Course Name: Machine Learning and Deep learning - Fundamentals and Applications

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-6

Lecture-24

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. Today I am going to discuss the concept of ensemble classifiers that is ensemble learning. If the performance of a classifier is not satisfactory, then I can consider multiple different classifiers. These classifiers can be trained with the help of the original training dataset or maybe these classifiers can be trained with the help of different training datasets. And after the training, the output of these classifiers can be combined by voting or averaging and after the combination, the performance of the overall system may improve. And this is the fundamental concept of the ensemble classifier.

Earlier I discussed the concept of bias and the variance. The problem is due to high bias and the high variance. And in this case, if I consider the ensemble classifiers, the bias and the variance can be reduced.

So I can show later on how to reduce the bias and

the variance with the help of the ensemble classifiers.

And also there are three popular ensemble techniques. One is stacking, another one is bagging and finally the boosting. So all these techniques I will be discussing in this class. So let us begin this class, the concept of ensemble classifiers. So here I have shown the one block diagram showing the concept of the ensemble classification.

So I have one input training dataset and I have the models, the model 1, model 2, so n number of models I am considering. And these models are trained with the help of this common training dataset and output of these models or the output of these classifiers are combined to get the final output that is the ensemble output. So this is the fundamental concept of the ensemble classifier. So this is the ensemble classifier.

So I can give one example regarding this concept.

Suppose the detection of actual mail or spam mail. So you can see suppose I have the training dataset X_n and I can consider different classifiers like support vector machine I can consider maybe the decision tree also I can consider or maybe I can consider that k nearest neighbor classifier also I can consider. And all these are trained with the help of the training dataset. And because I have two classes one is spam another one is not spam that is the classification of emails. So I am getting the outputs one is the spam and another one is the not spam or maybe output of the KNN is spam.

So I am getting this output and after this these all these outputs I am combining. So in this combination I can consider the concept of voting or averaging. So I can apply the concept of voting or maybe I can apply the concept of averaging. And finally I am getting the output. So if I consider the concept of voting then I will be getting the output that is the spam.

So that means all the outputs of the model 1 model 2 all the models are combined and I am getting the final output. So this is the fundamental concept of the ensemble classifier. So in this class I will be explaining the most popular ensemble techniques. So first I will be discussing the concept of stacking. Next I will be discussing the concept of bagging.

And finally I will be discussing the concept of boosting. So these are very important ensemble learning techniques. So these concepts I will be explaining now. So move to the next slide. So here I have shown the concept of stacking.

So in this case you can see I have a training data set and I have a number of models model 1 model 2 up to model n. And all these models are trained with the help of this training data set the common training data set. And output of these models I am considering to train the second level model that is the final model. And after the training I am getting the ensemble output. This model 1, model 2, model n these are first level learners.

And if you see the model the final model that I can consider as second level learner second level learner or sometimes it is called the meta learner. And this model 1 model 2 and this all these models I can consider as first level learners. So the output of all the models model 1 model 2 up to model n these I can consider as features to train the second level model that is the meta learner. And after this I am getting the output that is the ensemble output. So this is the fundamental concept of the stacking.

And in this case also I can give one example the problem is the same problem that is the mail detection email detection spam and the non spam. So suppose I have this training data set and I am considering the classifiers suppose the support vector machine decision tree and maybe we can consider the K nearest neighbor classifier. These are level 1 level 1 models and output of these models I am considering.

So output of this model is suppose spam non-spam and spam these are the outputs of the level 1 models. So all these are combined here.

So these we are combining so we are getting we are having the level 2 model. So level 2 model suppose I can consider maybe the decision tree the input to the level 2 models are the outputs of the level 1 models. So this I am getting the level 2 model and finally from this model I can do the prediction and the prediction maybe spam. So this is the fundamental concept of the stacking.

So you can see I have level 1 models and also I have level 2 models.

So this is the fundamental concept of stacking. Now let us discuss about the concept of bagging. So in this figure I have shown the concept of bagging. So this is bagging. So in this case I have a training dataset and from this training dataset I am getting number of training datasets like training set 1 training set 2 training set n.

So all these training sets I can obtain from the original training dataset and in this case I am applying the principle of sampling with replacement. So that concept I am explaining later on sampling with replacement. That is the random sampling with replacement. So after this what we can consider this model 1 model 2 all these models I can consider. So model number 1 is trained with the help of training dataset 1 model number 2 that is trained with the help of the training dataset 2 like this the model number n that is trained with the help of the training dataset n and after this we are considering we are combining the outputs of all these models model 1 model 2 model n and I am combining and I am getting the output.

So maybe I can consider the voting or averaging principle to combine. So this is the fundamental concept of the bagging but one thing it is important. So how to get the training set 1 training set 2 all these training set I can obtain from the original training dataset by the process of random sampling with replacement. So that concept I will explain after some slides. So this is the concept of bagging and finally I want to explain the concept of boosting.

