**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. In this setting, what is T?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| The process of the algorithm examining a large amount of historical weather data. |  |  |  |
| None of these. |  |  |  |
| The weather prediction task. |  |  |  |
| The probability of it correctly predicting a future date's weather. | Inorrect | 0.00 | This would be a reasonable measure P of measuring our weather predictions' accuracy. |
| Total |  | 0.00 / 1.00 |  |

**Question 2**

Suppose you are working on weather prediction, and you would like to predict whether or not it will be raining at 5pm tomorrow. You want to use a learning algorithm for this. Would you treat this as a classification or a regression problem?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Regression | Inorrect | 0.00 | Regression is appropriate when we are trying to predict a continuous-valued output; but in this problem we are trying to predict one of 2 possible discrete-valued outputs (raining or not) |
| Classification |  |  |  |
| Total |  | 0.00 / 1.00 |  |

**Question 3**

Suppose you are working on stock market prediction. You would like to predict whether the US Dollar will go up against the Euro tomorrow (i.e., whether a dollar will be worth more euros tomorrow than it is worth today). Would you treat this as a classification or a regression problem?

|  |  |  |  |
| --- | --- | --- | --- |
| **Your Answer** |  | **Score** | **Explanation** |
| Classification |  |  |  |
| Regression | Inorrect | 0.00 | Regression is appropriate when we are trying to predict a continuous-valued output. Here, there are two possible outcomes: That the US Dollar goes up (which we might designate as class 0, say) or that it does not (class 1). |
| Total |  | 0.00 / 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** |
| Given genetic (DNA) data from a person, predict the odds of him/her developing diabetes over the next 10 years. | Correct | 0.25 | This can be addressed as a supervised learning, classification, problem, where we can learn from a labeled dataset comprising different people's genetic data, and labels telling us if they had developed diabetes. |
| 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. |
| 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. |
| 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 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 field of allowing robots to act intelligently. |  |  |  |
| Machine learning is the science of programming computers. |  |  |  |
| Total |  | 1.00 / 1.00 |  |