Class SparseVector

source code

```
object --+
         |
        SparseVector
```

A simple sparse vector class for passing data to MLlib. Users may alternatively pass SciPy's {scipy.sparse} data types.

Instance Methods

	__init__(self, size, *args) Create a sparse vector, using either a dictionary, a list of (index, value) pairs, or two separate arrays of indices and values (sorted by index).	source code
	dot(self, other) Dot product with a SparseVector or 1- or 2-dimensional Numpy array.	source code
	squared_distance(self, other) Squared distance from a SparseVector or 1-dimensional NumPy array.	source code
	toArray(self) Returns a copy of this SparseVector as a 1-dimensional NumPy array.	source code
	__str__(self) str(x)	source code
	__repr__(self) repr(x)	source code
	__eq__(self, other) Test SparseVectors for equality.	source code
	__ne__(self, other)	source code

Inherited from object: __delattr__, __format__, __getattribute__, __hash__, __new__, __reduce__, __reduce_ex__, __setattr__, __sizeof__, __subclasshook__

Properties

Inherited from object: __class__

Method Details

*__init__(self, size, *args)* source code

(Constructor)

```
Create a sparse vector, using either a dictionary, a list of
(index, value) pairs, or two separate arrays of indices and
values (sorted by index).

@param size: Size of the vector.
@param args: Non-zero entries, as a dictionary, list of tupes,
       or two sorted lists containing indices and values.

>>> print SparseVector(4, {1: 1.0, 3: 5.5})
(4,[1,3],[1.0,5.5])
>>> print SparseVector(4, [(1, 1.0), (3, 5.5)])
(4,[1,3],[1.0,5.5])
>>> print SparseVector(4, [1, 3], [1.0, 5.5])
(4,[1,3],[1.0,5.5])
```

Overrides: object.__init__

dot(self, other)

source code

Dot product with a SparseVector or 1- or 2-dimensional Numpy array.

```
>>> a = SparseVector(4, [1, 3], [3.0, 4.0])
>>> a.dot(a)
25.0
>>> a.dot(array([1., 2., 3., 4.]))
22.0
>>> b = SparseVector(4, [2, 4], [1.0, 2.0])
>>> a.dot(b)
0.0
>>> a.dot(array([[1, 1], [2, 2], [3, 3], [4, 4]]))
array([ 22.,  22.])
```

squared_distance(self, other)

source code

Squared distance from a SparseVector or 1-dimensional NumPy array.

```
>>> a = SparseVector(4, [1, 3], [3.0, 4.0])
>>> a.squared_distance(a)
0.0
```

```
>>> a.squared_distance(array([1., 2., 3., 4.]))
11.0
>>> b = SparseVector(4, [2, 4], [1.0, 2.0])
>>> a.squared_distance(b)
30.0
>>> b.squared_distance(a)
30.0
```

source code

__str__(self)
(Informal representation operator)

str(x)

Overrides: object.__str__
(inherited documentation)

source code

__repr__(self)
(Representation operator)

repr(x)

Overrides: object.__repr__
(inherited documentation)

source code

__eq__(self, other)
(Equality operator)

Test SparseVectors for equality.

```
>>> v1 = SparseVector(4, [(1, 1.0), (3, 5.5)])
>>> v2 = SparseVector(4, [(1, 1.0), (3, 5.5)])
>>> v1 == v2
True
>>> v1 != v2
False
```