Certificate Course in Machine Learning using Python [6 Weeks]

Day 28 -Ensemble Learning

Ensemble learning

In the world of Statistics and Machine Learning, Ensemble learning techniques attempt to make the performance of the predictive models better by improving their accuracy. Ensemble Learning is a process using which multiple machine learning models (such as classifiers) are strategically constructed to solve a particular problem.

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem. Ensemble learning is primarily used to improve the classification, prediction, function, approximation, etc.

Example: I want to invest in a company XYZ. I am not sure about its performance though. So, I look for advice on whether the stock price will increase more than 6% per annum or not? I decide to approach various experts having diverse domain experience:

- **Employee of Company XYZ**

This person knows the internal functionality of the company and has the insider information about the functionality of the firm. But he lacks a broader perspective on how are competitors innovating, how is the technology evolving and what will be the impact of this evolution on Company XYZ's product. **In the past, he has been right 70% times**.

- **Financial Advisor of Company XYZ**

This person has a broader perspective on how companies strategy will fair of in this competitive environment. However, he lacks a view on how the company's internal policies are fairing off. **In the past, he has been right 75% times**.

- **Stock Market Trader**

This person has observed the company's stock price over past 3 years. He knows the seasonality trends and how the overall market is performing. He also has developed a strong intuition on how stocks might vary over time. **In the past, he has been right 70% times**.

- **Employee of a competitor**

This person knows the internal functionality of the competitor firms and is aware of certain changes which are yet to be brought. He lacks a sight of company in focus and the external factors which can relate the growth of competitor with the company of subject. **In the past, he has been right 60% of times**.

- **Market Research team in same segment**

This team analyzes the customer preference of company XYZ's product over others and how is this changing with time. Because he deals with customer side, he is unaware of the changes company XYZ will bring because of alignment to its own goals. **In the past, they have been right 75% of times**.

- **Social Media Expert**

This person can help us understand how has company XYZ has positioned its products in the market. And how are the sentiment of customers changing over time towards company. He is unaware of any kind of details beyond digital marketing. **In the past, he has been right 65% of times**.

In a scenario when all the 6 experts/teams verify that it's a good decision (assuming all the predictions are independent of each other), we will get a combined accuracy rate of

1 - 30%*25%*30%*40%*25%*35%

= 1 - 0.07875 = **99.92125%**

The assumption used here that all the predictions are completely independent is slightly extreme as they are expected to be correlated. However, you can see how we can be so sure by combining various forecasts together.

An ensemble is the art of combining a diverse set of learners (individual models) together to improvise on the stability and predictive power of the model. In the above example, the way we combine all the predictions collectively will be termed as Ensemble learning.

Models can be different from each other for a variety of reasons:

- There can be difference in the population of data.
- There can be a different modeling technique used.
- There can be a different hypothesis.

The following diagram presents a basic Ensemble structure:

Some Commonly used Ensemble learning techniques

- **Bagging**

Bagging tries to implement similar learners on small sample populations and then takes a mean of all the predictions. In generalized bagging, you can use different learners on different population. As you can expect this helps us to reduce the variance error.

- **Boosting**

Boosting is an iterative technique which adjusts the weight of an observation based on the last classification. If an observation was classified incorrectly, it tries to increase the weight of this observation and vice versa. Boosting in general decreases the bias error and builds strong predictive models. However, they may sometimes over fit on the training data.

- **Stacking**

This is a very interesting way of combining models. Here we use a learner to combine output from different learners. This can lead to decrease in either bias or variance error depending on the combining learner we use.

General speaking, ensemble learning is the technique of applying more than one Machine Learning algorithms on given problem.

Next

PREVIOUS ACTIVITY

◄ Input Image File

Jump to...

NEXT ACTIVITY

Elbow Method: Python File ►