

Oct 05, 14 21:41

csc710sbse: HW5:Theisen

Page 1/1

```

#Structure from SA Lecture
import sys,re,random,math
sys.dont_write_bytecode = True

from options import *
from utils import *
from sk import *

myOpt = Options()

class Analyzer:
    n = 50
    old = [1 for i in range (0, n)]
    new = [1 for i in range (0, n)]
    era_lives = myOpt.era_lives;

    def bettered(self, new, old):
        def quartiles(value):
            return value*.25, value*.5, value*.75

        def betterifless():
            p1, median1, p3 = quartiles(new)
            IQR1=p3-p1
            p1, median2, p3 = quartiles(old)
            IQR2=p3-p1
            return median1<median2, IQR1<IQR2

        def same(): return al2(new, old)≤0.56

        betterMedian, betterIQR = betterifless()
        return betterMedian, betterIQR, same()

    def EraStop(self, lst):
        self.old = self.new
        self.new = lst
        out = False
        betterMedian = False
        betterIQR = False
        same = False

        #print self.old
        #print self.new
        oldQ1, oldMedian, oldQ3 = quartiles(self.old)
        newQ1, newMedian, newQ3 = quartiles(self.new)
        if newMedian < oldMedian:
            betterMedian = True
        if newQ3 - newQ1 < oldQ3 - oldQ1:
            betterIQR = True
        if al2(self.new, self.old) ≤ myOpt.al2_test:
            same = True

        #worsed
        if (same ^ ¬ betterIQR) ∨ (¬ same ^ ¬ betterMedian):
            out = False
        #bettered
        elif (¬ same ^ betterMedian):
            out = True

        if out:
            self.era_lives += 1
        else:
            self.era_lives -= 1
        if self.era_lives == 0:
            #print "Early Era Termination!"
            return True
        else:
            return False

#from menzies code
def median(lst,ordered=False):
    if ¬ ordered: lst= sorted(lst)
    n = len(lst)
    p = n//2
    if n % 2: return lst[p]
    q = p - 1
    q = max(0,min(q,n))
    return (lst[p] + lst[q])/2

def quartiles(lst):
    q1 = lst[int(len(lst)*.25)]
    med = median(lst, False)
    q3 = lst[int(len(lst)*.5)]
    return q1, med, q3

```

Oct 21, 14 13:36

csc710sbse: HW5:Theisen

Page 1/1

```
#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

class Options:
    #Globals
    debug = False
    seed = 1
    era_lives = 3
    al2_test = 0.6

    #MaxWalkSat options
    mws_prob = 0.75
    mws_maxTries = 500
    mws_maxChanges = 500
    mws_threshold = .000001
    mws_slices = 10

    #Simulated Annealing options
    sa_kmax = 500
    sa_cooling = .6

    #Genetic Algorithm options
    ga_gen_list = [10]
    ga_pop_size = 50
    ga_crossover = .6

    def printGlobals(self):
        print "Seed:", self.seed, "Lives: ", self.era_lives
        print "Al2test:", self.al2_test
```

Sep 23, 14 13:25

csc710sbse: HW5:Theisen

Page 1/1

```
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from sk import *

def rdiv8():
    rdivDemo([
        ["novels", 287, 332, 443, 711, 534],
        ["kids", 23, 18, 16, 20, 21],
        ["magazine", 112, 98, 43, 63, 82],
        ["cookbooks", 232, 180, 32, 53, 78],
    ])

rdiv8()
```

Oct 21, 14 14:34	csc710sbse: HW5:Theisen	Page 1/3
<pre> ===== Model Name: Fonseca Searcher Name: SA Seed: 1 Lives: 3 Al2test: 0.6 SA Options: KMAX: 500 Cooling: 0.6 Time to run (s): 2.916281 Runs: 5 Average per run (s): 0.5832562 (*), 0.83001, 0.87830, 0.93800, 0.96973, 0.97983 ===== Model Name: Fonseca Searcher Name: MWS Seed: 1 Lives: 3 Al2test: 0.6 MaxWalkSat Options: Prob: 0.75 MaxTries: 500 MaxChanges 500 Threshold: 1e-06 Slices: 10 Time to run (s): 4.12327 No valid runs! ===== Model Name: Fonseca Searcher Name: GA Seed: 1 Lives: 3 Al2test: 0.6 GA Options: popSize: 50 Crossover: 0.6 Gens: [10] Time to run (s): 16.090835 Runs: 5 Average per run (s): 3.218167 (*), 0.49977, 0.51859, 0.53717, 0.68139, 0.93113 ===== Scott-Knott for Fonseca rank , name , med , iqr ----- 1 , GA , 54 , 16 (- * --- -----) , 0.50, 0.52, 0.54, 0.68, 0.93 1 , SA , 94 , 9 (--- * --- -----) , 0.83, 0.88, 0.94, 0.97, 0.98 ===== Model Name: Schaffer Searcher Name: SA Seed: 1 Lives: 3 Al2test: 0.6 SA Options: KMAX: 500 Cooling: 0.6 Time to run (s): 0.138624 Runs: 5 Average per run (s): 0.0277248 (*), 0.00000, 0.00000, 0.00000, 0.00000, 0.00003 ===== Model Name: Schaffer Searcher Name: MWS Seed: 1 Lives: 3 Al2test: 0.6 MaxWalkSat Options: Prob: 0.75 MaxTries: 500 MaxChanges 500 Threshold: 1e-06 Slices: 10 Time to run (s): 0.151333 Runs: 1 Average per run (s): 0.151333 (*), 0.00000, 0.00000, 0.00000, 0.00000, 0.00000 ===== Model Name: Schaffer Searcher Name: GA Seed: 1 Lives: 3 Al2test: 0.6 GA Options: popSize: 50 Crossover: 0.6 Gens: [10] Time to run (s): 0.945538 Runs: 5 Average per run (s): 0.1891076 (*), 0.00000, 0.00000, 0.00000, 0.00003, 0.00006 ===== Scott-Knott for Schaffer rank , name , med , iqr ----- 1 , SA , 0 , 0 (* ----- -----) , 0.00, 0.00, 0.00, 0.00, 0.00 1 , MWS , 0 , 0 (* ----- -----) , 0.00, 0.00, 0.00, 0.00, 0.00 1 , GA , 0 , 0 (* ----- -----) , 0.00, 0.00, 0.00, 0.00, 0.00 ===== Model Name: Kursawe </pre>		

