DataArrays: Name that axis!

Fernando Pérez

Helen Wills Neuroscience Instiute, U.C. Berkeley

Caltech
August 21 2009

Example: model atmospheric Temperature

$$T = T(lat, lon, z)$$

```
In [12]: T = reshape(arange(12),(2,2,3))


In [14]: T[0]
Out[14]:
array([[0, 1, 2],
[3, 4, 5]])

In [15]: T[:,0]
Out[15]:
array([[0, 1, 2],
[6, 7, 8]])

In [26]: for z in range(T.shape[2]):
   ....:     plane = T[:,:,z]
   ....:     print 'mean temp', plane.mean()
   ....:
mean temp 4.5
mean temp 5.5
mean temp 6.5
```

A better way?

```
In [3]: T = DataArray(T, names = ['lat', 'lon', 'z'])

In [5]: T.ax_lat[0]
Out[5]:
DataArray([[0, 1, 2],
[3, 4, 5]])

In [6]: T.ax_lon[0]
Out[6]:
DataArray([[0, 1, 2],
[6, 7, 8]])

In [8]: for plane in T.ax_z:
   ...:     print 'mean temp', plane.mean()
   ...:
mean temp 4.5
mean temp 5.5
mean temp 6.5
```

How does it work? Axis objects

```
class Axis(object):
    """Object to access a given axis of an array.

    Key point: every axis contains a  reference to its parent array!
    """

    def __init__(self, name, index, arr):
        self.name = name
        self.index = index
        self.arr = arr

    def __getitem__(self, key):
        # Here is the real work
```

Now what?

The code

`http://github.com/fperez/datarray`

Our plans

- Get it to work fully (broadcasting, arithmetic, ufuncs, ...)
- Use it in production for a while (NIPY)
- If it works, propose it to NumPy (a la MaskedArray)

Thanks to

- **Matthew Brett, Ariel Rokem, Emi Nomura, Caterina Gratton** - UC Berkeley
- **Gael Varoquaux** - Neurospin (Orsay, France)
- **JB Poline** - Neurospin (Orsay, France)