TECHNO INDIA
SILIGURI INSTITUTE
OF TECHNOLOGY
A Satyam Roychowdhury Initiative

Ensemble Learning

Bagging, Boosting, DECORATE, Active Learning with Ensemble

Ensemble Method

- Ensemble method trained a set of classifier/regression instead of single classifier/regression to provide the best performance.
- The combined models increase the accuracy of the results significantly.

Bootstrap Method

- Randomly select N samples from the training dataset with replacement.
- B number of samples dataset with the same size(N) as the original dataset is created.
- These B dataset called bootstrap dataset.
- Then fit a model for each of the dataset and assess the performance of the model over B sets.
- Then voting or average of probability of occurrence is used for the test instance and this process is known as Bagging.

Bootstrap/Bagging

- The aim of bagging is to reduce the variance of a model.
- Ensemble produce a better classifier than a single classifier.
- It is more useful for the cases of unstable classifier, it provide comparatively stable classifier.
- When we use bad classifier, then bagging produce worst classifier.
- Some ensemble method fit multiple model on the same dataset.
- Multiple classifier can be trained parallelly and independently on the dataset.

Committee Machine

- Assume that we have training dataset D, and the candidate models M_m.
- Committee machine take a simple unweighted average of the prediction of each model

Stacking

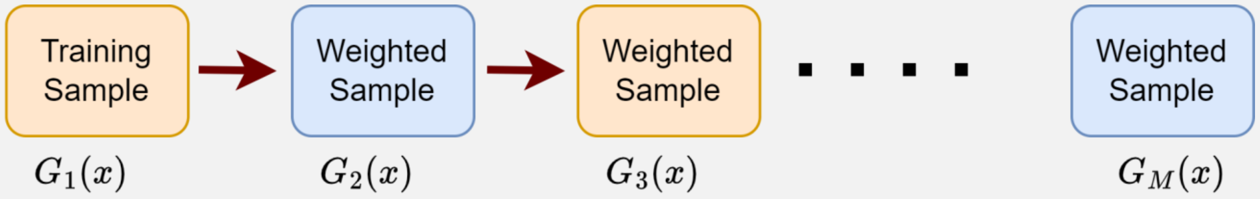
- Fit some set of classifiers on a set of datasets.
- Instead of arbitrarily assigning a weightage to each classifier, it will learn the weight.
- Use each of the classifiers to generate features.

Boosting

- Boosting is designed for the classification problem but can be extended for the regression problem.
- Motivation is to combine many weak classifier and produce a powerful committee.

AdaBoosting

- The goal of boosting is to sequentially apply weak classifiers to modified versions of the data, ultimately generating a collection of weak classifiers g_m.
- The predictions from all the models are then combined through a weighted majority vote to produce the final prediction.
- The more accurate model assigned a high weight, and the weak model has a low weight to indicate stronger and lower influence, respectively.



AdaBoost

1. Initialize the weights to the observation $w_i = \frac{1}{N}, i = 1,2,3 \dots, N$
2. for $m = 1$ to M
   a. fit a classifier $\mathrm{G_{m(x)}}$ to the training data using weights w_i
   b. compute $\mathcal{E}_m = \frac{\sum_i w_i \mathbb{I}(y_i \neq \mathcal{G}(x_i))}{\sum_i w_i}$
   c. compute $\alpha_m = \log \frac{1-\mathcal{E}_m}{\mathcal{E}_m}$
   d. set $w_i \leftarrow w_i . \exp(\alpha_m \mathbb{I}(y_i \neq \mathcal{G}(x_i)), i = 1,2, \dots., N$

3. Output

$$\mathcal{G}(x) = sign(\sum_m \alpha_m \mathcal{G}_m)$$

Diverse Ensemble Creation by Opposite Relabeling of Artificial Training Example(DECORATE)

- Classifiers are trained on the same training dataset.
- Each classifiers are successively trained on the original data with some artificial data.
- In each iteration the number data is generated artificially is specified as a fraction of the training data.
- The label is assigned to this artificial data in such a way that it differ maximally with the ensemble classifier.
- This artificial dataset is called diversity data. It tries to maintain diversity along with the training accuracy.

DECORATE

- Each base class classifier E_m in the ensemble E^* provides probability for the class membership x.
- If $P_{E_i,y}(x)$ is the probability of instance x belonging to the class y as per the classifier E_i, then class membership probabilities for the entire ensemble is $P_y(x) = \frac{\sum_{E_i \in E*} P_{E_i,y}(x)}{|E^*|}$
- Then select the most probable class as label for x i.e. $E^*(x) = \arg\max P_y(x)$

Construction of Artificial Data

- Generate artificial training data by randomly picking data points from an approximation of the training data distribution.
- For numeric attributes, compute the mean and standard deviation from the training set and generate values from the Gaussian distribution defined by these.
- Assume that the attributes are independent.
- Artificially generated examples are labeled based on the current ensemble.

