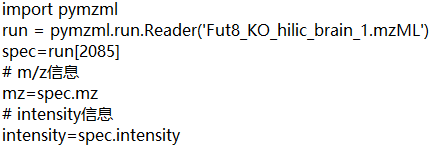
# 数据说明

1. 去除FUT8的鼠脑数据中不含核心岩藻糖  
   正常鼠脑数据中含有非核心岩藻糖和核心岩藻糖
2. mzML文件中包含原始谱图的信息，如谱图的scan号、m/z、intensity等。  
   xlsx文件中包含糖肽鉴定结果信息，如低能量谱图的scan号、糖链类型、是否核心岩藻糖等。
3. 读取mzML文件需要使用python库pymzml。  
   例如，读取Fut8\_KO\_hilic\_brain\_1.mzML文件中谱图scan号为2085的谱图。  
   

# 程序说明

CoreFucosylationModel.py文件中主要包含基于映射收敛算法的核心岩藻糖鉴定模型构建程序和基于自编码器的核心岩藻糖鉴定模型构建程序，具体函数介绍如下：

1. 函数MappingConvergence(labeled\_x, unlabeled\_x, filepath\_model)为基于映射收敛算法的核心岩藻糖鉴定模型构建程序。其中labeled\_x为已标记的数据集，unlabeled\_x为未标记的数据集，filepath\_model为模型保存的文件路径。
2. 函数AutoEncoder(train\_x, n\_encoder1, n\_encoder2, n\_decoder2, n\_decoder1, n\_latent=2)为基于自编码器的核心岩藻糖鉴定模型构建程序。其中train\_x为训练集，n\_encoder1,、n\_encoder2,、n\_decoder2,、n\_decoder1,和n\_latent=2为自编码器的各层的节点数量，函数返回已训练好的自编码器模型和阈值。
3. 函数AutoEncoderTest(reg, threshold, test\_x)为将已训练好的自编码器模型用于待测数据鉴定，其中reg为自编码器模型，threshold为阈值，test\_x为待测数据，函数返回鉴定结果。
4. 函数filter(x, filter\_type)为筛选数据集程序，其中x为数据集，filter\_type为筛选类型。
5. 主函数的程序主要是利用实际数据对模型进行训练，并验证训练好的模型，得到相关的实验结果。

# Demo

1. 通过训练自编码器模型，并验证训练好的模型，实验结果如图1和图2所示

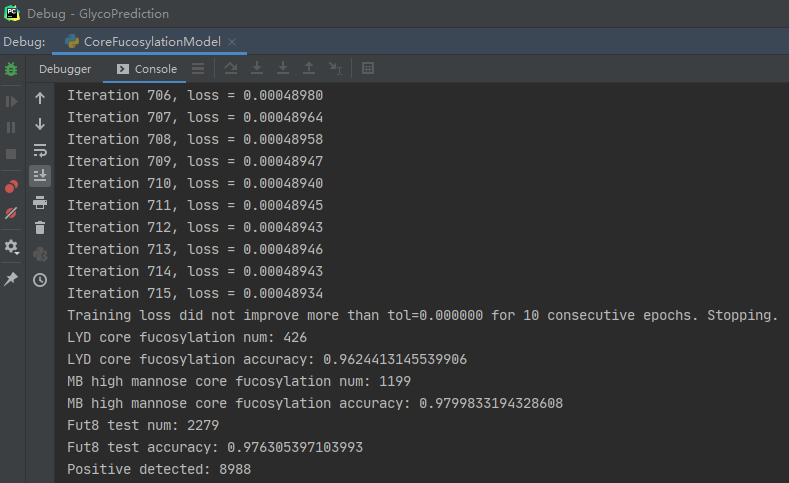


图1 基于自编码器的核心岩藻糖鉴定实验结果

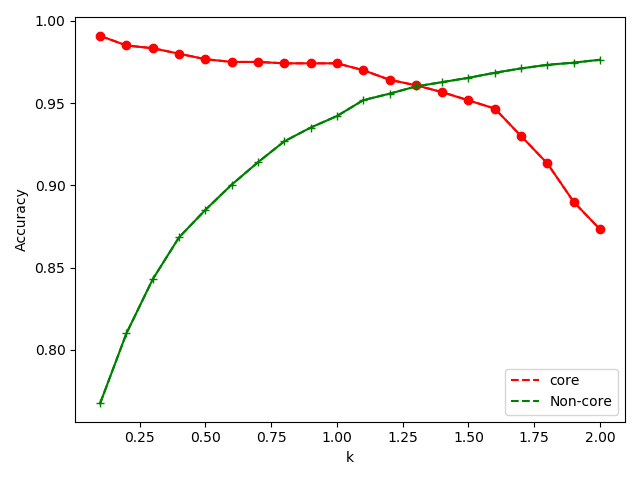


图2 参数k对于自编码器模型鉴定准确率的影响

1. 通过训练映射收敛算法模型，并验证训练好的模型，实验结果如图3

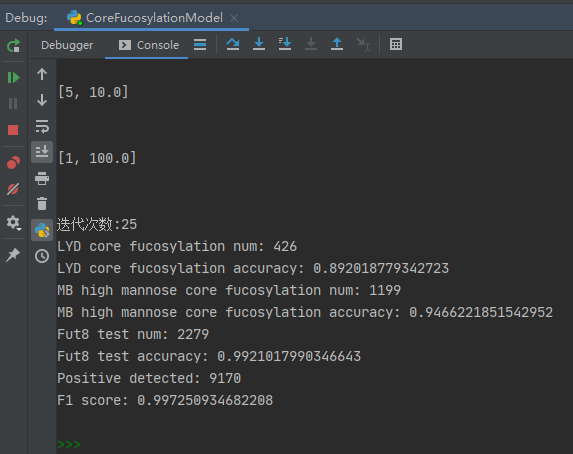


图3 基于映射收敛算法的核心岩藻糖鉴定实验结果