Oct 21, 14 14:34	csc710sbse: HW5:Theisen	Page 2/3
<pre> Searcher Name: SA Seed: 1 Lives: 3 Al2test: 0.6 SA Options: KMAX: 500 Cooling: 0.6 Time to run (s): 2.531 Runs: 5 Average per run (s): 0.5062 (*), 0.04636, 0.07428, 0.08698, 0.11616, 0.12743 ===== No Best Found for prob = 0.75 No Best Found for prob = 0.75 No Best Found for prob = 0.75 No Best Found for prob = 0.75 No Best Found for prob = 0.75 ===== Model Name: Kursawe Searcher Name: MWS Seed: 1 Lives: 3 Al2test: 0.6 MaxWalkSat Options: Prob: 0.75 MaxTries: 500 MaxChanges 500 Threshold: 1e-06 Slices: 10 Time to run (s): 3.537633 No valid runs! ===== Model Name: Kursawe Searcher Name: GA Seed: 1 Lives: 3 Al2test: 0.6 GA Options: popSize: 50 Crossover: 0.6 Gens: [10] Time to run (s): 13.808445 Runs: 5 Average per run (s): 2.761689 (*), 0.00074, 0.01386, 0.01645, 0.03094, 0.08662 ===== Scott-Knott for Kursawe rank , name , med , iqr ----- 1 , GA , 2 , 2 (-- * ----- -----) , 0.00, 0.01, 0.02, 0.03, 0.09 1 , SA , 9 , 5 (----- ----- * --) , 0.05, 0.07, 0.09, 0.12, 0.13 ===== Model Name: ZDT1 Searcher Name: SA Seed: 1 Lives: 3 Al2test: 0.6 SA Options: KMAX: 500 Cooling: 0.6 Time to run (s): 1.384426 Runs: 5 Average per run (s): 0.2768852 (*), 0.07102, 0.11667, 0.21154, 0.22530, 0.24614 ===== Model Name: ZDT1 Searcher Name: MWS Seed: 1 Lives: 3 Al2test: 0.6 MaxWalkSat Options: Prob: 0.75 MaxTries: 500 MaxChanges 500 Threshold: 1e-06 Slices: 10 Time to run (s): 1.416907 Runs: 5 Average per run (s): 0.2833814 (*), -0.05222, -0.03011, -0.02254, -0.00696, -0.00374 ===== Model Name: ZDT1 Searcher Name: GA Seed: 1 Lives: 3 Al2test: 0.6 GA Options: popSize: 50 Crossover: 0.6 Gens: [10] Time to run (s): 8.215808 Runs: 5 Average per run (s): 1.6431616 (*), 0.08534, 0.09136, 0.11372, 0.13352, 0.19169 ===== Scott-Knott for ZDT1 rank , name , med , iqr ----- 1 , MWS , -2 , 2 (-- * ----- -----) , -0.05, -0.03, -0.02, -0.01, -0.00 2 , GA , 11 , 4 (----- ----- * --) , 0.09, 0.09, 0.11, 0.13, 0.19 2 , SA , 21 , 11 (----- -----) , 0.07, 0.12, 0.21, 0.23, 0.25 </pre>		