So here I am showing the concept of boosting. This is also a very important ensemble learning technique. So here you can see I have a training dataset and I am considering all the models model 1 model 2 model 3 all these are classifiers. So model 1 is trained with the help of the training dataset and this model 1 can classify the samples of the training dataset and some of the samples will be misclassified. So I am giving importance to the misclassified samples that means the samples which are misclassified by the model number 1 they are given maximum importance and so that is why I am getting the weighted sample 1.

So I am getting another training dataset considering the weighted samples. Weighted sample means the samples which are misclassified are given maximum importance and these are weighted. So the samples which are correctly classified I am not giving the importance. So I am getting the weighted sample 1 that is the training dataset. This model number 2 it is trained with the help of the weighted sample 1 training dataset and in this case also I cannot perfectly classify all the samples.

There will be there will be some misclassifications. So that is why I have to weight the samples all the misclassified samples are weighted based on the importance. So that means I am giving the maximum importance to the misclassified samples. So I am getting the weighted sample 2 and the model number 3 like the previous one it is trained with the help of the weighted sample 2 training dataset. And again in this case there may be some misclassification.

So I have to give the importance to the misclassified samples. So that is why I am getting the weighted sample 3 training dataset. So like this I have to train all the models. So this model 1 model 2 model 3 these are actually called the weak classifiers or I can say these are weak learners. So this model 1 model 2 this model 3 these are the weak learners and the output of these weak learners are combined to get a strong classifier or the strong learner.

So this is strong classifier or I can say the learner. So output of all these weak learners are combined in the strong classifier. So these are inputs to the classifier and these are combined and I am getting the ensemble output. So this concept is used in the AdaBoost algorithm. So AdaBoost this algorithm I will be explaining in my next class.

So what is the meaning of AdaBoost adaptive boosting. So adaptive boosting. So the objective is to generate a strong classifier from the weak classifiers. That is the objective of the boosting.

So in case of the bagging so I explained the concept of random sampling with replacement.

So that concept I am going to explain what is this principle. So here I have shown the concept of sample random sampling with replacement. So I have shown in a vase there are 10 balls 1 2 3 4 5 6 7 8 9 10 and I am doing the sampling randomly I am doing the sampling. So if you see in the figure first figure in unit number 1 I am just drawing a ball from out of this 10 balls.

So first suppose the ball 5 is sample. So out of 10 balls suppose I am drawing a particular ball and suppose the ball is 5 the number ball is 5. So what is the probability of obtaining the ball 5 because I have 10 balls. So it is 1 divided by 10 that is 0.1. In unit number 2 you can see I am replacing the ball the ball 5 is placed in the vase again and in the unit number 2 I am drawing the another another ball that is randomly and the ball is suppose 7.

So this is the random sampling. So in the second unit if you see the unit number 2 I am drawing the ball 7. So if you see in unit number 3 in the figure 1 the ball 7 already I have drawn and that is replaced if you see the ball 7 is replaced that is placed in the vase I am drawing a ball randomly and suppose this ball is 10. So what is the probability in this case the probability will be same it is $\frac{1}{10}$. Similarly in the unit number 4 I am again drawing the ball randomly and suppose the ball is 5 and what is the probability in this case also if the probability will be 1 divided by 10 because out of 10 balls I am selecting 1 randomly. So this is the concept of random sampling with replacement.

So with this concept I am generating the training dataset. So I have the original training dataset and I am applying this concept to get the other training datasets. The second figure I have shown a random sampling without replacement. So in this case first you see in unit number 1 I am drawing a ball randomly the ball is suppose 5. In unit number 2 how to get the unit number 2 in this case I have not replaced the ball 5.

In case of the unit number 1 what is the probability of obtaining the ball 5 this is $\frac{1}{10}$. But in case of the unit number 2 in the figure 2 what is the probability of getting the ball 7 this is $\frac{1}{9}$ because now I have 9 balls. And in the unit number 3 that means how to get the unit number 3 I am drawing the ball number 10. But if you see I have not replaced the ball number 7. So now you can determine the probability and similarly the unit 4 that is the 4 the 4th ball I can draw that is 1 I can draw and in this all these cases I am not doing the replacement that is the without replacement.

So you can see the fundamental difference between these 2 techniques one is random sampling with replacement and another one is random sampling without replacement. So we are considering this concept the concept is random sampling with replacement. So I have suppose these samples all these samples are available all these are samples these are samples. So this is a single training dataset so that means is a complete training dataset complete training set and this complete training set we employed in case of the stacking. The next is I am considering the random sampling with replacement.

So the replacement concept already I have explained so that I am getting the training dataset and that is with the help of random sampling with replacement. So this is I am obtaining this is random sampling with replacement. So this concept we applied in case of the bagging bootstrap aggregation the meaning of bagging is bootstrap aggregation. And finally in the third figure what I want I want to consider random sampling with replacement over weighted data. So I can write random sampling with replacement over weighted data that means we are considering weighted samples.

So these are weighted samples. So I am giving the importance to the misclassified samples the concept of weighted samples I have explained in case of the boosting. So I am considering the weighted samples in this case also. And this principle we have applied in case of the boosting. So here you can see I am showing the concept of stacking bagging and also the boosting and also I am explaining the concept of random sampling with replacement.

