

```
import sys, csv
```

```
INPUT_FILE_GCOL_TRAIN = "ALL_AML_gcol.train.tmp.cutcomma"
INPUT_FILE_NAME_TRAIN = "ALL_AML_idclass.train.txt.cutright"
OUTPUT_FILE_NAME_TRAIN = "ALL_AML_gcol_class.train.csv"
INPUT_FILE_GCOL_TEST = "ALL_AML_gcol.test.tmp.cutcomma"
INPUT_FILE_NAME_TEST = "ALL_AML_idclass.test.txt.cutright"
OUTPUT_FILE_NAME_TEST = "ALL_AML_gcol_class.test.csv"
```

```
def stitchFiles(input_file_name_samples, input_file_name_id_class, output_file_name):
    #https://stackoverflow.com/questions/3277503/how-do-i-read-a-file-line-by-line-into-a-list
    with open(input_file_name_samples) as samplesFile:
        modified_lines = []

        #which will
        input_file_id_class = open(input_file_name_id_class)

        #open the file which will be the train output; we don't want to append
        out_file = open(output_file_name, "w")

        #create a list of lines, stripped of the newline
        sample_content = [line.rstrip('\n') for line in samplesFile]

        #create a list of "ID Class" strings, stripped of the newline
        id_class_content = [line.rstrip('\n') for line in input_file_id_class]

        print("len of sample_content: " + str(len(sample_content)))
        print("len of id_class_content: " + str(len(id_class_content)))

        i = 0

        for i in range(0, len(sample_content)):
            sample_line = sample_content[i]
            id_class_line = id_class_content[i]

            id_class_lst = id_class_line.split(" ")

            modified_line = id_class_lst[0] + "," + sample_line + "," + id_class_lst[1] + "\n"

            #join the stitched line
            modified_lines.append(modified_line)

        out_file.writelines(modified_lines)

    return modified_lines
```

```
modified_lines_train = stitchFiles(INPUT_FILE_GCOL_TRAIN, INPUT_FILE_NAME_TRAIN,
OUTPUT_FILE_NAME_TRAIN)
modified_lines_test = stitchFiles(INPUT_FILE_GCOL_TEST, INPUT_FILE_NAME_TEST,
OUTPUT_FILE_NAME_TEST)
```

```
# for sample_line, id_class_line in zip(sample_content,id_class_content):
    # # if sample_line == "" or id_class_line == "" or sample_line.isspace() or
    id_class_line.isspace():
        # # continue
    # modified_line = ""
    # i += 1
    # #print(str(i) + ": " + sample_line + "\n")

    # id_class_lst = id_class_line.split(" ")

    # modified_line = id_class_lst[0] + "," + sample_line + "," + id_class_lst[1] + "\n"

    # #join the stitched line
    # modified_lines.append(modified_line)

# # i = 0
# # for sample_line in sample_content:
#     # # id_class_line = id_class_content[i]
#     # # if sample_line == "" or id_class_line == "" or sample_line.isspace() or
#     id_class_line.isspace():
#         # # continue
#     # # modified_line = ""
#     # #print(str(i) + ": " + sample_line + "\n")

#     # # id_class_lst = id_class_line.split(" ")

#     # # modified_line = id_class_lst[0] + "," + sample_line + "," + id_class_lst[1] +
#     "\n"

#     # # #join the stitched line
#     # # modified_lines.append(modified_line)

# out_file.writelines(modified_lines)

# return modified_lines
```