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-1

Lecture-2

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. In my first class, I discussed the concept of pattern classification and the concept of machine learning. Today I am going to discuss about the performance evaluation of classification. So for performance evaluation, I can consider some metrics like accuracy, recall, precision, F1 score. So all these metrics I can consider for performance evaluation and mainly all these parameters I can determine from the confusion metrics. So what is a confusion metrics? Suppose a particular class I want to recognize.

So how many times a particular class is recognized correctly that I can determine and how many times a particular class is not recognized correctly that also I can determine and based on this I can determine the confusion metrics. And from the confusion metrics I can determine the percentage accuracy I can determine, misclassification rate I can determine and also I can determine the metrics like accuracy, recall, precision. So all these parameters all these metrics I can determine from the confusion metrics.

So let us discuss about the performance evaluation of classification.

So let us start this class. So the performance of classification. So for determining the performance what actually we need to consider. First I have to train the model using the learning algorithm and for this we are considering a training set.

And after this we have to consider testing that is nothing but validation.

So we are considering a validation set and with the help of this validation set I can determine the performance of a classifier. And for this training we have to do hyper parameter tunings. So we have to tune the parameters of the models. So that is nothing but the hyper parameter tuning.

And after this we have to stop the training because we have to do the training iteratively.

So we have to stop the training to avoid overfitting. And finally I want to evaluate the performance of the model with the help of the testing samples that is the validation set. So in a pattern classification system first I have to do the training and I will be getting the learned model the train model and with the help of the train model I have to do the classification I have to do the prediction. So that is the concept of the classifier in machine learning. So for classification metrics I will be discussing the concept of the confusion metrics, accuracy, precision, recall, F1 score and area under the ROC curve.

So these are the metrics with the help of this I can determine the performance of a classifier the performance of classification. So all these metrics I am going to discuss and mainly the important point is the confusion metrics. So from the confusion metrics I can determine the rest of the parameters the rest of the metrics. So what is confusion metrics here you can see the confusion metrics is nothing but a table to summarize how successfully the classification model can predict examples belonging to various classes. So I will be considering c number of classes and how many times a particular class is recognized correctly that I can determine and based on this I can determine the confusion metrics.

So let us move to the next slide here you can see suppose for a binary classification corresponding to an email I have two classes one is spam another one is not spam. So here you can see in this metrics that is a confusion metrics we are showing the actual classes spam and not spam and also I have shown the prediction. So the prediction is spam and not spam. So here you can see in the first we are determining the true positive that means how many times the spam is recognized as spam. So that is the true positive.

So how many times it is recognized as spam and what is the false negative. So how many times the spam is recognized as not spam. So that is the false negative. So from this I can determine the false negative and you can see another class is suppose not spam.

So how many times the not spam is recognized as spam.

So that is nothing but the false positive and how many times not spam is recognized as not spam. That is the true negative. So you can see we are determining all these parameters that true positive false positive and false negative and true negative. So this is the confusion metrics. So how many times a particular class is correctly recognized and how many times a particular class is not correctly recognized based on this I can determine all these all these parameters like true positive false negative, false positive, true negative.

So all this we can determine. So let us consider another example consider a cricket tournament and I want to show the mapping in the confusion metrics. So you had predicted that India would win and it won. So that means it is the true positive. Number two is you had predicted that England would not win and it lost and that corresponds to true negative.

And number three is you had predicted that England would win but it lost that is nothing but false positive. And number four is you had predicted that India would not win but it won that is the false negative. So you can see all these parameters the true positive, false negative, false positive, true negative all this you can determine from the confusion metrics. So this is the confusion metrics. So like this I can consider another example.

So actual classification the actual classes we are showing 1 1 1 1 1 1 1and 0 0 0 0. And what is the predicted classification you can see the predicted classification I am showing in second row *0 0 1 1 1 1 1 1 1 0 0 0*. So corresponding to the actual class one what is the predicted classification the predicted classification is 0 corresponding to the actual class the class is 1 the predicted class is 0 corresponding to the actual class 1 the predicted class is 1. So that I want to mean like this so corresponding to this you can see and this is false negative the first one is the false negative because the actual class is 1 and the predicted class is 0 that is the false negative. The second point you can see the second column that is actual class is 1 but the predicted class is 0 so that is the false negative and if you see the third column so 1 is correctly recognized as 1 so actual class is 1 the predicted class is 1 that is the true positive and similarly another if you see 1 is recognized as 1 that is the true positive and what is false positive the actual class is 0 and that is predicted as 1 that is the false positive and what is the true negative.

