**ASSIGNMENT 1**

Q1. WHAT DOES ONE MEAN BY THE TERM "MACHINE LEARNING"?

ANS. MACHINE LEARNING IS DEFINED BY PROGRAMMING MACHINE TO LEARN FROM DATA TO SOLVE A SPECIFIC PROBLEM.

Q2. CAN YOU THINK OF 4 DISTINCT TYPES OF ISSUES WHERE IT SHINES?

ANS. MACHINE LEARNING IS GREAT FOR PROBLEMS WHOSE SOLUTION REQUIRES A GREAT DEAL OF WORK OR A LONG LIST OF RULES, COMPLEX PROBLEMS THAT ARE HARD TO GET A SOLUTION OF USING A TRADITIONAL METHOD, FLUCTUATING ENVIRONMENTS AND GETTING INSIGHTS ABOUT COMPLEX PROBLEMS AND LARGE DATA.

Q3. WHAT IS A LABELED TRAINING SET, AND HOW DOES IT WORK?

ANS. LABELED TRAINING SET IS A SET OF TRAINING DATA WHICH HAS A SOLUTION TO THE PROBLEM OR TASK. THE TRAINING SET IS USED TO TRAIN THE ALGORITHM, AND THEN YOU USE THE TRAINED MODEL ON THE TEST SET TO PREDICT THE RESPONSE VARIABLE VALUES THAT ARE ALREADY KNOWN. THE FINAL STEP IS TO COMPARE THE PREDICTED RESPONSES AGAINST THE ACTUAL (OBSERVED) RESPONSES TO SEE HOW CLOSE THEY ARE.

Q4.WHAT ARE THE TWO MOST IMPORTANT TASKS THAT ARE SUPERVISED?

ANS. TWO MOST COMMON SUPERVISED TASKS ARE CLASSIFICATION AND REGRESSION.

Q 5.CAN YOU THINK OF FOUR EXAMPLES OF UNSUPERVISED TASKS?

ANS. FOUR COMMON UNSUPERVISED TASKS INCLUSED CLUSTERING, VISUALIZATION, DIMENSIONALITY REDUCTION , AND ASSOCIATION RULE LEARNING.

Q6. STATE THE MACHINE LEARNING MODEL THAT WOULD BE BEST TO MAKE A ROBOT WALK THROUGH VARIOUS UNFAMILIAR TERRAINS?

ANS. THE BEST MACHINE LEARNING ALGORITHM TO ALLOW A ROBOT TO WALK IN UNKNOWN TERRAIN IS REINFORCED LEARNING, WHERE THE ROBOT CAN LEARN FROM RESPONSE OF THE TERRAIN TO OPTIMIZE ITSELF.

Q7. WHICH ALGORITHM WILL YOU USE TO DIVIDE YOUR CUSTOMERS INTO DIFFERENT GROUPS?

ANS. CLUSTERING ALGORITHM . IT IS AN ALGORITHM USED IN CUSTOMER SEGMENTATION WHICH IS A PROCESS OF CLASSIFYING SIMILAR CUSTOMERS INTO SAME SEGMENT

Q8. WILL YOU CONSIDER THE PROBLEM OF SPAM DETECTION TO BE A SUPERVISED OR UNSUPERVISED LEARNING PROBLEM?

ANS. SPAM DETECTION IS A SUPERVISED LEARNING PROBLEM.

Q9. WHAT IS THE CONCEPT OF AN ONLINE LEARNING SYSTEM?

ANS. ONLINE LEARNING SYSTEM IS A LEARNING SYSTEM IN WHICH THE MACHINE LEARNS AS DATA IS GIVEN IN SMALL STREAMS CONTINUOUSLY.

Q10. WHAT IS OUT-OF-CORE LEARNING, AND HOW DOES IT DIFFER FROM CORE LEARNING?

ANS. OUT-OF-CORE LEARNING SYSTEM IS A SYSTEM THAT CAN HANDLE DATA THAT CANNOT FIT INTO YOUR COMPUTER MEMORY. IT USES ONLINE LEARNING SYSTEM TO FEED DATA IN SMALL BITS.

Q11. WHAT KIND OF LEARNING ALGORITHM MAKES PREDICTIONS USING A SIMILARITY MEASURE?

ANS. INSTANCE BASED ALGORITHM.

Q.12. WHAT'S THE DIFFERENCE BETWEEN A MODEL PARAMETER AND A HYPERPARAMETER IN A LEARNING ALGORITHM?

