# Note

Feng-Yang Hsieh

## 1 Bootstrapping

The bootstrapping method involves an iterative process of training a classifier on mixed datasets. We start by training a classifier on the initial mixed datasets. The trained classifier is then used to reclassify the data, creating a new mixed training set. This process is repeated multiple times to increase the sample fraction differences in mixed datasets. We want to explore whether this process can improve CWoLa's performance.

We consider the events sampled from the normal distribution for the testing and implement this method. We found this method is unsuccessful. The model initially achieved the best performance, then worsened at subsequent iterations.

The reason is the reclassification step breaks the key assumption of the CWoLa approach: the signal and background events should have the same distributions in both mixed datasets. Figure 1 shows the initial signal and background distributions. Signal has the same distribution in mixed dataset $M_1$ and $M_2$, as does the background. Figure 2 shows signal and background distributions after the reclassification. We could observe the signal events have different distributions in $M_1$ and $M_2$. As a result, the assumption of the CWoLa approach is violated, leading to the bootstrapping method failure.

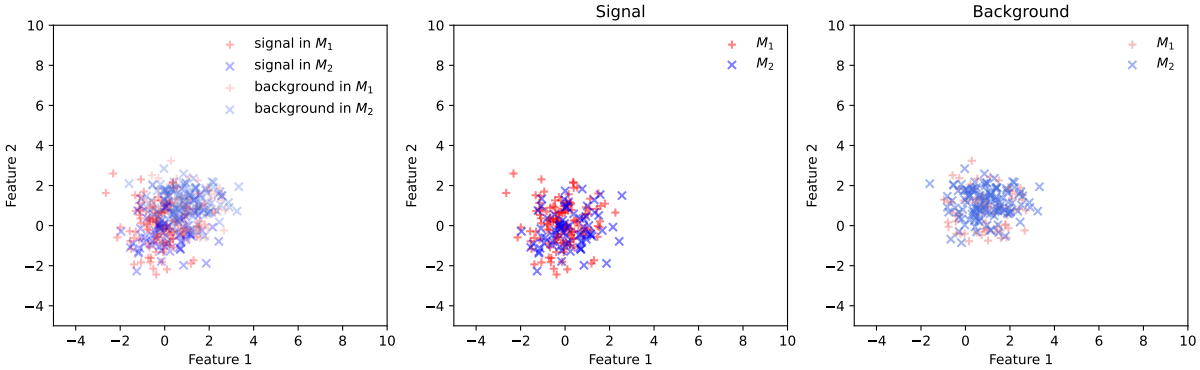


Figure 1: The signal and background samples distributions. The signal and background events are sampled from different two-dimensional normal distributions. They are randomly assigned to the mixed datasets $M_1$ or $M_2$.

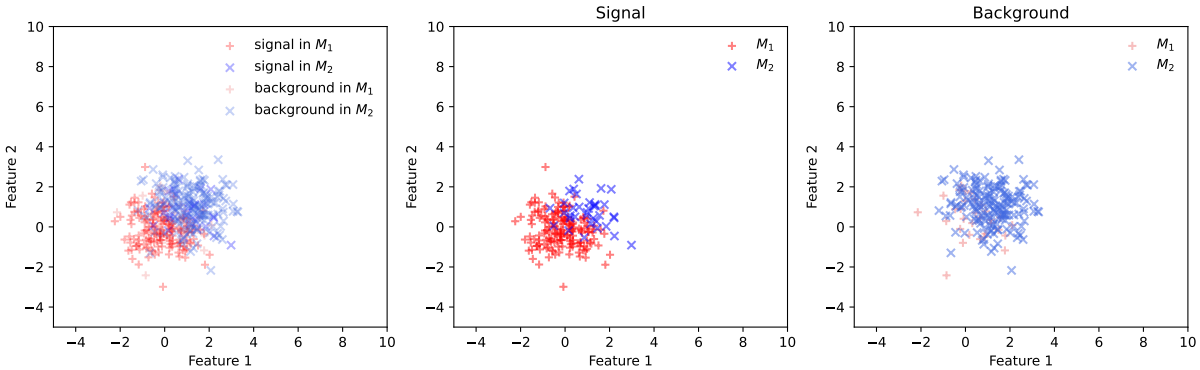


Figure 2: The signal and background samples distributions. The signal and background events are sampled from different two-dimensional normal distributions. They are assigned to the $M_1$ or $M_2$ by the trained classifier.