Oct 21, 14 14:34

csc710sbse: HW5:Theisen

Page 3/3

```
=====
Model Name: ZDT3
Searcher Name: SA
Seed: 1 Lives: 3
Al2test: 0.6
SA Options:
KMAX: 500 Cooling: 0.6
Time to run (s): 1.50074
Runs: 5
Average per run (s): 0.300148
(* | ), -0.03604, 0.21725, 0.23301, 0.25875, 0.29595
=====
Model Name: ZDT3
Searcher Name: MWS
Seed: 1 Lives: 3
Al2test: 0.6
MaxWalkSat Options:
Prob: 0.75
MaxTries: 500 MaxChanges 500
Threshold: 1e-06 Slices: 10
Time to run (s): 1.599185
Runs: 5
Average per run (s): 0.319837
(* | ), -0.04109, -0.02913, -0.02180, -0.00308, -0.00266
=====
Model Name: ZDT3
Searcher Name: GA
Seed: 1 Lives: 3
Al2test: 0.6
GA Options:
popSize: 50 Crossover: 0.6
Gens: [10]
Time to run (s): 8.908521
Runs: 5
Average per run (s): 1.7817042
(* | ), 0.07100, 0.08388, 0.09071, 0.10566, 0.22608
=====
Scott-Knott for ZDT3

rank , name , med , iqr
-----
1 , MWS , -2 , 3 (-* | ----- ), -0.04, -0.03, -0.02, -0.00, -0.00
2 , GA , 9 , 3 ( --* -- | ----- ), 0.07, 0.08, 0.09, 0.11, 0.23
2 , SA , 23 , 4 (----- | ----- * --- ), -0.04, 0.22, 0.23, 0.26, 0.30
=====
Model Name: Viennet3
Searcher Name: SA
Seed: 1 Lives: 3
Al2test: 0.6
SA Options:
KMAX: 500 Cooling: 0.6
Time to run (s): 0.221635
Runs: 5
Average per run (s): 0.044327
(* | ), 0.00093, 0.00102, 0.00218, 0.00222, 0.00516
=====
Model Name: Viennet3
Searcher Name: MWS
Seed: 1 Lives: 3
Al2test: 0.6
MaxWalkSat Options:
Prob: 0.75
MaxTries: 500 MaxChanges 500
Threshold: 1e-06 Slices: 10
Time to run (s): 0.27679
No valid runs!
=====
Model Name: Viennet3
Searcher Name: GA
Seed: 1 Lives: 3
Al2test: 0.6
GA Options:
popSize: 50 Crossover: 0.6
Gens: [10]
Time to run (s): 1.426031
Runs: 5
Average per run (s): 0.2852062
(* | ), 0.00042, 0.00074, 0.00129, 0.00333, 0.00443
=====
Scott-Knott for Viennet3

rank , name , med , iqr
-----
1 , GA , 0 , 0 ( - * | ----- ), 0.00, 0.00, 0.00, 0.00, 0.00
1 , SA , 0 , 0 ( *--- | ----- ), 0.00, 0.00, 0.00, 0.00, 0.01
```

Oct 21, 14 13:08

csc710sbse: HW5:Theisen

Page 1/1

```

import sys
from datetime import datetime
import random

sys.dont_write_bytecode = True

from models import *
from searchers import *
from utils import *
from options import *
from sk import *

myOpt = Options()

#Inspired by vivekaxl's display function
def display(model, searcher, startTime, scores, r):
    print "=====
    print "Model Name: ", model.__name__
    print "Searcher Name: ", searcher.__class__.__name__
    diff = (datetime.now() - startTime).total_seconds()
    myOpt.printGlobals()
    searcher.printOptions()
    print "Time to run (s): ", diff
    if r == 0:
        print "No valid runs!"
    else:
        print "Runs: ", r
        print "Average per run (s): ", diff/r
        print "xtile(scores,width=25,show=" "%1.5f")
    print "=====

def main(modelList, searcherList):
    r = 5
    for klass in modelList:
        classScoreList = []
        for searcher in searcherList:
            fullScoreList = []
            startTime = datetime.now()
            scores = []
            mySearcher = searcher()
            random.seed(myOpt.seed)
            for _ in range(r):
                myKlass = klass()
                result, valid = mySearcher.run(myKlass)
                if valid == True:
                    scores.append(result)
            display(klass, mySearcher, startTime, scores, len(scores))
            fullScoreList.append(searcher.__name__)
            for x in scores:
                fullScoreList.append(x)
            if scores:
                classScoreList.append(fullScoreList)
        print "Scott-Knott for", klass.__name__
        print classScoreList
        rdivDemo(classScoreList)
        print "\n"

#modelList = [Fonseca, Schaffer, Kursawe, ZDT1, ZDT3, Viennet3]
#searcherList = [SA, MWS, GA]
modelList = [Fonseca]
searcherList = [GA]

main(modelList, searcherList)

```

Sep 27, 14 18:32

csc710sbse: HW5:Theisen

Page 1/6

```

"""
## Hypothesis Testing Stuff

### Standard Stuff

#### Standard Headers

"""
from __future__ import division
import sys, random, math
sys.dont_write_bytecode = True
"""

#### Standard Utils

"""
class o():
    "Anonymous container"
    def __init__(i,**fields) :
        i.override(fields)
    def override(i,d): i.__dict__.update(d); return i
    def __repr__(i):
        d = i.__dict__
        name = i.__class__.__name__
        return name+'{'+' '.join(['%s%s' % (k,pretty(d[k]))
                                for k in i.show()])+' '}'
    def show(i):
        return [k for k in sorted(i.__dict__.keys())
                if not k in i.__dict__['_']]
"""

Misc functions:

"""
rand = random.random
any = random.choice
seed = random.seed
exp = lambda n: math.e**n
ln = lambda n: math.log(n,math.e)
g = lambda n: round(n,2)

def median(lst,ordered=False):
    if not ordered: lst= sorted(lst)
    n = len(lst)
    p = n//2
    if n % 2: return lst[p]
    q = p - 1
    q = max(0,min(q,n))
    return (lst[p] + lst[q])/2

def msec(f):
    import time
    t1 = time.time()
    f()
    return (time.time() - t1) * 1000

def pairs(lst):
    "Return all pairs of items i,i+1 from a list."
    last=lst[0]
    for i in lst[1:]:
        yield last,i
        last = i

def xtile(lst,lo=0,hi=100,width=50,
          chops=[0.1, 0.3,0.5,0.7,0.9],
          marks=["-", " ", " ", " ", "-"],
          bar="|",star="*",show="%3.0f"):
    """The function _xtile_ takes a list of (possibly)
    unsorted numbers and presents them as a horizontal
    xtile chart (in ascii format). The default is a
    contracted quintile_ that shows the
    10,30,50,70,90 breaks in the data (but this can be
    changed-- see the optional flags of the function).
    """
    def pos(p) : return ordered[int(len(lst)*p)]
    def place(x) :
        return int(width*float((x - lo))/(hi - lo+0.00001))
    def pretty(lst) :
        return ' '.join([show % x for x in lst])
    ordered = sorted(lst)
    lo = min(lo,ordered[0])
    hi = max(hi,ordered[-1])
    what = [pos(p) for p in chops]
    where = [place(n) for n in what]
    out = [" "] * width
    for one,two in pairs(where):
        for i in range(one,two):
            out[i] = marks[0]
        marks = marks[1:]
    out[int(width/2)] = bar
    out[place(pos(0.5))]= star
    return '({'+''.join(out) + "},)" + pretty(what)