So you can see actual class is 0 and that is also predicted as 0 so that is the true negative. So from this you can see how many times you are getting true positive so 6 times you are getting true positive 2 times you are getting false negative 1 time you are getting false positive and 3 times you are getting true negative and that is the confusion matrix. So move to the next slide so this is another example and that is the multi class classification and that is the emotion classification. So we are considering these emotions like happy, sad, angry, surprised, disgust and neutral. So we have shown this actual classes

and also the predicted classes.

So how many times happy is recognized as happy how many times sad is recognized as sad how many times angry is recognized as angry. So this we can determine from the results from the classification results we can determine these values and how many times suppose the happy is recognized as sad or the happy is recognized as angry or the happy is recognized as surprised or disgust or neutral that also we can determine from the classification results. So we can fill up this table so this is the table is nothing but the confusion matrix. So move to the next slide here you can see based on this actually we can determine the true positive false negative false positive and true negative and one metric that is very important the accuracy.

So how many times a particular class is recognized correctly so that is the accuracy.

So accuracy means the number of correctly classified examples divided by total number of classified examples. So what is the total number of classified example that you can determine from the confusion matrix that is nothing but how many times true positive how many times true negative how many times false positive and how many times false negative that is nothing but the total number of classified examples. So now we have to determine how many times the particular class is correctly classified that is nothing but the true positive and true negative. So true positive plus true negative divided by tp plus tn plus fp plus fn that is the accuracy.

So accuracy also we can determine from this confusion matrix.

So in this case here you can see accuracy is given by the number of correctly classified examples divided by total number of classified examples. So in this table or in this matrix and the confusion matrix I have shown the true positive is equal to 6 false negative is equal to 2 false positive is equal to 1 and true negative is equal to 3 and from this you can determine the accuracy. After this the next matrix that is the next parameter I can determine from the confusion matrix. That is a precision.

So precision is the ratio of correct positive predictions to the overall number of positive predictions.

What is the overall number of positive predictions that is nothing but the true positive and the false positive. So in the table you can see in the confusion matrix you can see one is the true positive another one is the false positive. That is the overall number of positive predictions and you can see it is the true positive we are considering. So precision is nothing but that the ratio of correct positive predictions to the overall number of positive predictions.

So suppose the false positive is 0 then that means I am getting the precision is equal to 1.

So suppose in a classification problem I need I want to give the importance of the false positive then you have to consider this metric. The metric is the precision. So false positive whenever I want to consider then I have to consider the this metric or this parameter the parameter is the precision. The another parameter is the recall.

So you can see this recall also I can determine from the confusion matrix.

Recall is nothing but the ratio of correct positive predictions to the overall number of positive examples. So the overall number of positive examples is nothing but true positive plus false negative that is the overall number of positive examples and Tp divided by Tp plus Fn and that is nothing but recall. So in a classification problem suppose if I consider or if I give more importance to false negative then I have to consider this metric that is the recall. So you can see the recall can be determined from this values true positive false negative and true positive and that is you can obtain from the confusion matrix.

The another metric is so another parameter is the F1 score.

So F1 score is nothing but the harmonic mean of the precision and the recall because in case of the precision we are considering Fp. Fp means the false positive. In the recall we are considering false negative. Okay? So in one case that means in precision we are giving the importance to the false positive.

In case of the recall we are giving the importance to the false negative.

But in this F1 score we are considering the harmonic mean of the precision and the recall. So that means both are considered. So if I consider both F1 F1 score is nothing but harmonic mean. So it is 2 divided by 1 by recall into 1 by precision.

So you will be getting this one. So that is nothing but the true positive divided by true positive plus 1 by 2 false positive plus false negative. So that means we are considering both the false positive and the false negative we are considering together. So if the false positive is 0 and if the false negative is also 0 then F1 score will be 1. So that is the perfect model.

I am getting. So for a perfect model the F score is 1. So this is the concept of the F1 score. In this case I am considering the importance to the false positive and the false negative and we are taking the average of the false positive and false negative because it is 1 by 2. So move to the next slide and that visualization of the precision and the recall. So in this figure I have shown here that the false negative this part is the false negative and you can see the true positives false positives and the true negatives.

So corresponding to this we can determine the precision. So how many selected items are relevant. So that means you can see the precision is nothing but true positive divided by true positive plus false positive.

