Example1: If the temperature is below 0 degrees, students will think it is cold and get up after 12 o'clock, otherwise students think it is hot and get up before 12 o'clock. Assuming that the indoor and outdoor temperatures are the same without air conditioning, if the air conditioning is on, the indoor temperature will increase by 10 degrees. I want to predict whether the students get up before 12 o’clock or not based on the following two features: outdoor temperature and whether the air conditioner is on.

The reason is that the target label has no linear correlation with the features. In such cases, logistic regression can’t predict targets with good accuracy.

Example2: Many modern companies and schools are using face recognition technology, but using logistic regression predictions are not accurate.

This is because face data is linearly indistinguishable and logistic regression requires the construction of decision boundaries to classify it.

1. The MLP has a back-propagation feature. The error after the output is used to estimate the error in the direct leading layer of the output layer, and then this error is used to estimate the error in the layer further ahead, and so on back-propagating layer by layer, the error estimates for all the other layers are obtained. The MLP thus uses the output error to improve the prediction accuracy of the model, whereas the kNN does not have this property.
2. The kNN requires the choice of the value of k. The choice of k has a large impact on the prediction results and there is no theoretical choice of the optimal k-value size, often the optimal k-value choice is obtained in combination with k-fold cross-validation. However the MLP does not need to choose parameters, and therefore the performance of the MLP is more stable than that of the kNN.
3. The MLP is fully connected and therefore requires a large number of parameters such as network topology, initial values for weights and thresholds, long training time and good classification results. kNN is simple in principle, short training time, but not as good as MLP, especially when the sample is not balanced, the prediction bias is larger. For example, there are fewer samples in one category and more samples in other categories.
4. A common advantage of MLP and kNN is their robustness to noise and fault tolerance, with no requirement for linear correlation of the data.

K-fold cross validation randomly splits the dataset into K copies at a time, one copy is kept for validating the model and the other K-1 copies are used for training. Because of the random feature of each selection, the data in a dataset is used repeatedly and each piece of data can be trained, which corresponds to obtaining as much valid information as possible from a limited amount of data and reduces the effect of noise in the data. When the amount of data is not large enough, it is easy to overfit the model if all the data is used to train the model. By dividing the data through K-fold cross validation and integrating the evaluation results, we can effectively reduce the variance in model selection. In other words, we expect the model to perform well on multiple subsets of the training set, rather than on the entire training data set alone. Therefore it evaluates the generalisation performance of a given algorithm after training on a given dataset and can reduce overfitting to a certain extent.

Let us first think of two edge cases.

1) The edge case where no cross-validation is used at all is when K = 1. At this point, all of the data is used for training, the model is prone to overfitting and therefore prone to being low bias and high variance.

2) The leave-one-out method is the other edge case of the K-fold, i.e. K = n. As the value of K increases, the variance in the evaluation of a single model gradually increases while the bias decreases. However, from the overall model perspective, the bias increases while the variance decreases.

So when the value of K wanders between 1 and n, it can be interpreted as a compromise between variance and bias. Taking K=10 as an example, at training time we have only 90% of the training data in the training set. In contrast to the case where no cross-validation is used, this makes the bias go up, but for the average of the results it reduces the model variance, and the final result is better or not depending on the degree of variation between the two. Therefore, as a rule of thumb, K = 5 or 10 performs well in most cases.

4.

In time series analysis, the lagged values of a series can be used as features to help predict the future values of the series. For example, consider a time series of the number of people entering and leaving Trinity University Dublin's West Gate. To use this time series to predict future numbers, we can construct an eigenmatrix using the lagged values of the temperature series.

Assume that the sequence of the number of people entering and exiting the West Gate is as follows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Day1 | Day2 | Day3 | Day4 | Day5 | Day6 | Day7 |
| 2013 | 2942 | 2502 | 2949 | 5021 | 3521 | 4452 |

We can create an eigenmatrix containing the lagged values of the sequence of people entering and leaving Westgate as follows.

|  |  |  |  |
| --- | --- | --- | --- |
| Day | Number (n) | Number(n-1) | Number(n-2) |
| 1 | 2013 | NA | NA |
| 2 | 2943 | 2013 | NA |
| 3 | 2502 | 2943 | 2013 |
| 4 | 2949 | 2502 | 2943 |
| 5 | 5021 | 2949 | 2502 |
| 6 | 3521 | 5021 | 2949 |
| 7 | 4452 | 3521 | 5021 |

In this example, the lagged values of the sequence of the number of people entering and leaving the West Gate are used as features to help predict the number of people entering and leaving the West Gate in the future. For example, the number of people entering and exiting the west gate on day 5 is 5021. the feature matrix shows that the number of people entering and exiting the west gate on the previous day was 2949 and the day before that was 2502. this information can be used to construct a model to predict the number of people entering and exiting the west gate on day 6.

Lagged values can be useful features for time series data as they capture some of the time dependence in the data. For example, in the series above, the number of people entering and exiting the west gate on day 5 may be influenced by the number of people entering and exiting the west gate on days 4 and 3. By using the lagged values as features, we can capture this time dependence in the model.