrast Thompson sampling with other bandit strategies.**   Ans: Thompson Sampling takes a different approach to these other methods. Instead of simply maintaining an estimate of the reward, it gradually refines a model of the probability of the reward for each action and actions are chosen by sampling from this distribution.   1. **Why is Thompson sampling referred to as Bayesian Bandits?**   Ans: Thompson Sampling (also sometimes referred to as the Bayesian Bandits algorithm) takes a slightly different approach; rather than just refining an estimate of the mean reward it extends this, to instead build up a probability model from the obtained rewards, and then samples from this to choose an action. |

**Student Work Area**

**Algorithm/Flowchart/Code/Sample Outputs**

**Graphical user interface, text, application

Description automatically generated**

**Chart

Description automatically generated**

**EXPERIMENT NO. 5**

|  |
| --- |
| **Student Name and Roll Number: Mukul 20CSU355** |
| **Semester /Section: CSE-V-AB** |
| **Link to Code:** |
| **Date:** |
| **Faculty Signature:** |
| **Marks/Grade:** |

|  |
| --- |
| **Objective(s):** Write python program to implement Q-Learning |
| **Outcome(s):** To understand Q-Learning |
| **Problem Statement:** Implement Q-Learning using Python |
| **Background Study:** Q-learning is a model-free reinforcement learning algorithm to learn the value of an action in a particular state. It does not require a model of the environment (hence "model-free"), and it can handle problems with stochastic transitions and rewards without requiring adaptations.  For any finite Markov decision process (FMDP), *Q*-learning finds an optimal policy in the sense of maximizing the expected value of the total reward over any and all successive steps, starting from the current state. *Q*-learning can identify an optimal action-selection policy for any given FMDP, given infinite exploration time and a partly random policy. "Q" refers to the function that the algorithm computes – the expected rewards for an action taken in a given state. |
| **Question Bank:**   1. **Differentiate between policy based and value-based reinforcement learning.**   Ans: **Value-Based:** learn the state or state-action value. Act by choosing the best action in the state. Exploration is necessary.  **Policy-Based:** Directly learn the stochastic policy function that maps state to action. Act by sampling policy.   1. **What are off-policy and on-policy learners?**   Ans: An off-policy learner learns the value of the optimal policy independently of the agent's actions. Q-learning is an off-policy learner. An on-policy learner learns the value of the policy being carried out by the agent including the exploration steps.   1. **What is the Bellman equation?**   Ans: Bellman equation decomposes the value function into two parts, the immediate reward plus the discounted future values.   1. **What will be the effect(s) of changing the learning rate in Q-Learning?**   Ans: Learning rate controls how fast we modify our estimates, on changing the learning rate our speed of estimating the estimates will vary but overall, we won’t have any change on the final estimates. |

**Student Work Area**

**Algorithm/Flowchart/Code/Sample Outputs**

**Calendar

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**Graphical user interface, text

Description automatically generated with medium confidence**

**Text

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Annexure 2**

**Applied Computational Statistics**

**Project Report**



Faculty name Student name : Mukul

Roll No.: 20 csu 355

Semester: 5

Group:AI-B

Department of Computer Science and Engineering

The NorthCap University, Gurugram- 122001, India

Session 2021-2022

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S.No** |  | **Page No.** |
| **1.** | **Project Description** |  |
| **2.** | **Problem Statement** |  |
| **3.** | **Analysis**  **3.1 Hardware Requirements**  **3.2 Software Requirements** |  |
| **4.** | **Design**  **4.1 Data/Input Output Description:**  **4.2 Algorithmic Approach / Algorithm / DFD / ER diagram/Program Steps** |  |
| **5.** | **Implementation and Testing (stage/module wise)** |  |
| **6.** | **Output (Screenshots)** |  |
| **7.** | **Conclusion and Future Scope** |  |