/*

Author: Wang Jinge

Date: 2/19/2019

*/

1. [18 points] Explain the following concepts:

1) **supervised learning,**

Supervised learning is a method used to enable machines to classify objects, problems or situations based on related data fed into the machines. Machines are fed with data such as characteristics, patterns, dimensions, color and height of objects, people or situations repetitively until the machines are able to perform accurate classifications. Supervised learning is a popular technology or concept that is applied to real-life scenarios. Supervised learning is used to provide product recommendations, segment customers based on customer data, diagnose disease based on previous symptoms and perform many other tasks.

2) unsupervised learning,

Unsupervised learning is a method used to enable machines to classify both tangible and intangible objects without providing the machines any prior information about the objects. The things machines need to classify are varied, such as customer purchasing habits, behavioral patterns of bacteria and hacker attacks. The main idea behind unsupervised learning is to expose the machines to large volumes of varied data and allow it to learn and infer from the data. However, the machines must first be programmed to learn from data

3) online learning,

online machine learning is a method of machine learning in which data becomes available in a sequential order and is used to update our best predictor for future data at each step, as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set at once. Online learning is a common technique used in areas of machine learning where it is computationally infeasible to train over the entire dataset, requiring the need of out-of-core algorithms. It is also used in situations where it is necessary for the algorithm to dynamically adapt to new patterns in the data, or when the data itself is generated as a function of time, e.g. stock price prediction. Online learning algorithms may be prone to catastrophic interference, a problem that can be addressed by incremental learning approaches.

4) batch learning,

Batch learning refers to situations where the program is not operating and taking in new information in real time. Instead, it has a static set of input data. The opposite is online learning, where the machine learning program is working in real time on data that comes in.

5) model-based learning,

6) instance-based learning.

Instance-based learning (sometimes called memory-based learning) is a family of learning algorithms that, instead of performing explicit generalization, compares new problem instances with instances seen in training, which have been stored in memory.

It is called instance-based because it constructs hypotheses directly from the training instances themselves.This means that the hypothesis complexity can grow with the data:in the worst case, a hypothesis is a list of n training items and the computational complexity

of classifying a single new instance is O(n). One advantage that instance-based learning has over other methods of machine learning is its ability to adapt its model to previously unseen data. Instance-based learners may simply store a new instance or throw an old instance away.

2. [6 points] What is over-fitting of training data? What is regularization?

①Overfitting refers to a model that models the training data too well.

Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize.

Overfitting is more likely with nonparametric and nonlinear models that have more flexibility when learning a target function. As such, many nonparametric machine learning algorithms also include parameters or techniques to limit and constrain how much detail the model learns.

For example, decision trees are a nonparametric machine learning algorithm that is very flexible and is subject to overfitting training data. This problem can be addressed by pruning a tree after it has learned in order to remove some of the detail it has picked up.

② Constraining a model to make it simpler and reduce the risk of overfitting is called regularization.

3. [6 points] Prove Bayes' Theorem.

The probability of two events A and B happening, $P(A \cap B)$, is the probability of A, P(A), times the probability of B given that A has occurred, $P(B|A)$.

$$P(A \cap B) = P(A)P(B|A) \quad (1)$$

On the other hand, the probability of A and B is also equal to the probability of B times the probability of A given B.

$$P(A \cap B) = P(B)P(A|B) \quad (2)$$

Equating the two yields:

$$P(B)P(A|B) = P(A)P(B|A) \quad (3)$$

and thus

$$P(A|B) = P(A)\ P(B|A)\ P(B) \quad (4)$$

This equation, known as Bayes Theorem is the basis of statistical inference.