**Car Dataset:**

First, the dataset is loaded from the specified path and converted into a DataFrame format. Upon inspection, no missing values or anomalies were found in the dataset, so no special handling is required. The task type is then set to regression, and the label column strings are converted into numerical values in ascending order. Considering the dataset lacks purely numerical features, the string '5more' in the 5th column is replaced with the numerical value 5, converting the 5th column into a purely numerical feature. The dataset is then split into training, validation, and test sets according to a set ratio, with a fixed random seed to ensure reproducibility. Subsequently, features and labels are extracted from the processed dataset and saved into separate files according to their types, for example, saving numerical features as X\_num\_train.npy and categorical features as X\_cat\_test.npy. Finally, important information about the dataset is collected and saved into a JSON file.

**Autos Dataset:**

First, the dataset is loaded from the specified path and converted into a DataFrame format. Upon inspection, missing values were found in the dataset, which we replaced with -2024. The dataset also contains anomalous data; therefore, a Boolean index is created to remove the anomalies in the 14th column containing 'dohcv'. The task type is then set to regression. For label columns with negative values, they are converted to positive values in ascending order. Additionally, the 15th column representing the number of cylinders is converted from a string to a purely numerical feature to facilitate the regression task. The dataset is then split into training, validation, and test sets according to a set ratio, with a fixed random seed to ensure reproducibility. Subsequently, features and labels are extracted from the processed dataset and saved into separate files according to their types, for example, saving numerical features as X\_num\_train.npy and categorical features as X\_cat\_test.npy. Finally, important information about the dataset is collected and saved into a JSON file.

car数据集：首先从指定路径加载数据集，并将其转换为DataFrame格式，经检查数据集未发现任何缺失值或异常数据，因此无需进行特殊处理。随后将任务类型设置为回归，将标签列字符串按照评价从低到高的顺序转化为从小到大的数值，考虑到数据集中缺少纯数值特征，将第5列评分中的字符串’5more’替换为数值5，将第5列转变为纯数值特征。之后将数据集按比例划分为训练、验证和测试集，并设置了固定的随机种子保证实验的可复现性，随后从处理后的数据集中提取特征和标签，并根据它们的类型分别保存为不同的文件，例如将数值型特征保存为X\_num\_train.npy，将类别型特征保存为X\_cat\_test.npy。最后，收集有关数据集的重要信息，并将这些信息保存为JSON文件。

autos数据集： 首先从指定路径加载数据集，并转换为DataFrame格式，经检查数据集存在缺失，我们将缺失值替换为-2024，同时数据集中也存在异常数据，为此我们创建布尔索引，用于删除第14列中包含'dohcv'的异常数据。随后我们将任务类型设置为回归，针对标签列带有负数标签的情况，我们重新将其转化为从小到大的正数标签。此外，我们还将15列气缸数量从字符串转化为纯数值特征，方便回归任务进行。之后将数据集按比例划分为训练、验证和测试集，并设置了固定的随机种子保证实验的可复现性，随后从处理后的数据集中提取特征和标签，并根据它们的类型分别保存为不同的文件，例如将数值型特征保存为X\_num\_train.npy，将类别型特征保存为X\_cat\_test.npy。最后，收集有关数据集的重要信息，并将这些信息保存为JSON文件。