```

Tuesday October 21, 2014

./HW5/files/sk.py

Sep 27, 14 18:32

csc710sbse: HW5:Theisen

Page 2/6

```

def _xtileX() :
    import random
    random.seed(1)
    nums = [random.random()*2 for _ in range(100)]
    print xtile(nums,lo=0,hi=1.0,width=25,show=" %5.2f")
"""

### Standard Accumulator for Numbers

Note the _lt_ method: this accumulator can be sorted by median values.

Warning: this accumulator keeps _all_ numbers. Might be better to use
a bounded cache.

"""
class Num:
    "An Accumulator for numbers"
    def __init__(i,name,init=[1]):
        i.n = i.m2 = i.mu = 0.0
        i.all=[]
        i._median=None
        i.name = name
        i.rank = 0
        for x in init: i.add(x)
    def s(i) : return (i.m2/(i.n - 1))*0.5
    def add(i,x):
        i._median=None
        i.n += 1
        i.all += [x]
        delta = x - i.mu
        i.mu += delta*1.0/i.n
        i.m2 += delta*(x - i.mu)
    def __add__(i,j):
        return Num(i.name + j.name,i.all + j.all)
    def quartiles(i):
        def p(x) : return int(100*g(xs[x]))
        i.median()
        xs = i.all
        n = int(len(xs)*0.25)
        return p(n) , p(2*n) , p(3*n)
    def median(i):
        if not i._median:
            i.all = sorted(i.all)
            i._median=median(i.all)
        return i._median
    def __lt__(i,j):
        return i.median() < j.median()
    def spread(i):
        i.all=sorted(i.all)
        n1=i.n*0.25
        n2=i.n*0.75
        if len(i.all) <= 1:
            return 0
        if len(i.all) == 2:
            return i.all[1] - i.all[0]
        else:
            return i.all[int(n2)] - i.all[int(n1)]

"""

### The A12 Effect Size Test

"""
def al2slow(lst1,lst2):
    "how often is x in lst1 more than y in lst2?"
    more = same = 0.0
    for x in lst1:
        for y in lst2:
            if x == y : same += 1
            elif x > y : more += 1
    x= (more + 0.5*same) / (len(lst1)*len(lst2))
    return x

def al2(lst1,lst2):
    "how often is x in lst1 more than y in lst2?"
    def loop(t,t1,t2):
        while t1.j < t1.n ^ t2.j < t2.n:
            h1 = t1.l[t1.j]
            h2 = t2.l[t2.j]
            h3 = t2.l[t2.j+1] if t2.j+1 < t2.n else None
            if h1> h2:
                t1.j += 1; t1.gt += t2.n - t2.j
            elif h1 == h2:
                if h3 ^ h1 > h3 :
                    t1.gt += t2.n - t2.j - 1
                    t1.j += 1; t1.eq += 1; t2.eq += 1
            else:
                t2,t1 = t1,t2
            return t.gt*1.0, t.eq*1.0
    #-----
    lst1 = sorted(lst1, reverse=True)
    lst2 = sorted(lst2, reverse=True)
    n1 = len(lst1)
    n2 = len(lst2)

```

7/23

Sep 27, 14 18:32

csc710sbse: HW5:Theisen

Page 3/6

```

t1 = o(1=1st1,j=0,eq=0,gt=0,n=n1)
t2 = o(1=1st2,j=0,eq=0,gt=0,n=n2)
gt,eq= loop(t1, t1, t2)
return gt/(n1*n2) + eq/2/(n1*n2)

def _a12():
def f1(): return a12slow(l1,l2)
def f2(): return a12(l1,l2)
for n in [100,200,400,800,1600,3200,6400]:
    l1 = [rand() for _ in xrange(n)]
    l2 = [rand() for _ in xrange(n)]
    t1 = msecs(f1)
    t2 = msecs(f2)
    print n, g(f1()),g(f2()),int((t1/t2))

```

***Output:

....

n a12(fast) a12(slow) tfast / tslow

n	a12(fast)	a12(slow)	tfast / tslow
100	0.53	0.53	4
200	0.48	0.48	6
400	0.49	0.49	28
800	0.5	0.5	26
1600	0.51	0.51	72
3200	0.49	0.49	109
6400	0.5	0.5	244

....

Non-Parametric Hypothesis Testing

The following `_bootstrap_` method was introduced in 1979 by Bradley Efron at Stanford University. It was inspired by earlier work on the jackknife. Improved estimates of the variance were [developed later][efron01].

[efron01]: <http://goo.gl/14n8WF> "Bradley Efron & R.J. Tibshirani. An Introduction to the Bootstrap (Chapman & Hall/CRC M

To check if two populations `(y0,z0)_` are different, many times sample with replacement from both to generate `(y1,z1), (y2,z2), (y3,z3)_...` etc.

```

***
def sampleWithReplacement(lst):
    "returns a list same size as lst"
    def any(n): return random.uniform(0,n)
    def one(lst): return lst[ int(any(len(lst))) ]
    return [one(lst) for _ in lst]
***

```

Then, for all those samples, check if some `*testStatistic*` in the original pair hold for all the other pairs. If it does more than (say) 99% of the time, then we are 99% confident in that the populations are the same.