Construction of Artificial Data

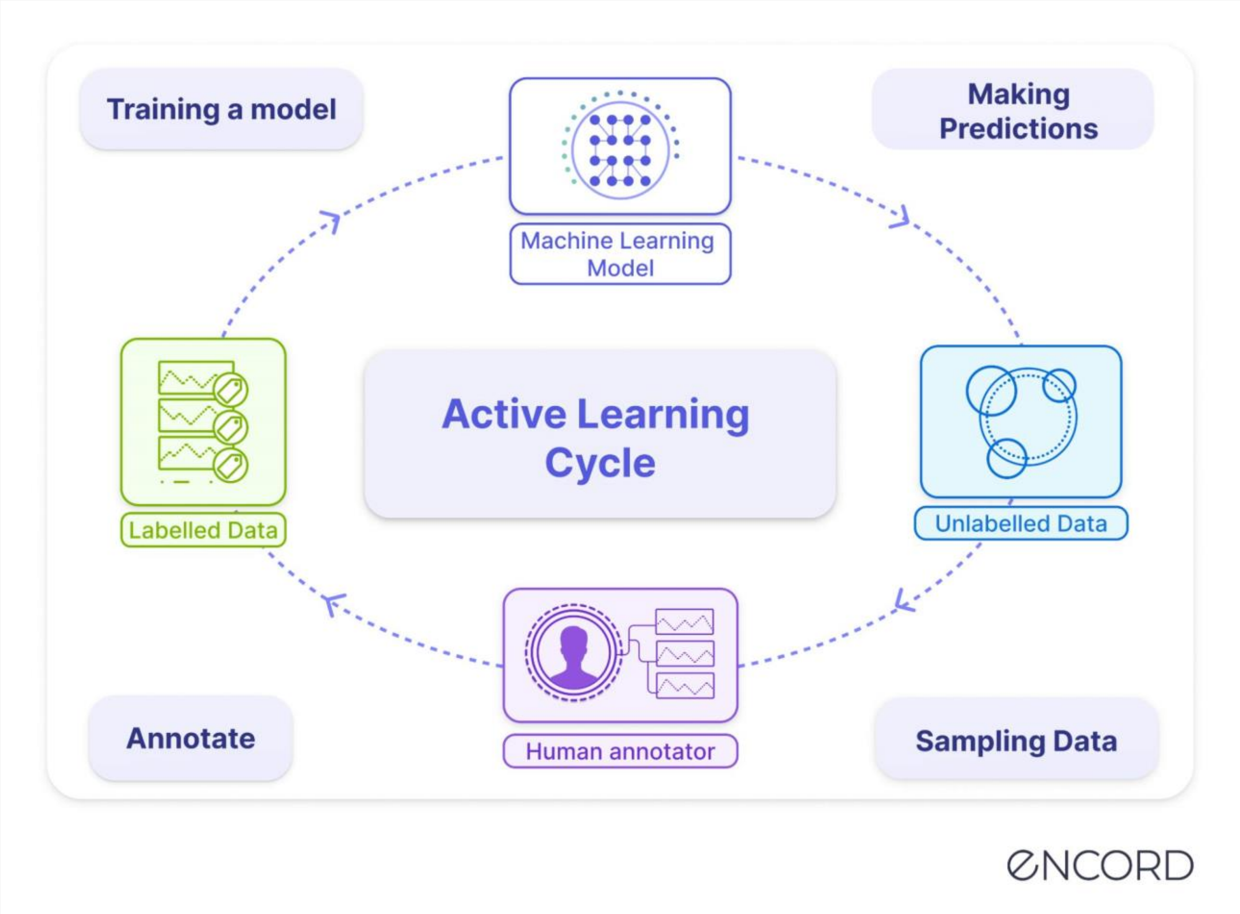
- First, find the class membership probabilities predicted by the ensemble.
- Labels are then selected so that the probability of selection is inversely proportional to the current ensemble's predictions.
- If the current ensemble predicts the class membership probabilities $P_y(x)$, then the new label is selected based on the distribution

$$P_y(x) = \frac{1/P_y(x)}{\sum_y 1/P_y(x)}$$

Active Learning

- Lower cost of smartphone, internet, and IoT devices, making ample data available to the data scientist.
- This vast dataset requires lots of time to analyze.
- Active learning helps users label the data interactively with the desired outputs.
- Active learning is a subset of machine learning in which a learning algorithm can query a user interactively to label data with the desired outputs.

Active Learning



Active Learning with Ensemble

- Initially, we utilize ensemble learning methods to construct our initial classification model using the initial labeled training data.
- The class of the newly arrived unlabelled examples is predicted using the initial model.
- Next, the expert evaluates the predicted values. The expert chooses and labels instances with high prediction probabilities and then appends them to the training set.
- Using the ensemble learning in this step guides the expert toward the most informative labeled example by using multiple classifiers.