Algorithm and Terminology Quiz

1. In Supervised learning, you need to provide several examples input data and correct answer
    a. True
    b. False
2. In supervised learning, labeled data refers to training set and unlabeled data refers to evaluation set
    a. True
    b. False
3. Typical practice is to use all labeled data for training and reuse that data for evaluation
    a. True
    b. False
4. Evaluation Set is the data that is reserved for testing quality of trained model
    a. True
    b. False
5. Labeled data is easiest type of data to get
    a. True
    b. False
6. Movie training data has a feature 'Movie Type' that can contain any of these values: G – General Audiences, PG – Parental Guidance Suggested, PG-13 – Parents Strongly Cautioned, R – Restricted, NC-17 – Adults Only.
   What is an appropriate data type for this?
    a. Categorical
    b. Text
    c. Numeric
7. A training data has a field "Assigned_Seating" and the valid values are: Yes, No. What is the appropriate data type for this field?
    a. Categorical
    b. Text
    c. Binary
    d. Numeric
8. A field contain shoe sizes: 10,10.5,11,11.5 up to 14 in 0.5 increments. What data type is suitable for this?
    a. Categorical
    b. Text
    c. Binary
    d. Numeric
9. A training set consists of text columns. Does AWS Machine Learning models support text as input for training models?
    a. Yes
    b. No
10. We need to train a model that predicts any value that is between 0 and 1. What type of model is suitable for this problem?
    a. Linear Regression
    b. Binary Classifier

Answers

1. True – In supervised learning, you provide known correct answers to let machine learning algorithm learn the relationship between input features and target variable
2. False – Labeled data refers to data with known correct answers. Evaluation set is part of labeled data that is reserved for verifying predictive accuracy of a model using a data set that was not seen by ML Training algorithm before.
3. False - Evaluation data needs to be reserved and not used for training. Evaluation set is used to verify predictive quality of the algorithm
4. True
5. False - Labeling process can be time consuming and resource intensive
6. Categorical – These predefined set of categories
7. Binary - Since there are only two states, binary would be appropriate datatype
8. Categorical – These are fixed set of values. Categorical would be a better choice
9. Yes - Text support is one of the powerful features offered by AWS ML
10. Linear Regression - Note that the question says any value between 0 and 1; it does not say 0 and 1 are the only possible outputs