So now let us consider the concept of bagging. So what is actually the bagging. So bagging concept already I have explained. Now just I want to write the algorithm for bagging. So bagging so the first step I can say take original dataset D with n number of training examples after this I am creating m copies of this training dataset.

So $\{\tilde{D_m}\}_{m=1}^{M}$. So how to create this one, each $\tilde{D_m}$ is generated from D from the original dataset from D by sampling with replacement. The second point is each dataset that is the dataset is $\tilde{D_m}$ has the same number of examples as in the dataset D. So that is the original data set. So each dataset that is the $\tilde{D_m}$ has the same number of examples as in the dataset D. After this after considering this we are considering the training of the models train models.

So what are the models h_1 suppose h_2 like this all these models I am trained I am doing the training using the training datasets $\tilde{D_1}, \tilde{D_2}, \tilde{D_m}$ respectively. After this we have to combine all these outputs so we can consider averaging so use an averaged so average model is h and we have m number of models just I am doing the averaging $h = \frac{1}{M}\sum_{m=1}^{M} h_m$ and that is the final model and this is the final model. So this concept is used for the concept of bagging is useful for models with high variance. So it is useful for models with high variance and also it can consider noisy data. So you already I have discussed about the high variance if the variance is very high that means we are considering a very complex model and that is nothing but the overfitting.

So the overfitting can be reduced with this concept that is the bagging model that is a bagging ensemble model. So this is the concept of the bagging. So pictorially I can show the concept of bagging so this is the illustration. So in this figure you can see I am considering the original data so this is the original data and in the middle we are considering 3 models. So these 3 models are learned using 3 datasets and that is selected by considering the random sampling with replacement.

So you can see that D_1 another one is D_2 another one is D_3 these are the training datasets and I am considering 3 models which are trained with the help of the training dataset D_1, D_2 and D_3. And after this you can see I am considering the Everest model. So the final one is the Everest model. So this models in the middle row 3 models are trained with the help of the training dataset D_1, D_2 and D_3. And after this the output of these models I am combining that means I am averaging and I am getting the final Everest model.

So that is the concept of the bagging. So now let us discuss one important concept that is the random forest classifier and this is the concept is based on the bagging. So this concept already I have explained in one of my classes and here you can see I have the training dataset and from this training dataset I am obtaining the training dataset 1 training dataset 2 training dataset n. So all these training dataset I am obtaining. After this I am considering the decision trees decision tree 1 decision tree 2 decision tree n.

So all these decision tree I am considering. The decision tree 1 is trained with the help of the training dataset 1 the decision tree 2 it is trained with the help of the training dataset 2 like this we are considering and the output of all these decision trees are combined and maybe we can consider voting or averaging and after combination I can do the prediction. So this is the fundamental concept of the random forest. So we are considering number of decision trees and all these decision trees are trained with the help of the training dataset 1 training dataset 2 training dataset n respectively and after the training I am considering the outputs of all decision trees and we can combine these outputs by voting or averaging and finally I am getting the ensemble output and that is the prediction. So this is the fundamental concept of the random forest. So here also I have shown the concept of the random forest I have 3 decision trees.

These decision trees are trained with the help of different training datasets. So all these training datasets I am obtaining with the help of the principle the principle is random sampling with replacement. So in this case I can consider like this and this is the ensemble classifier and we are considering all these decision trees. So given a total of d number of features, each decision tree uses root d randomly chosen features. So since we are considering randomly selected features that means all these decision trees are uncorrelated.

So this point is very important I can repeat this sentence because we are considering randomly selected features because of this all these decision trees will be uncorrelated and all these decision trees have the same depth. So these two points better to write randomly chosen features make the decision trees DTs uncorrelated and all the DTs usually have the same depth and you can see here and each DT will split the training dataset differently at the leaves and the prediction for a test sample how to do the prediction we have to consider the voting or the averaging. So corresponding to the first decision tree I can do the prediction the prediction is $P(Y|X)$. So that prediction I can obtain the $p_1(y|x)$ I can determine from the first decision tree from the second decision tree I can determine again the $P(Y|X)$ that is the $p_2(y|x)$ and similarly from the third decision tree I can also predict a $P(Y|X)$. So this I can combine I can consider averaging or voting and if you see the structure of the decision trees they are not same they are different but the depth is same.

So if you see the structure of these decision trees because this is the root node this is the root node and you can see these are the leaf node and these are the internal nodes internal nodes and this is the leaf nodes. So the structure of this decision trees are quite different and this is the fundamental concept of the random forest. So in this class I explained the concept of ensemble learning that is the ensemble classifiers and I explained the concept of three techniques one is stacking another one is bagging and finally the boosting and one important concept is random sampling with replacement. So we can obtain different training datasets from the original dataset with the principle the principle is random sampling with replacement and finally I explained the concept of random forest. In my next class I will be explaining the concept of boosting and also the concept of the adaboost classifier. So let me stop here today. Thank you.