

HW1

Tim Demetriades

AAI 695 WS2 – Shucheng Yu

Question 1

Explain the following concepts:

1. Supervised Learning

Supervised learning is a type of machine learning system where the training set you feed to the algorithm includes the desired solutions (labels). Essentially you are dealing with labeled data. Classification is a common type of supervised learning. For example, given a set of cars, you have to figure out if they are cheap or expensive given a number of attributes. When training, the model is able to use the labels to make correlations between the attributes and the labels. Then, the model is used to predict the labels of new data (that does not have labels).

2. Unsupervised Learning

Unsupervised learning is a type of machine learning where the training set you feed to the algorithm does not include the desired solutions (labels). Essentially you are dealing with unlabeled data. The system is trying to learn the correlations of the data on its own. One type of unsupervised learning is clustering, such as K-means clustering. With these algorithms, the model tries to identify similarities in the data and group similar data together into clusters. Another form of unsupervised learning is PCA, which is used for dimensionality reduction.

3. Online Learning

Online learning is where you train a system incrementally by feeding it data instances sequentially, either individually or in small groups. This is useful because as new data comes in, the model can be adjusted without having to retrain the whole model with all of the data, saving time and resources. This makes it ideal for systems where there is a continuous flow of data, and the model may need to be constantly adjusted.

4. Batch Learning

Batch learning is the opposite of online learning. With batch learning, the system is incapable of learning incrementally. Instead, the model is trained with all of the data at once, which can take a lot of time and resources. As a result, this is generally done offline, hence the alternative name, offline learning. First the model is trained with all of the data, and then the model is applied and no more learning is done. If new data comes in, the model must be re-trained from scratch on the entire dataset. This is why online learning is the preferred method.

5. Model-Based Learning

Model-based learning is a way to generalize from a set of examples by building a model of these examples and then using that model to make predictions. For example, logistic regression is model-based learning. The learning happens at the training step by discovering patterns and correlations in the training data and it gives you weights to be able to make predictions on new data using these parameters.

6. Instance-Based Learning

Instance-based learning is another way to generalize from a set of examples. With instance-based learning, the system learns the examples by heart, then generalizes to new cases by using a similarity measure to compare them to the learned examples. For example, the K-nearest neighbors algorithm is a type of instance-based learning. To be able to classify a new data point, the algorithm calculates distances between the new data point and all the points in your dataset and predicts the class which the new data point belongs to – the class with the k nearest neighbors. Essentially, the learning happens at prediction time rather than training time.