

```

1 import os
2 import sys
3 import numpy as np
4
5 def numlines(filename) :
6     with open(filename) as fi:
7         for cnt, line in enumerate(fi):
8             n = cnt
9             n += 1
10    return n
11
12 def select_random_p_from_n(p, n) :
13     permutation = np.random.permutation(n)
14     k = int(p*n + 0.5)
15     firstk = permutation[:k]
16     return firstk
17
18 def select_random_p(p, filename) :
19     n = numlines(filename)
20     permutation = np.random.permutation(n)
21     k = int(p*n + 0.5)
22     firstk = permutation[:k]
23     return firstk
24
25 def select_from_file(f_in, f_out, selection) :
26     fo = open(f_out, "w");
27     with open(f_in) as fi:
28         for cnt, line in enumerate(fi):
29             if(cnt in selection) :
30                 fo.write(line)
31     fo.close()
32
33 def select_from_two_files(f1_in, f1_out, f2_in, f2_out, selection) :
34     select_from_file(f1_in, f1_out, selection)
35     select_from_file(f2_in, f2_out, selection)
36
37 def out_name(in_name,p,seed) :
38     s = str(seed) if seed >= 0 else ""
39     r = str(int(p*100))
40     basename = os.path.basename(in_name);
41     name, extension = os.path.splitext(basename)
42     return(name + "_" + s + "_" + r + extension)
43
44 # insist on 3 or 4 arguments
45 if len(sys.argv) != 4 and len(sys.argv) != 5 :
46     print(sys.argv[0], "takes 4 or 5 arguments. Not ", len(sys.argv)-1)
47     print("Arguments: file_x file_y fraction [seed]. Example:",
48           sys.argv[0], "training.txt testing.txt 0.3 7")
49     sys.exit()
50
51 f_x = sys.argv[1]
52 f_y = sys.argv[2]

```

```

53 p = float(sys.argv[3])
54 seed = -1
55 if(len(sys.argv) == 5) :
56     seed = int(sys.argv[4])
57     np.random.seed(seed)
58
59 lines_f_x = numlines(f_x)
60 lines_f_y = numlines(f_y)
61 assert lines_f_x == lines_f_y, "f_x,f_y must have same length"
62 selection = select_random_p_from_n(p, lines_f_x)
63
64 f_x_out = out_name(f_x,p,seed)
65 f_y_out = out_name(f_y,p,seed)
66
67 select_from_two_files(f_x, f_x_out, f_y, f_y_out, selection)

```