**Question 1**

A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for P?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| None of these. |  |  |  |
| The probability of it correctly predicting a future date's weather. |  |  |  |
| The weather prediction task. |  |  |  |
| The process of the algorithm examining a large amount of historical weather data. | Inorrect | 0.00 | It is by examining the historical weather data that the learning algorithm improves it's performance, so this is the experience E. |
| Total |  | 0.00 / 1.00 |  |

**Question 2**

Suppose you are working on weather prediction, and use a learning algorithm to predict tomorrow's temperature (in degrees Centigrade/Fahrenheit). Would you treat this as a classification or a regression problem?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Classification |  |  |  |
| Regression | Correct | 1.00 | Regression is appropriate when we are trying to predict a continuous-valued output, such as the temperature. |
| Total |  | 1.00 / 1.00 |  |

**Question 3**

Suppose you are working on stock market prediction, Typically tens of millions of shares of Microsoft stock are traded (i.e., bought/sold) each day. You would like to predict the number of Microsoft shares that will be traded tomorrow. Would you treat this as a classification or a regression problem?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Classification |  |  |  |
| Regression | Correct | 1.00 | Regression is appropriate when we are trying to predict a continuous-valued output. Even though the number of shares traded is technically discrete-valued (i.e., an integer), it would be more natural to treat it as a continuous value (similar to the housing prices example in the lecture). |
| Total |  | 1.00 / 1.00 |  |

**Question 4**

Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. Which of the following would you apply supervised learning to? (Select all that apply.) In each case, assume some appropriate dataset is available for your algorithm to learn from.

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number of groups of essays that are somehow "similar" or "related". | Correct | 0.25 | This is an unsupervised learning/clustering problem (similar to the Google News example in the lectures). |
| Given data on how 1000 medical patients respond to an experimental drug (such as effectiveness of the treatment, side effects, etc.), discover whether there are different categories or "types" of patients in terms of how they respond to the drug, and if so what these categories are. | Correct | 0.25 | This can be addressed using an unsupervised learning, clustering, algorithm, in which we group the 1000 patients into different clusters based on their responses to the drug. |
| Given historical data of childrens' ages and heights, predict children's height as a function of their age. | Correct | 0.25 | This is a supervised learning, regression problem, where we can learn from a training set to predict height. |
| Have a computer examine an audio clip of a piece of music, and classify whether or not there are vocals (i.e., a human voice singing) in that audio clip, or if it is a clip of only musical instruments (and no vocals). | Correct | 0.25 | This can be addressed using supervised learning, in which we learn from a training set of audio clips which have been labeled as either having vocals or not. |
| Total |  | 1.00 / 1.00 |  |

**Question 5**

Which of these is a reasonable definition of machine learning?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Machine learning means from labeled data. |  |  |  |
| Machine learning is the field of allowing robots to act intelligently. |  |  |  |
| Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed. | Correct | 1.00 | This was the definition given by Arthur Samuel (who had written the famous checkers playing, learning program). |
| Machine learning is the science of programming computers. |  |  |  |
| Total |  | 1.00 / 1.00 |  |