**Guys, copy the 3 files named :  
  
task12\_part1\_lr\_rdd  
housing.csv  
and the Tar file into /home/cloudera folder**

**Step 1 — Download the correct NumPy tarball on Windows**

Download this file (compatible with Python 2.6):  
🔗 NumPy 1.6.2 source tar.gz

Save it to your Windows machine (e.g., Desktop).

**Step 2 — Copy to your VM using WinSCP**

1. Open **WinSCP** → connect to your Cloudera VM.
2. In the left panel (Windows), navigate to where you saved numpy-1.6.2.tar.gz.
3. In the right panel (VM), navigate to /home/cloudera/.
4. Drag & drop numpy-1.6.2.tar.gz into /home/cloudera/.

👉 Now confirm it’s there by running in VM:

ls -l /home/cloudera/numpy-1.6.2.tar.gz

If you see the file, we’re good.

**Step 3 — Extract the tarball**

cd /home/cloudera

tar -xvzf numpy-1.6.2.tar.gz

This should create a folder:

/home/cloudera/numpy-1.6.2

**Step 4 — Install NumPy**

cd /home/cloudera/numpy-1.6.2

sudo python setup.py install

This will take some time as it compiles. If no errors → NumPy is installed.

1. Go back to your home directory:

cd ~

Now:

1. python -c "import numpy; print(numpy.\_\_version\_\_)"

If installation was successful, you should see:

1.6.2

👉 Once you confirm NumPy works, you can immediately run your Spark program:

spark-submit task12\_part1\_lr\_rdd.py