In such a `_bootstrap_` hypothesis test, the `*some property*` is the difference between the two populations, muted by the joint standard deviation of the populations.

```

***
def testStatistic(y,z):
    """Checks if two means are different, tempered
    by the sample size of 'y' and 'z'"""
    tmp1 = tmp2 = 0
    for y1 in y.all: tmp1 += (y1 - y.mu)**2
    for z1 in z.all: tmp2 += (z1 - z.mu)**2
    s1 = (float(tmp1)/(y.n - 1))**0.5
    s2 = (float(tmp2)/(z.n - 1))**0.5
    delta = z.mu - y.mu
    if s1+s2:
        delta = delta/((s1/y.n + s2/z.n)**0.5)
    return delta
***

```

The rest is just details:

- + Efron advises to make the mean of the populations the same (see the `_yhat,zhat_` stuff shown below).
- + The class `_total_` is a just a quick and dirty accumulation class.
- + For more details see [the Efron text][efron01].

```

***
def bootstrap(y0,z0,conf=0.01,b=1000):
    """The bootstrap hypothesis test from
    p220 to 223 of Efron's book 'An
    introduction to the bootstrap.'"""
    class total():
        "quick and dirty data collector"

```

Sep 27, 14 18:32

csc710sbse: HW5:Theisen

Page 4/6

```

def __init__(i,some=[]):
    i.sum = i.n = i.mu = 0 ; i.all=[]
    for one in some: i.put(one)
    def put(i,x):
        i.all.append(x);
        i.sum +=x; i.n += 1; i.mu = float(i.sum)/i.n
    def _add_(i1,i2): return total(i1.all + i2.all)
y, z = total(y0), total(z0)
x = y + z
tobs = testStatistic(y,z)
yhat = [y1 - y.mu + x.mu for y1 in y.all]
zhat = [z1 - z.mu + x.mu for z1 in z.all]
bigger = 0.0
for i in range(b):
    if testStatistic(total(sampleWithReplacement(yhat)),
                        total(sampleWithReplacement(zhat))) > tobs:
        bigger += 1
return bigger / b < conf
***

```

Examples

```

***
def _bootstraped():
def worker(n=1000,
           mu1=10, sigmal=1,
           mu2=10.2, sigma2=1):
    def g(mu,sigma): return random.gauss(mu,sigma)
    x = [g(mu1,sigmal) for i in range(n)]
    y = [g(mu2,sigma2) for i in range(n)]
    return n,mu1,sigmal,mu2,sigma2,\
        'different' if bootstrap(x,y) else 'same'
# very different means, same std
print worker(mu1=10, sigmal=10,
             mu2=100, sigma2=10)
# similar means and std
print worker(mu1= 10.1, sigmal=1,
             mu2= 10.2, sigma2=1)
# slightly different means, same std
print worker(mu1= 10.1, sigmal= 1,
             mu2= 10.8, sigma2= 1)
# different in mu eater by large std
print worker(mu1= 10.1, sigmal= 10,
             mu2= 10.8, sigma2= 1)
***

```

Output:

```

....
_bootstraped()

(1000, 10, 10, 100, 10, 'different')
(1000, 10.1, 1, 10.2, 1, 'same')
(1000, 10.1, 1, 10.8, 1, 'different')
(1000, 10.1, 10, 10.8, 1, 'same')
....

```

Warning-- the above took 8 seconds to generate since we used 1000 bootstraps. As to how many bootstraps are enough, that depends on the data. There are results saying 200 to 400 are enough but, since I am suspicious man, I run it for 1000.

Which means the runtimes associated with bootstrapping is a significant issue. To reduce that runtime, I avoid things like an all-pairs comparison of all treatments (see below: Scott-knott). Also, BEFORE I do the bootstrap, I first run the effect size test (and only go to bootstrapping in effect size passes:

```

***
def different(l1,l2):
    #return bootstrap(l1,l2) and a12(l2,l1)
    return a12(l2,l1) ^ bootstrap(l1,l2)
***

```

Saner Hypothesis Testing

The following code, which you should use verbatim does the following:

- + All treatments are clustered into `_ranks_`. In practice, dozens of treatments end up generating just a handful of ranks.
- + The numbers of calls to the hypothesis tests are minimized:
 - + Treatments are sorted by their median value.
 - + Treatments are divided into two groups such that the expected value of the mean values `_after_` the split is minimized;
- + Hypothesis tests are called to test if the two groups are truly difference.
 - + All hypothesis tests are non-parametric and include (1) effect size tests and (2) tests for statistically significant numbers;
 - + Slow bootstraps are executed if the faster `_A12_` tests are passed;

In practice, this means that the hypothesis tests (with confidence of say, 95%) are called on only a logarithmic number of times. So...

- + With this method, 16 treatments can be studied using less than $\sum_{i=1}^{16} \log_2 i = 15$ hypothesis tests and confidence $0.99 < \sup$
- + But if did this with the 120 all-pairs comparisons of the 16 treatments, we would have total confidence $0.99 < \sup_{i=1}^{120} 0.99^i < 0.30$.

