**IE406 Midterm Exam (Take-home)**

20161190 이민재

**1\_1. (Self-driving cars)**

Laser scanner, radar, GPS computer in trunk see and consider the visual data, so car can figure out what to do. Self-driving cars use this radars and camera data, and it also plans routes through GPS. In addition, it uses data from human driver to conduct training. Since the road situation is very complicated, the application of deep learning is also being studied. Perception and object detection(object is human or some trash)will be permanently move on to the next levels by deep learning.

**1\_2. (Illegal parking detection)**

Illegal parking detection using data from the camera(video capture module). They recognizes image through SSD detection module. And the vehicle is specified through the tracking module. Then Illegal judgment module determine the vehicle is illegal parking or not. The algorithm include some question like Vehicle in ROI?, Previous vehicle? and so on at data read frame by frame.

**1\_3. (Credit card fraud detection)**

This is the anomaly detection type. It detects errors using data such as date, time, payment type, place of payment, amount, etc. If the data used by the user is out of range, it is judged as an anomaly detection. For example, may-22, 1:14pm, FOOD, Monaco Café, $1,127.80. The payment value $1,127.80 is too high value, so it judge the value is anomaly.

**1\_4. (Spam e-mail classification)**

The spam-mail problem is a text classification problem. This is a supervised learning method that gives a label(this sample is spam e-mail, this sample is not spam e-mail) to model. A typical method is the Naive bays method. When a particular word is found(using natural language processing), the probability of classifying it as spam increases or decreases. In this way, the overall probability is calculated, and if the value exceeds a threshold value, it is classify as spam e-mail. If not, classify as no spam e-mail.

**1\_5. (Product display in retail store)**

This is related to the association rule. The association rule is a theory that people who buy late-night snacks at stores are more likely to buy alcohol. It can express how much each item is related by calculating support, confidence, and lift from item sets. Based on this information, the retail store can increase sell amount by displaying related items nearby.

**2.**

There are many ways to solve this problem. The most obvious way is to re-learn using a replaced sensor. In fact  I would learn again with a new sensor because self-driving-related learning is directly related to human life. But learning can be uneconomical if learning costs are too much. Then, we should think about other ways besides re-learning. Since errors were found in some of the 20 sensors, it may be possible to modify the classify algorithm that reduces the weight of sensors that had errors in classification and increase the weight of right learned sensor. In addition, there will be a way to modify the process of classify algorithms by analyzing how the previously mis-learned information affects the outcome in which process.

**3\_1.**

Classification work is fatal in various fields such as autonomous driving and healthcare and so on. Another typical example is anomaly detection. Let's judge whether a company's product is defective or not, and if it is defective, assume it will not sell. The problem of determining it as a defective product when it is not a defective product is not occur a big problem. However, if it is a defective product and it is determined to be normal product, it can cause a critical problem.

**3\_2.**

These problems affect the training process and evaluation protocols. It is important to resolve errors that lead to fatal problems, which require learning in other ways, changing parameters, or changing evaluation methods. After testing the model, we should try various methods and try to reduce the number of critical mistakes.

**4.**

The main difference between traditional machine learning and deep learning is feature extraction. At traditional machine learning, feature exctraction is done by human processing data. But at deep learning, feature extraction takes place from the model. It means, deep learning is to find out directly the patterns and features that data have. In the figure below, machine learning is feature extraction (human) -> classification (model), and deep learning is feature extraction + classification (model).

테이블이(가) 표시된 사진

자동 생성된 설명

**5.**

The difference between DBSCAN and K-means is that determine the number of clusters(value of k) or not. Also, k-mean is center based, and DBSCAN is density based. K-means method set the value of k in advance and perform classification. As the performance varies depending on the value of k, it is important to select k value well. On the other hand, DBSCAN mean density-based spatial clustering of applications with noise. This method classify using two parameters, Eps and Minpts. Eps is Maximum radius of the neighborhood. MinPts is minimum number of points in an Eps-neighborhood of that point. Since it is classify using these two parameters, DBSCAN does not need to determine the k value in advance.

**6.**

Apriori principle mean if an itemset is frequent, then all of its subsets must also be frequent. If we analysis d item set, we consider (number of possible association rules) number of item. If d increase, then the computational cost is too expensive. To solve this problem, we set the minimum support. If some item does not over the minimum support value, we pruned superset. By this way we can reduce much computational time. In summary, Apriori Principle is a theory that if a certain value(example, min support) is not exceeded, then not considering the super set of the item. So the amount of calculation can be greatly reduce.

**7.**

Classification and regression are both supervised learning, but there are many differences. First of all, classification is to predict which class it belongs to when an input is received. There are binary classifications that classify two classes and multi classifications that classify several classes. Regression is the prediction of continuous values, not classes, when an input is received. There are linear regression, logistic regression, and ridge regression.

**8.**

I solve at good note(I-pad) and attach the picture.

테이블이(가) 표시된 사진

자동 생성된 설명

**9.**

The Apriori algorithm requires a large amount of computation when large set of items given and a the minimum support is low. Therefore, the FP-Growth algorithm(Frequent Pattern Growth Algorithm) was proposed after the Apriori algorithm. After generating the FP-Tree, this algorithm extracts a set of frequently targeted items using divide and conqure method. Search the database only twice and ensure faster calculation speed than Apriori(spend many time when scan the database).

**10.**

I solve at good note(I-pad) and attach the picture.

텍스트, 영수증, 서류이(가) 표시된 사진

자동 생성된 설명