ANS.

|  |  |
| --- | --- |
| **MODEL PARAMETER** | **HYPERPARAMETER** |
| MODEL PARAMETERS ARE THE CONFIGURATION MODEL, WHICH ARE INTERNAL TO THE MODEL. | HYPERPARAMETERS ARE THE EXPLICITLY SPECIFIED PARAMETERS THAT CONTROL THE TRAINING PROCESS. |
| PARAMETERS ARE ESSENTIAL FROM= MAKING PREDICTIONS | HYPERPARAMETERS ARE ESSENTIAL FOR OPTIMIZING THE MODEL |
| THESE ARE SPECIFIED OR ESTIMATED WHILE TRAINING THE MODEL | THESE ARE SET BEFORE THE BEGINNING OF THE TRAINING F THE MODEL. |
| IT IS INTERNAL TO THE MODEL | IT IS EXTERNAL TO THE MODEL |
| IT IS LEARNED AND SET BY THE MODEL BY ITSELF | IT IS SET MANUALLY BY THE MACHINE LEARNING ENGINEER |

Q13.WHAT ARE THE CRITERIA THAT MODEL-BASED LEARNING ALGORITHMS LOOK FOR? WHAT IS THE MOST POPULAR METHOD THEY USE TO ACHIEVE SUCCESS? WHAT METHOD DO THEY USE TO MAKE PREDICTIONS?

ANS. MODEL BASED LEARNING ALGORITHM SEARCH FOR **THE OPTIMAL VALUE OF PARAMETERS IN A MODEL THAT WILL GIVE THE BEST RESULTS FOR THE NEW INSTANCES**. WE OFTEN USE A COST FUNCTION OR SIMILAR TO DETERMINE WHAT THE PARAMETER VALUE HAS TO BE IN ORDER TO MINIMIZE THE FUNCTION. THE MODEL MAKES PREDICTION BY USING THE VALUE OF THE NEW INSTANCE AND THE PARAMETERS IN ITS FUNCTION.

Q14. CAN YOU NAME FOUR OF THE MOST IMPORTANT MACHINE LEARNING CHALLENGES?

ANS. FOUR MAIN CHALLENGES IN MACHINE LEARNING INCLUDE OVERFITTING THE DATA (USING A MODEL TOO COMPLICATED), UNDERFITTING THE DATA (USING A SIMPLE MODEL), LACKING IN DATA AND NONREPRESENTATIVE DATA.

Q15. WHAT HAPPENS IF THE MODEL PERFORMS WELL ON THE TRAINING DATA BUT FAILS TO GENERALIZE THE RESULTS TO NEW SITUATIONS? CAN YOU THINK OF THREE DIFFERENT OPTIONS?

ANS. IF THE MODEL PERFORMS POORLY TO NEW INSTANCES, THEN IT HAS OVERFIT PROBLEM ON THE TRAINING DATA. TO SOLVE THIS, WE CAN DO ANY OF THE FOLLOWING THREE: GET MORE DATA, IMPLEMENT A SIMPLER MODEL, OR ELIMINATE OUTLIERS OR NOISE FROM THE EXISTING DATA SET.

Q16.WHAT EXACTLY IS A TEST SET, AND WHY WOULD YOU NEED ONE?

*ANS. TEST SET IS A SET THAT YOU TEST YOUR MODEL (FIT USING TRAINING DATA) TO SEE HOW IT PERFORMS. TEST SET IS NECESSARY SO THAT YOU CAN DETERMINE HOW GOOD (OR BAD) YOUR MODEL PERFORMS.*

17.WHAT IS A VALIDATION SET'S PURPOSE?

ANS. VALIDATION SET IS A SET USED TO COMPARE BETWEEN DIFFERENT TRAINING MODELS.

Q18. WHAT PRECISELY IS THE TRAIN-DEV KIT, WHEN WILL YOU NEED IT, HOW DO YOU PUT IT TO USE?

Ans. A TRAIN DEV KIT IS A COLLECTION OF TOOLS , LIBRARIES OR RESOURCES THAT ASSIST IN BOTH THE TRAINING AND DEVELOPMENT STAGES OF MACHINE LEARNING PROJECTS. WE NEED TRAIN DEV KIT WHEN WORKING ON DEVELOPING AND REFINING MACHINE LEARNING MODELS. TO PUT A TRAIN DEV KIT TO USE BY FOLLOWING THESE STEPS INSTALLATION , MODEL TRAINING , EVALUATION & ANALYSIS AND ITERATIVE DEVELOPMENT

Q19. WHAT COULD GO WRONG IF YOU USE THE TEST SET TO TUNE HYPERPARAMETERS?

ANS. IF YOU TUNE HYPERPARAMETERS USING THE TEST SETS, THEN IT MAY NOT PERFORM WELL ON THE OUT-OF-SAMPLE DATA BECAUSE THE MODEL IS TUNED JUST FOR THAT SPECIFIC SET.