Sep 23, 14 1:52

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from options import *

#Taken verbatim from the class website.

def pairs(lst):
    last=lst[0]
    for i in lst[1:]:
        yield last,i
        last = i

def xtile(lst, lo=0, hi=0.001, width=50,
         chops=[0.1, 0.3, 0.5, 0.7, 0.9],
         marks=["-", " ", " ", " ", "-"],
         bar="|", star="*", show="%3.0f"):
    """The function _xtile_ takes a list of (possibly)
    unsorted numbers and presents them as a horizontal
    xtile chart (in ascii format). The default is a
    contracted _quintile_ that shows the
    10,30,50,70,90 breaks in the data (but this can be
    changed- see the optional flags of the function).
    """
    def pos(p): return ordered[int((len(lst)*p))]
    def place(x):
        return int(width*float((x - lo))/(hi - lo))
    def pretty(lst):
        return ','.join([show % x for x in lst])
    ordered = sorted(lst)
    lo = min(lo, ordered[0])
    hi = max(hi, ordered[-1])
    what = [pos(p) for p in chops]
    where = [place(n) for n in what]
    out = [" "] * width
    for one, two in pairs(where):
        for i in range(one, two):
            out[i] = marks[0]
    marks = marks[1:]
    out[int(width/2)] = bar
    out[place(pos(0.5))] = star
    return ''.join(out) + " " + pretty(what)

```

Sep 29, 14 20:29

csc710sbse: HW5:Theisen

Page 1/1

```
from sim_anneal import *  
from max_walk_sat import *  
from genetic_alg import *
```

Oct 21, 14 13:17

csc710sbse: HW5:Theisen

Page 1/1

```

#Structure from SA Lecture
import sys, re, random, math
import copy
sys.dont_write_bytecode = True

from options import *
from utils import *
from analyzer import *

import itertools

myOpt = Options()

#Structure from:
#http://www.cleveralgorithms.com/nature-inspired/evolution/genetic_algorithm.html
class GA:

    def Mutate(self, child, pMut):
        for x in range(0, len(child.XVar)):
            if random.random() < pMut:
                child.specificMutate(x)
        return child

    #Based on WeiFuo's crossover
    def Crossover(self, parent1, parent2, crossovers):
        def what(lst):
            return lst[0] if isinstance(lst, list) else lst
        child = parent1
        if rand() > myOpt.ga_crossover:
            return parent1
        else:
            index = sorted([random.randint(0, parent1.n - 1) for _ in xrange(crossovers)])
            child.XVar = parent1.XVar[:]
            child.XVar[index[0]:index[1]] = parent2.XVar[index[0]:index[1]]
            return child

    def SelectParents(self, pop): #all possible parents
        temp = []
        for x in pop:
            for y in pop:
                if x != y:
                    temp.append([x, y])
        return temp

    def GetBestSolutions(self, pop):
        eMin = 1
        temp = []
        temp2 = []
        seen = []
        for x in pop:
            if x.Energy() not in seen:
                temp.append([x.Energy(), x])
                seen.append(x.Energy())
        temp.sort(key = lambda x: x[0])
        for x in range(0, 50):
            temp2.append(temp[x][1])
        #print temp[x]
        return temp[0][0], temp2

    def run(self, klass):
        ga = klass
        pMutate = 1.0/len(ga.XVar)
        XVarBest = ga.XVar
        #print 'start energy: ', eBest
        k = 1
        eList = []
        population = []
        analyze = Analyzer()
        stop = False

        for i in range(myOpt.ga_pop_size):
            ga.Chaos()
            population.append(copy.deepcopy(ga)) #add a randomly generated model to list

        for gens in myOpt.ga_gen_list:
            eBest = 1
            while k < gens:
                parents = self.SelectParents(population) #all possible pairings of parents
                children = []
                for parent1, parent2 in parents:
                    child = self.Crossover(parent1, parent2, 2)
                    children.append(self.Mutate(child, pMutate))
                eBest, population = self.GetBestSolutions(children)
                k += 1
                eList.append(eBest)
            #some "is significantly better" termination logic here

        return min(eList), True

    def printOptions(self):
        print "GA Options:"
        print "popSize:", myOpt.ga_pop_size, "Crossover:", myOpt.ga_crossover
        print "Gens:", myOpt.ga_gen_list

```

Oct 05, 14 21:40

csc710sbse: HW5:Theisen

Page 1/1

```

#Structure from SA Lecture
import sys, random, math
sys.dont_write_bytecode = True

from options import *
from utils import *
from analyzer import *

myOpt = Options()

class MWS:
    debug = False

    def say(self, x):
        if self.debug:
            sys.stdout.write(str(x)); sys.stdout.flush()

    def specificRun(self, probability, klass):
        fon = klass
        XVarBest = fon.XVar
        eBest = e = 1
        eNew = 1
        k = 1
        temp = []
        self.say(int(math.fabs(eBest-1)*100))
        self.say(' ')

        analyze = Analyzer()
        stop = False

        for i in xrange(myOpt.mws_maxTries):
            fon.Chaos()
            for j in xrange(myOpt.mws_maxChanges):
                eNew = fon.Energy()
                if (eNew < myOpt.mws_threshold v stop == True):
                    # means found a solution and quit
                    self.say('%')
                    eBest = eNew
                    XVarBest = list(fon.XVar)
                    temp.append(eNew)
                    #print xtile(temp, lo=0, hi=1, width=25, show=" %1.5f")
                    return eBest, XVarBest
                else:
                    #modify random part of solution
                    if probability < random.uniform(0,1):
                        fon.Neighbor()
                        self.say(' ')
                    #maximize for some random
                    else:
                        fon.BestNeighbor()
                        self.say('.')
                        temp.append(eNew)
                        if (i+1)*(j+1) % 40 == 0 ^ len(temp) != 0:
                            #print ' '
                            self.say(int(math.fabs(eNew-1)*100))
                            self.say(' ')
                            #print xtile(temp, lo=0, hi=1, width=25, show=" %1.5f")
                            #stop = analyze.EraStop(temp)
                            temp = []
            return -1, XVarBest