So precision already I have shown the precision is nothing but true positive plus false negative.

So from the figure you can see. So if I consider this part the overall part that part is nothing but the true positive plus false negative you can see from the figure and the true positive already I have shown in the figure. So this part that this green part is the true positive. So you can understand or you can visualize what is the precision and what is the recall. Move to the next slide. So this recall and the precision depends on the problem the classification problem.

So suppose the problem is the diagnosis of cancer. So what metrics I should consider. So suppose the cancer detection. So cancer is detected as cancer that is the perfect cancer is not detected the actual cancer is not detected as cancer that is this is the value and no cancer is detected as cancer. So that is the false positive and no cancer is detected as no cancer that is the perfect that is the true negative.

So the first one is a perfect means it is the true positive cancer is detected as cancer there is a true positive cancer is detected as no cancer that is the false negative and no cancer is detected as cancer that is the false positive and no cancer is detected as no cancer that is the perfect that is the true negative. So corresponding to this what matrix I should consider. Now in this case we raise a false alarm but the actual positive cases should not go undetected. So that means we raise a false alarm but the actual positive cases should not go undetected. So corresponding to this case the best matrix will be recall.

Recall is nothing but the true positive divided by true positive plus false negative. So that means corresponding to this problem the diagnosis of cancer I have to consider the matrix the matrix is the recall. So suppose another problem I am giving another example detecting if an email is spam or not spam. So this is the detection of the email. So in this case it is important in emails where it is more important that we do not miss any important email as spam then receiving an occasional spam as no spam.

That means our importance is we should not miss any important emails as spam. So corresponding to this example corresponding to this application what will be the best matrix. So you can see in this confusion matrix the spam is detected as spam that is nothing but the true positive perfect. The spam is detected as no spam that is the okay that is nothing but the false negative and no spam is detected as spam.

So that is nothing but the false positive and no spam is detected as no spam that is the true negative that that we should consider.

So corresponding to this what matrix I should consider because here we are considering okay that is not a problem the spam is detected as no spam that does not matter. The matter is that we have to we should not miss the important emails as spam that is the importance we have to give. So corresponding to this example the matrix should I should consider as the precision. So corresponding to this problem the matrix the best matrix I should consider as the precision that is nothing but the true positive divided by true positive plus false positive.

So where I should give the importance and based on this I can select the recall I can select the precision.

So corresponding to this problem you can see in the confusion matrix I have shown t p is equal to 6 f n is equal to 2 f p is equal to 1 and t n is equal to 3 and corresponding to this I can determine this matrix or these parameters the precision recall and f1 score I can determine. Just you try to determine this one is a very simple one and corresponding to this multiclass classification you can see I am showing all the emotions so happy sad angry surprised disgust and neutral. So how many times the happy is recognized as a happy that value I can determine from the classification results. How many times sad is recognized as sad that also I can determine from the classification results. How many times it is misclassified that means how many times a happy is recognized as sad happy is recognized as angry happy is recognized as surprised happy is recognized as disgust and happy is recognized as neutral that is nothing but the misclassification.

So corresponding to this I can determine this two matrix one is the recall another one is the precision. So you can see the recall for determining the recall I have to consider this row because recall is nothing but $\frac{t_p}{t_p+f_n}$. So that means if I consider this row I can determine the recall because this is the first one is the tree t p and the rest is the false negative these are all false negative. So all these are false negative these are false negative. So recall is nothing but you know recall is nothing but the t p divided by t p plus f n that means I have to consider that row in the confusion matrix.

And what about the precision? Precision is nothing but the t p and it is $\frac{t_p}{t_p+f_p}$So this is the t p the how many times happy is recognized as happy and what is the false positive sad is recognized as happy that is the false positive angry is recognized as happy that is also the false positive surprised is recognized as happy that is also false positive disgust is is recognized as happy that is the false positive and neutral is recognized as happy that is the false positive. So that means I have to consider this column to determine the precision. So for determining the recall I have to consider the row and for determining the precision I have to consider this column in the confusion matrix. So based on this I can determine the accuracy how to determine the accuracy.

So you can see this is the true positive happy is recognized as happy sad is recognized as sad angry is recognized as angry that is the true positive surprised is recognized as surprised disgust is recognized as disgust and neutral is recognized as neutral. So I have to consider these diagonal entries and that is the diagonal entries means it the all these are true positives. So from these entries that is the true positives I can determine the accuracy. So you can see I am getting this diagonal values and that is nothing but the true positive and from this we can determine the true positive.

