Evaluating Logistic Regression Models

This lesson will focus on how to evaluate logistic regression models.

WE'LL COVER THE FOLLOWING

Evaluating classification models

Just like there were many ways to evaluate linear regression models, there are many ways to evaluate the performance of classification models. Accuracy is one of the techniques. But it is not a sufficient metric alone. Why?

Think about a scenario where our model predicts a rare disease that is present only in 0.01% of the data. If our model always predicts that no disease is present, it will still be accurate 99% of the time but it would not diagnose correctly when it matters the most.

Classification report

A classification report is a table that calculates different metrics to evaluate our model. We can obtain the table using the function `classification_report` in `sklearn.metrics`

We will make the same model that we made in the last lesson.

```
import pandas as pd
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,classification_report

df = pd.read_csv('credit_card_cleaned.csv')
# Make data
X = df.drop(columns = ['default.payment.next.month','MARRIAGE','GENDER'])
Y = df[['default.payment.next.month']]
# Fit model
lr = LogisticRegression()
lr.fit(X,Y)
# Get predictions and accuracy
preds = lr.predict(X)
acc = accuracy_score(y_true = Y,y_pred = preds)

print('accuracy = ',acc)
print(classification_report(y_true = Y,y_pred = preds))
```

By printing the classification report we can see different metrics listed for both classes. Let's look at these one by one. But before that, we need to know the concept of positive class and a negative class. A **positive class** is the one in which we are interested. If we are interested in predicting customers that will not default, then `no` is the positive class and `yes` is the negative class. However, we can invert this and call `yes` a positive class and `no` the negative class. In our example above, `no` is the positive class. Some of the concepts associated with positive and negative classes are:

- **True Positive**: A true positive is an outcome where the model correctly predicts the positive class.
- **False Positive**: A false positive is an outcome where the actual class was the negative class, but the model predicts the positive class.
- **True Negative**: A true negative is an outcome where the model correctly predicts the negative class.
- **False Negative**: A false negative is an outcome where the actual class was the positive class but the model predicts the negative class.

Precision

Precision is a measure of how well the model performs when it predicts the positive class. It tells us that for all instances classified positive, what percent was correct. It is defined as:

$$precision = \frac{TP}{TP + FP}$$

Recall

Recall is a measure of how well the model predicts the positive class. It tells us that for all instances that were actually positive, what percent was classified correctly.

$$recall = \frac{TP}{TP + FN}$$

F1-Score

If precision increases, recall decreases and vice versa. Therefore, a new measure was introduced called F1-score. It is the harmonic mean of precision and recall. The best score is 1.0, whereas the worst score is 0.0.

$$F1 = 2\frac{precision * recall}{precision + recall}$$

Support

Support is the number of instances predicted in each class.

We need to look at all of these measures when evaluating our logistic regression models. The above model that we created using logistic regression does not do a very good job of predicting. The accuracy is 77%. Precision is on the lower side. Also, the F1-score is on the lower side, as well. A reason could be **imbalanced classes**. In our dataset, there are more negative examples than positive ones. This affects the outcome of the model, as well.

Multiclass classification

We can use the same code with different data for multi-class classification as well. The `fit` function will handle the details itself, and we would not need to do anything extra.

This brings an end to this lesson. In the next lesson, you will be given a challenge that you need to complete.