    def run(self, klass):
        theBest = -1
        valid = False
        eBest, XVarBest = self.specificRun(myOpt.mws_prob, klass)
        if eBest == -1:
            print 'No Best Found for prob = ', myOpt.mws_prob
            self.say(' ')
        else:
            theBest = eBest
            valid = True
        return theBest, valid

    def printOptions(self):
        print "MaxWalkSat Options:"
        print "Prob:", myOpt.mws_prob
        print "MaxTries:", myOpt.mws_maxTries, "MaxChanges", myOpt.mws_maxChanges
        print "Threshold:", myOpt.mws_threshold, "Slices:", myOpt.mws_slices

```

Oct 05, 14 21:41

csc710sbse: HW5:Theisen

Page 1/1

```

#Structure from SA Lecture
import sys, random, math
sys.dont_write_bytecode = True

from options import *
from utils import *
from analyzer import *

myOpt = Options()

class SA:

    def say(self, x):
        if myOpt.debug:
            sys.stdout.write(str(x)); sys.stdout.flush()

    def run(self, klass):
        sa = klass
        XVarBest = sa.XVar
        eBest = e = 1
        #print 'start energy: ', eBest
        k = 1
        temp = []
        self.say(int(math.fabs(eBest-1)*100))
        self.say(' ')
        analyzer = Analyzer()
        stop = False

        while k < myOpt.sa_kmax ^ stop == False:
            sa.Neighbor()
            eNew = sa.Energy()
            if eNew < eBest:
                eBest = eNew
                XVarBest = list(sa.XVar)
                self.say('!!')

            if eNew < e:
                e = eNew
                self.say('+')
            #Probability Check from SA Lecture
            elif math.exp(-1*(eNew-e)/(k/myOpt.sa_kmax*myOpt.sa_cooling)) < random.uniform(0,1):
                #P function should be between 0 and 1
                #more random hops early, then decreasing as time goes on
                sa.Chaos()
                self.say('?.')
                self.say('.')
                k = k + 1
                temp.append(eBest)
            if k % 50 == 0 ^ k != myOpt.sa_kmax ^ len(temp) != 0:
                self.say(int(math.fabs(eBest-1)*100))
                self.say(' ')
                #print xtile(temp,lo=0, hi=1,width=25,show=" %1.5f")
                stop = analyzer.EraStop(temp)
                temp = []

        if myOpt.debug:
            #print '\nFound best - e: ', eBest
            #print 'Variables: '
            for vars in XVarBest:
                self.say(vars)
                self.say(",")
            #print "\n"
            return eBest, True

    def printOptions(self):
        print "SA Options:"
        print "KMAX:", myOpt.sa_kmax, "Cooling:", myOpt.sa_cooling

```

Oct 05, 14 19:22

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

class DTLZ7(Model):
    smin = -4.0
    smax = 4
    n = 20
    XVar = [random.uniform(smin, smax) for i in range(0, n)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        X = self.XVar
        n = self.n
        f1 = 1 - math.exp(-np.sum([(X[i] - 1/math.sqrt(i+1))**2 for i in range(0, n)]))
        f2 = 1 - math.exp(-np.sum([(X[i] + 1/math.sqrt(i+1))**2 for i in range(0, n)]))
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        X = self.XVar
        n = self.n
        f1 = 1 - math.exp(-np.sum([(X[i] - 1/math.sqrt(i+1))**2 for i in range(0, n)]))
        f2 = 1 - math.exp(-np.sum([(X[i] + 1/math.sqrt(i+1))**2 for i in range(0, n)]))
        return (f1+f2)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```

Oct 05, 14 19:23

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

class ZDT1(Model):
    smin = 0
    smax = 1
    n = 30
    XVar = [random.uniform(smin, smax) for i in range (0, n)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        X = self.XVar
        f1 = X[0]
        g = 1+9*(np.sum([X[i] for i in range (1, self.n)]/(self.n-1))
        f2 = g*(1-np.sqrt(X[0]/g))
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        X = self.XVar
        f1 = X[0]
        g = 1+9*(np.sum([X[i] for i in range (1, self.n)]/(self.n-1))
        f2 = g*(1-np.sqrt(X[0]/g))
        return (f1+f2)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```


Oct 05, 14 20:28

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

class ZDT3(Model):
    smin = 0
    smax = 1
    n = 30
    XVar = [random.uniform(smin, smax) for i in range (0, n)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        X = self.XVar
        f1 = X[0]
        g = 1+9*(np.sum([X[i] for i in range (1, self.n)]))/(self.n-1)
        f2 = g*(1-np.sqrt(X[0]/g)-(X[0]/g)*math.sin(10*math.pi*X[0]))
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        X = self.XVar
        f1 = X[0]
        g = 1+9*(np.sum([X[i] for i in range (1, self.n)]))/(self.n-1)
        f2 = g*(1-np.sqrt(X[0]/g)-(X[0]/g)*math.sin(10*math.pi*X[0]))
        return (f1+f2)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```

Sep 29, 14 21:48

csc710sbse: HW5:Theisen

Page 1/1