So these values are true positives. So how many times a particular class is recognized correctly. So this is nothing but the true positive and this is the false negative. So how many times the happy is recognized as sad. So these are the false negative. So that means from this this true positive I can determine the accuracy percentage how many times how many times a particular class is recognized correctly that is the percentage accuracy I can determine and the percentage misclassification rate percentage misclassification rate percentage misclassification also I can determine because I have the full I know the false negative values.

So how many times a particular class is wrongly classified. So that is the false negative. So from the false negative values I can determine the misclassification percentage and another percentage I can consider the rejection percentage the rejection rate. So rejection rate I can consider suppose I am considering the problem of suppose the character recognition suppose the character is suppose A. So if I input this A and the classifier can recognize A correctly and suppose if I write even the small letter that classifier can recognize correctly if I write A something like this the classifier can recognize correctly this is the problem is the alphabet recognition.

So if I write A something like this the classifier can also recognize but if I write A something like this in this case it is not available in the learned model.

So that means it is not correctly recognized or suppose in the alphabet recognition because we have 26 alphabets. So suppose if I give so this type of alphabet English alphabet the classifier cannot recognize that will be rejected.

So how many times it is rejected. So based on this I can determine the rejection rate.

So how many times a particular input is rejected. So based on this I can determine the rejection rate. So in a alphabet recognition system if I give some these type of alphabets that is not the English alphabet. So the classifier will reject that one. So you can see accuracy rate or accuracy percent is the misclassification percentage and the rejection percentage or rejection rate we can determine from the confusion matrix.

So this is the example. So actual class level I have shown here these are the actual class levels and these are the predicted classes. So you can see the diagonal entries this is 137, 55. So high value we are getting that means these are the true positive values the true positives. So from this you can see that how many times one is recognized as one it is 137 times.

How many times two is recognized as two 55 times. So these are the correct decisions the diagonal elements and from this you can determine the percentage accuracy you can determine. And if you see here this 13, 3 all these this is nothing but the false negative and from this you can determine the misclassification rate you can determine.

So this is the confusion matrix. Now after this I am considering another matrix or another parameter and that is the area under the ROC curve.

So what is the ROC? ROC means the receiver operating characteristics. So this term comes from the radar engineering. So receiver operating characteristics. So ROC curve I can plot or I can determine from these two parameters one is the true positive rate that is nothing but the recall and another parameter is the false positive rate. So from these two parameters one is the true positive rate another one is the false positive rate I can draw the ROC the receiver operating characteristics.

So move to the next slide. So what is the true positive rate that is nothing but the recall $\frac{t_p}{t_p+f_n}$ that is the true positive rate I can determine like this. So what is the false positive rate that is the false positive divided by false positive plus true negative. So that means the proportion of negative examples predicted incorrectly that is the false positive rate. What is the true positive rate? The proportion of positive examples predicted correctly and that is nothing but the recall.

So you can see how to determine the Tp and the Fp the true positive rate and the false positive rate and based on this I can draw the ROC the receiver operating characteristics.

And another point I want to mention here that area under the ROC curve that is a very important parameter to determine the classification performance. So corresponding to this confusion matrix we have shown what is the true positive that is the recall $\frac{t_p}{t_p+f_n}$ that is the true positive. So from this here you can from the table you can see there is a confusion matrix and the false positive rate I have already explained in my previous slide that is the false positive divided by false positive plus true negative. From this you can see another parameter I can determine that is nothing but specificity that is nothing but 1 minus a false positive rate and move to the next slide.

So based on this true positive rate and the false positive rate you can draw the ROC and also we can see or we can determine the area under the ROC curve.

So based on the area under the ROC curve I can determine the performance of a classifier.

So here I have shown the curve that is the ROC curve and that is the curve between true positive rate and the false positive rate. So this false positive rate it is from 0 to 1 you can see in the x axis and the true positive rate it is from 0 to 1 in the y axis and for a classification I have to take one threshold and based on the threshold I have to determine the true positive and the false positive that means the classification decision I have to take based on the threshold.

So suppose based on a particular condition if this condition is satisfied that means if this parameter is greater than a particular threshold that means a particular class is correctly classified that means for a classification decision I have to consider a threshold.

