

```
In [1]: # imports  
import numpy as np
```

```
In [2]: # load the data  
arr = np.loadtxt(open('covtype.data', "rb"), delimiter=",")
```

```
In [3]: # find the largest class  
labels = arr[:, -1]  
counts = {  
    1: 0,  
    2: 0,  
    3: 0,  
    4: 0,  
    5: 0,  
    6: 0,  
    7: 0  
}  
for label in labels:  
    counts[label] += 1  
print(counts)  
  
{1: 211840, 2: 283301, 3: 35754, 4: 2747, 5: 9493, 6: 17367, 7: 20510}
```

```
In [4]: # largest class is 2  
# according to the paper, set the largest class to positive and the rest  
to negative  
for i in range(len(labels)):  
    arr[i, -1] = 1 if arr[i, -1] == 2 else 0
```

```
In [5]: # check that our labels are correct  
count = 0  
arr[:, -1]  
for label in arr[:, -1]:  
    count += 1 if label == 1 else 0  
assert count == counts[2]
```

```
In [6]: # save in npy format for use by classifiers  
np.save('covtype.npy', arr)
```