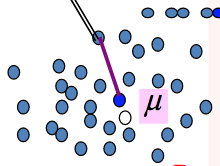
Midterm Thesis

Zhang, Tuxin

1155101156

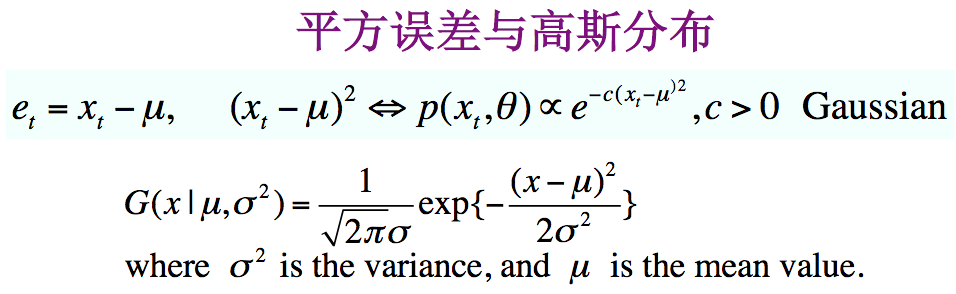
Machine learning is the field or subject that researches how to make computers learn. In other words, a machine learning algorithm is an algorithm that runs in the computer that trains themselves so that human don’t need to explicitly describe the task human want to achieve.

To understand how machine learns, we first need to understand what is law(规律). Law has characteristics such as rhythmical, constancy, unadorned, impersonality. Everything in the world are constrained by law. But once we are familiar with law, we can actually learn some pattern or knowledges from it.

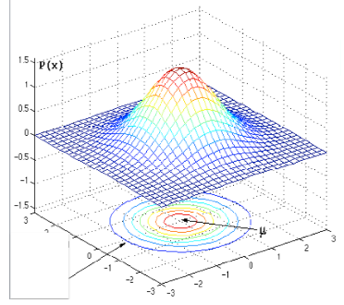


But how exactly can computer learn? A data point (a sample) is the key element of how machine is learning. A point (can be treated as a sample in realistic) is very useful. A point can represent a point, even a cluster (set of points). In vector form it can also represent a line, even a plane. Easily speaking a point can represent a structure. Start from the beginning, if a point representing a point, easy, it is what a point supposed to do. But in the cases that a point is representing a structure other than point, it becomes complicated. How to decide which is the “best” point to represent let’s say a cluster (set of points)? How to find, this is called implementation, we need algorithm. Obviously, we want the point with the lowest error. The lower the error, the better the point is. The next question pops up is how to compute the error? What kind of error measurement method we want to use. Actually, learning is just to decide which theory (best theory, lowest error) to use, what is your learner(model), and finally implementation (code the algorithm). The way we are trying to find the optimal point to represent the cluster is also kind of learning.

Common error for points and points are square error and absolute error. Although we have discussed this in the assignment, but these two errors are quite different. And these two errors should be applied only in correct place. Square error is measured by so called Euclidean distance, basically means how far are these two points is there error. Absolute error focus more on the “coordinate error” which is the difference inside the coordinate system.



Assuming all points are independent and squared error are used, we learned that the mean point μ is the optimal (lowest error) point (can be derived from MLE). The error are added up together, importantly, not all error can be added together to find the minimum. The error can added up together if and only if the points are independent. We can also treat error in another way, how to find the better point, by the error, but we know the very large error appear less, small error appear a lot. Then we turned this into probability now. Probability can then be develop further. That it, Gaussian distribution, set of points can have many different distribution, the most famous one is Gaussian distribution. If a cluster’s points are Gaussian distribution, the mean point μ we just computed and squared error method are related to Gaussian distribution. By central limit theorem we know that the points tend to close to mean with high probability, far from mean point with lower probability. From MLE we can obtain the formula to compute the variance. Together with all parameters we can obtain the completed Gaussian distribution formula and model. (MLE is a method of estimating the parameters of a statistical model such as Gaussian distribution, given the observation, in our case, observation are points. MLE attempt to find the parameters that maximize the likelihood function, minimum error can be viewed as largest likelihood, which means theory can be talk in error manner or likelihood manner).



OK, what if use a point to represent a cluster of points is not enough, we want something more.

Reference:

1.All PTT provided by Prof. Xu lei

2.https://www.quora.com/What-is-machine-learning-4

3.Assignments solution provided by TA

4.Wiki