How to get the cumulative distribution function with NumPy?

Asked 8 years, 8 months ago Active 4 years, 4 months ago Viewed 87k times

33

12

I want to create a CDF with NumPy, my code is the next:

```
histo = np.zeros(4096, dtype = np.int32)
for x in range(0, width):
   for y in range(0, height):
      histo[data[x][y]] += 1
      q = 0
   cdf = list()
   for i in histo:
      q = q + i
      cdf.append(q)
```

I am walking by the array but take a long time the program execution. There is a built function with this feature, isn't?

python numpy histogram

Share Improve this question Follow

asked May 17 '12 at 17:44

omar

1,334 3 16 34

5 Answers

Active | Oldest | Votes

21

I'm not really sure what your code is doing, but if you have `hist` and `bin_edges` arrays returned by `numpy.histogram` you can use `numpy.cumsum` to generate a cumulative sum of the histogram contents.

```
>>> import numpy as np
>>> hist, bin_edges = np.histogram(np.random.randint(0,10,100), normed=True)
>>> bin_edges
array([ 0. ,  0.9,  1.8,  2.7,  3.6,  4.5,  5.4,  6.3,  7.2,  8.1,  9. ])
>>> hist
array([ 0.14444444,  0.11111111,  0.11111111,  0.1       ,  0.1       ,
        0.14444444,  0.14444444,  0.08888889,  0.03333333,  0.13333333])
>>> np.cumsum(hist)
array([ 0.14444444,  0.25555556,  0.36666667,  0.46666667,  0.56666667,
        0.71111111,  0.85555556,  0.94444444,  0.97777778,  1.11111111])
```

Share Improve this answer Follow

answered May 17 '12 at 19:15

user545424

14k 10 49 66

15 However, this introduces a binning step that would not be necessary for a cumulative distribution. – hans_meine Feb 20 '14 at 9:32

2 "This keyword, `normed` is deprecated in Numpy 1.6 due to confusing/buggy behavior. It will be removed in Numpy 2.0. "There is a bug in the code if bin is not in `[0,1]` . Add x=np.cumsum(hist); x=(x - x.min()) / x.ptp() – ArtificiallyIntelligence Jun 23 '16 at 19:29