You can see the true positive versus false positive rate for different thresholds. So you can see this is the value the true positive versus false positive rate corresponding to a particular decision threshold and similarly if I consider this one that is the value of the true positive versus false positive rate corresponding to one decision threshold. So that means the true positive rate and the false positive rate depends on the classification threshold or the decision threshold. So already I told you that area under the ROC curve is quite important. So with the help of this parameter that is the area under the curve area under the ROC curve we can compare different classifiers and it can be useful to summarize the performance of the classifiers into a single measure. So you can see in this figure I have shown the plot of the ROC and you can also see the area under the curve.

So corresponding to the random classifier I am showing the red dotted line. So that is the plot between the true positive rate and the false positive rate and if you see this purple line that is for the perfect classifier that is the ROC curve for the perfect classifier and you can see other colors like blue color yellow colors or the green colors these are the ROC curves for different classifiers and based on this ROCs we can see the area under the curve. So for the perfect classifier the area under the curve is one because x axis it is from 0 to 1 and y axis from 0 to 1. So if I consider this purple ROC curve you can see the area under the perfect classifier ROC curve is 1. So move to the next slide you can see.

So this area under the curve AUC ranges in the value of 0 to 1. So a model whose predictions are 100% wrong that means the area under the curve will be 0. So if all the predictions are wrong 100% wrong then the area under the curve will be 0 and one whose predictions are 100% correct that means I have shown the purple curve that is the perfect classifier the AUC will be 1 and this AUC that is the area under the curve is classification threshold invariant and it is very important for comparisons. So it is very suitable for comparisons.

So this is the best parameter that the AUC I can consider for comparing different classifiers or different classification methods.

So here you can see I am considering this confusion matrix. So the problem already I have explained the problem of the email spam and the not spam and here you can see we are considering the true positive the true positive is 10 false positive is 10 false negative is 0 and true negative is 0. So from this you can determine the TPR the true positive rate you can determine and false positive rate you can determine and corresponding to this you can determine the point in the ROC curve. So you try where will be the point in the ROC curve corresponding to these values you can try. Similarly if I consider this case the spam is recognized as spam that is 0 times the spam is recognized as not spam 10 times and similarly the not spam is recognized as spam 0 times not spam is recognized as not spam 10 times that means it is the this is the false negative and this is the true negative. So from this you can determine TPR and FPR that is the true positive rate false positive rate you can determine and corresponding to this you can determine the point in the ROC curve.

So you try to do this so what will be the point in the ROC curve corresponding to this case and corresponding to this case that is true positive I have shown along the diagonals so 10 times the spam is recognized as spam and 10 times not spam is recognized as not spam so that means we have this true positive and the true negative. So you can see these values are 0 0 the rest of the values so from this also you can determine the true positive rate and the false positive rate and corresponding to this you can determine the point in the ROC curve so that also you can determine. So for the perfect classifier the area under the ROC curve will be 1 so corresponding to this case whether you will get the area under the curve will be 1 or not that you can check. And corresponding to this example also I have shown the true positive is 5 and the false negative is 5 and false positive is 5 and true negative is 5 and again you can determine the TPR value, FPR value and

corresponding to this you can determine the ROC curve you can determine the what is the point in the ROC curve that also you can determine.

So just you try to do this. So that means in summary you can see what are the parameters or what are the metrics we can determine the true positive that means number of correct matches that we can determine false negative so matches that are not correctly detected that is the false negative the false positive matches that are incorrect that is the false positive true negative non-messes that are correctly rejected true positive rate already I have explained TPR you can determine false positive rate you can determine accuracy you can determine by using this formula precision also you can determine the recall you can determine and specificity you can determine and also I have mentioned the F1 score also you can determine. So these are the parameters or these are the metrics with the help of this metrics you can determine the classification performance or you can determine the performance of the classification techniques and one important point already I have explained that is the receiver operating characteristics and another one is the area under the curve AUC. So this you can determine. So in this class I discussed the concept of performance evaluation of classification. So for this I have explained how to determine the confusion metrics and from the confusion metrics you can determine some of the metrics like accuracy you can determine precision you can determine recall also you can determine specificity you can determine F1 score you can determine so all these parameters we can determine from the confusion metrics and after this I discussed the concept of the ROC the receiver operating characteristics this is one important curve that is the curve between the true positive rate and the false positive rate that is the receiver operating characteristics and the another parameter another important parameter is the area under the ROC curve which is independent of the classification threshold.

So this area under the ROC curve also you can determine and based on this you can compare the performance of the classification techniques. So this is about the performance evaluation of classifications. So let me stop here today. Thank you.