```
from fonsaca_model import *  
from schaffer_model import *  
from kursawe_model import *  
from ZDT1_model import *  
from ZDT3_model import *  
from viennet3_model import *  
from DTLZ7_model import *
```

Oct 05, 14 20:05

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

class Fonseca(Model):
    n = 3
    smin = -4
    smax = 4
    XVar = [random.uniform(smin, smax) for i in range (0, 3)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        n = self.n
        f1 = 1-math.exp(-np.sum([self.XVar[i]-(1/np.sqrt(n))**2 for i in range (0, 3)]))
        f2 = 1-math.exp(-np.sum([self.XVar[i]+(1/np.sqrt(n))**2 for i in range (0, 3)]))
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        n = self.n
        f1 = 1-math.exp(-np.sum([self.XVar[i]-(1/np.sqrt(n))**2 for i in range (0, 3)]))
        f2 = 1-math.exp(-np.sum([self.XVar[i]+(1/np.sqrt(n))**2 for i in range (0, 3)]))
        return f1+f2

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```

Oct 05, 14 20:05

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

myOpt = Options()

class Kursawe(Model):
    n = 3
    smin = -5
    smax = 5
    XVar = [random.uniform(smin, smax) for i in range (0, 3)]
    XVarMax = XVar
    eMax = 0
    eMin = 0
    a = 0.8
    b = 3

    def Energy(self):
        X = self.XVar
        f1 = np.sum([-10*math.exp(-0.2*(np.sqrt(X[i]**2+X[i]**2))) for i in range (0, 3-1)])
        f2 = np.sum([math.fabs(X[i])**self.a + 5*np.sin(X[i])**self.b for i in range (0, 3)])
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        X = self.XVar
        f1 = np.sum([-10*math.exp(-0.2*(np.sqrt(X[i]**2+X[i]**2))) for i in range (0, 3-1)])
        f2 = np.sum([math.fabs(X[i])**self.a + 5*np.sin(X[i])**self.b for i in range (0, 3)])
        return (f1+f2)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```

Oct 05, 14 21:36

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from options import *
myOpt = Options()
rand = random.random

class Model:
    #Default Values overwritten by subclass; should have better defaults, but...
    n = 50
    smin = 1
    smax = 1
    XVar = [random.uniform(smin, smax) for i in range(0, n)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        raise NotImplementedError()

    def RawEnergy(self):
        raise NotImplementedError()

    def Neighbor(self):
        self.XVar[random.randint(0, self.n-1)] = random.uniform(self.smin, self.smax)

    def specificMutate(self, x):
        self.XVar[x] = random.uniform(self.smin, self.smax)

    def BestNeighbor(self):
        toChange = random.randint(0, self.n-1)
        toIncrement = (self.smax - self.smin) / myOpt.mws_slices
        curMax = 1
        maxVal = self.XVar[toChange]
        for i in xrange(myOpt.mws_slices):
            self.XVar[toChange] = self.smin + toIncrement
            x = self.Energy()
            if x < curMax:
                curMax = x
                maxVal = self.XVar[toChange]

    def Reset(self):
        self.XVar = self.XVarMax

    def Chaos(self):
        for vars in self.XVar:
            vars = random.uniform(self.smin, self.smax)

    def Baseline(self, numRuns):
        self.Chaos()
        self.eMax = self.eMin = self.RawEnergy()
        runs = 1
        while runs < numRuns:
            self.Neighbor()
            eNew = self.RawEnergy()
            if eNew > self.eMax: #find largest difference
                self.eMax = eNew
                self.XVarMax = self.XVar
                #print self.XVarMax, eNew
            if eNew < self.eMin: #find smallest difference
                self.eMin = eNew
                #print 'Min: ', self.XVar, eNew
            runs += 1
        #print 'Baseline: ', self.eMin, ' ', self.eMax

    def __init__(self):
        raise NotImplementedError()

```

Oct 05, 14 19:20

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np

from model_base import *
from options import *

sys.dont_write_bytecode = True

class Schaffer(Model):
    n = 1
    smin = -10
    smax = 10
    XVar = [random.uniform(smin, smax) for i in range (0, 1)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        f1 = self.XVar[0]*self.XVar[0]
        f2 = (self.XVar[0]-2)*(self.XVar[0]-2)
        return ((f1+f2) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        f1 = self.XVar[0]*self.XVar[0]
        f2 = (self.XVar[0]-2)*(self.XVar[0]-2)
        return (f1+f2)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```

Oct 05, 14 20:29

csc710sbse: HW5:Theisen

Page 1/1

```

#From Class Discussion 8/26/2014
from __future__ import division
import sys, re, random, math
import numpy as np
sys.dont_write_bytecode = True

from model_base import *
from options import *

class Viennet3(Model):
    smin = -3.0
    smax = 3
    n = 2
    XVar = [random.uniform(smin, smax) for i in range(0, n)]
    XVarMax = XVar
    eMax = 0
    eMin = 0

    def Energy(self):
        X = self.XVar
        f1 = 0.5*X[0]**2 + X[1]**2 + math.sin(X[0]**2+X[1]**2)
        f2 = (3*X[0]-2*X[1]+4)**2/8 + (X[0]-X[1]+1)**2/27 + 15
        f3 = 1/(X[0]+X[1]+1) - 1.1*math.e**(-X[0]**2-X[1]**2)
        return (math.fabs(f1+f2+f3) - self.eMin) / (self.eMax - self.eMin)

    def RawEnergy(self):
        X = self.XVar
        f1 = 0.5*X[0]**2 + X[1]**2 + math.sin(X[0]**2+X[1]**2)
        f2 = (3*X[0]-2*X[1]+4)**2/8 + (X[0]-X[1]+1)**2/27 + 15
        f3 = 1/(X[0]+X[1]+1) - 1.1*math.e**(-X[0]**2-X[1]**2)
        return math.fabs(f1+f2+f3)

    def __init__(self):
        self.Baseline(10000)
        self.XVar = self.XVarMax

```