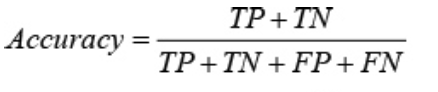
**Instructions**: Please complete and submit your work to the appropriate folder in LumiNUS. You may work in study groups, but each student must be responsible for their own submission.

Please submit the completed Word file named as StudentID-Name-HW4.docx (with all results).

1. Consider the metrics accuracy, precision, and recall.
   1. Give one example when accuracy would not be a good performance metric. Give a numerical example.
   2. Given one example of a supervised machine learning classification problem when higher precision is desired. Please give a different example than the ones given in class. This need not be a numerical example but must be clearly defined classification problem and dataset.
   3. Given one example of a supervised machine learning classification problem when higher recall is desired. Please give a different example than the ones given in class. This need not be a numerical example but must be clearly defined classification problem and dataset.
2. Suppose you are given the same test dataset and two binary classifiers. Give a numerical example such that Classifier 1 has higher accuracy than Classifier 2, but Classifier 2 has both higher precision and higher recall than Classifier 1? Hint: Give a hypothetical 2x2 confusion matrix for each classifier.
3. (a) Accuracy is simply defined as the fraction of the number of correct predictions to the total number of samples. By formula,



For instance, given a problem of classifying tumors as malignant (positive class) or benign (negative class) through a predictive model, we have the following result:

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion Matrix** | | **Actual Class** | |
| **Positive (Malignant)** | **Negative (Benign)** |
| **Predicted Class** | **Positive (Malignant)** | TP = 1 | FP = 1 |
| **Negative (Benign)** | FN = 8 | TN = 90 |

Using the given formula and values in the confusion matrix, accuracy is equal to 0.91. This gives us a high accuracy value. However, if we take a closer look at our model performance, results show otherwise. Out of the 9 actual cases of malignant tumors (TP+FN), the predictive model only successfully identifies 1 as malignant tumor (TP), whereby, 8 out of 9 malignant tumors are undiagnosed (FN). It can also be said that the recall of this predictive model is 1/9 = 0.111.

Although the accuracy of this predictive model is very high, it still produces a terrible outcome where 8 out of 9 malignant tumors are undiagnosed. Recall should have been considered in this case. Thus, accuracy alone would not be a good performance metric in this scenario.

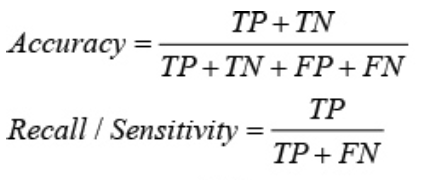
(b) One example of a supervised machine learning classification problem where higher precision is desired is in the case of spam email detection, where a spam email will be labelled as positive class and a non-spam email will be labelled as negative class. Precision will answer the following question: “If the predictive model says this email is spam, what is the chance that it’s actually a spam email?”. On the other hand, recall will address this question: “Out of all the actual spam emails, what fraction of the spam emails did the predictive model correct detects?” Certainly, precision is highly desirable here, whereby, precision is emphasized over recall. It is fine if an actual spam email is left undetected and doesn’t go into the spam folder. However, if an email is non-spam, it must not go into the spam folder. Precision is more imperative because if the predictive model predicts an email as spam, it better be inside the spam folder. Else, we may miss important emails. Thus, precision is highly desired in this situation.

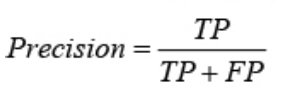
(c) One example of a supervised machine learning classification problem where higher recall is desired is in the clinical trial of a vaccine. A vaccine that results in side effects is labelled as positive class, while a vaccine that does not result in side effects is labelled negative class. In this case, recall is highly desirable because we must account for false negative cases. False negative cases refer to cases while the predictive model predicts the vaccine to have no side effects, but in reality, the vaccine has side effects. These side effects might be harmful to individual, or even worse, result in side effects that is contagious. Hence, recall is imperative and desired here, where we really want to ensure we are certainly capturing positive cases, as the cost of missing a positive case is more problematic than the cost of including a negative.

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion Matrix**  **for Classifier 1** | | **Actual Class** | |
| **Positive** | **Negative** |
| **Predicted Class** | **Positive** | TP = 20 | FP = 30 |
| **Negative** | FN = 28 | TN = 22 |

1. There are 100 data points.

|  |  |  |  |
| --- | --- | --- | --- |
| **Confusion Matrix**  **for Classifier 2** | | **Actual Class** | |
| **Positive** | **Negative** |
| **Predicted Class** | **Positive** | TP = 30 | FP = 42 |
| **Negative** | FN = 18 | TN = 10 |





Using the formula and confusion matrices above, the metrics for each classifier is calculated as show:

**Classifier 1**: Accuracy = 0.42

Recall = 0.416

Precision = 0.4

**Classifier 2**: Accuracy = 0.40

Recall = 0.625

Precision = 0.416

Thus, Classifier 1 has higher accuracy than Classifier 2, but Classifier 2 has both higher precision and higher recall than Classifier 1.