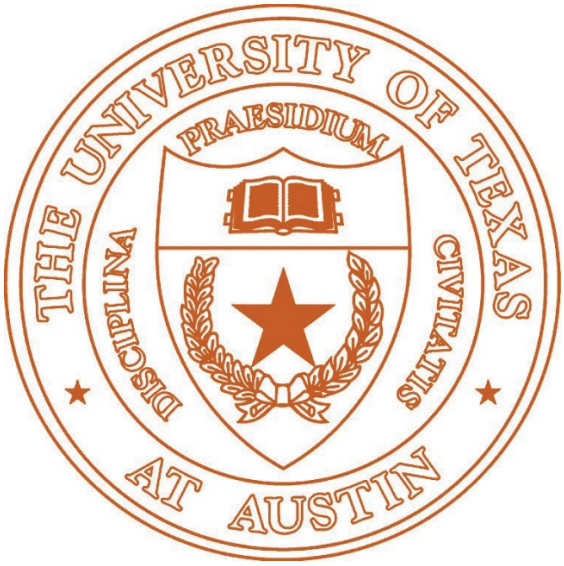
University of Texas at Austin, Cockrell School of Engineering
Data Mining – EE 380L



Problem Set # 1

February 08, 2016

Gabrielson Eapen
EID: EAPENGP

Discussed Homework with Following Students:
1. Mudra Gandhi
2. Rayo Landeros

Q1]

1. Download this dataset about professors in CS departments:
   http://users.ece.utexas.edu/~cmcaram/EE380L-2016/cs52.csv

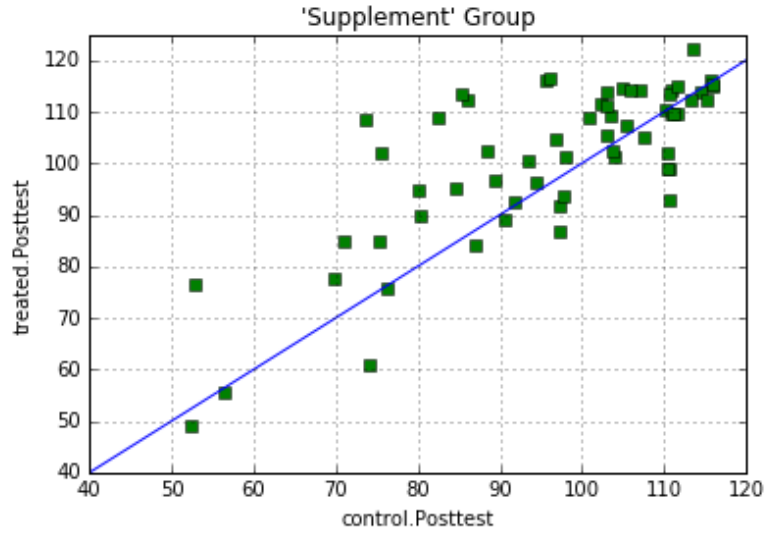
   Use Pandas to load this dataset in Python and count many professors teaching in UC Berkeley obtained their Ph.D. in UC Berkeley. Submit your python code through canvas. Name your file hw1.py and assume that cs52.csv will be in the same folder.

```
In [6]: df[df.University.str.contains("University of California - Berkeley") & df.Doctorate.str.contains
Out[6]: 13

In [7]: filtered_df = df[df.University.str.contains("University of California - Berkeley") & df.Doctorat

In [14]: print(filtered_df.Name.count())
         13
```

See iPython notebook file (**hw1-q1.ipynb**)

Q2 Part a]

(a) Plot `treated.Posttest` vs. `control.Posttest` for all students (so across all grades and cities) in the 'Supplement' group, and find the sample average of Treated.Posttest - Control.Posttest for these students.



sample average of 'Treated.Posttest - Control.Posttest' is: 4.8245901639.

```
In [1]: %matplotlib inline

        import pandas as pd
        import matplotlib
        import numpy as np
        import matplotlib.pyplot as plt

In [2]: df = pd.read_csv('electric.dat', sep='\s+')
        df.shape

Out[2]: (96, 7)
```

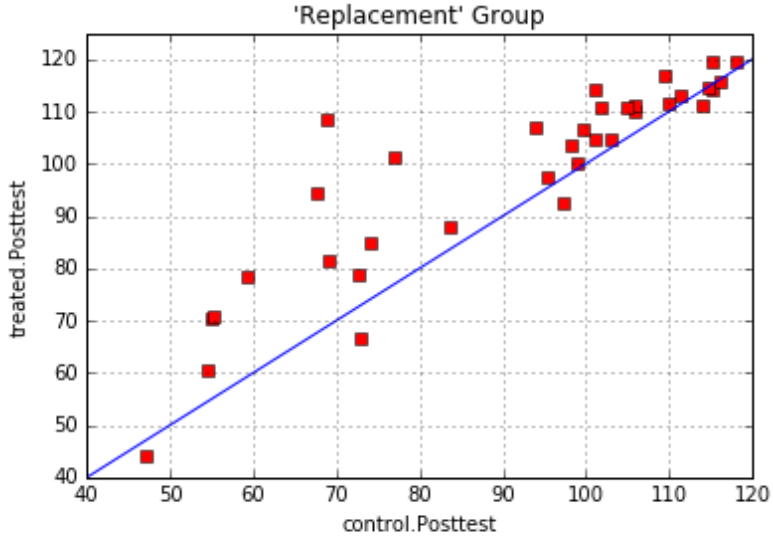
```
In [4]: sup_df = df[(df["Supplement?"] == 'S')]
        sup_df.shape

Out[4]: (61, 7)

In [5]: part_a, = plt.plot(sup_df["control.Posttest"],sup_df["treated.Posttest"],'gs')
        plt.grid(True)
        plt.ylabel("treated.Posttest")
        plt.xlabel("control.Posttest")
        plt.title("'Supplement' Group")
        plt.axis([40,120,40,125])

        # Add y=x
        plt.plot([40,130],[40,130], 'b-')
        plt.show()
        sample_avg = (sup_df["treated.Posttest"] - sup_df["control.Posttest"]).mean()
        # Print Sample Average
        print("sample average of 'Treated.Posttest - Control.Posttest' is: {}.".format(round(sample_avg,10)))
```

Code:

```
%matplotlib inline
import pandas as pd
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('electric.dat', sep='\s+')
sup_df = df[(df["Supplement?"] == 'S')]
sup_df.shape

part_a, = plt.plot(sup_df["control.Posttest"], sup_df["treated.Posttest"],'gs')
plt.grid(True)
plt.ylabel("treated.Posttest")
plt.xlabel("control.Posttest")
plt.title("'Supplement' Group")
plt.axis([40,120,40,125])

# Add y=x
plt.plot([40,130],[40,130], 'b-')
plt.show()
sample_avg = (sup_df["treated.Posttest"] - sup_df["control.Posttest"]).mean()
# Print Sample Average
print("sample average of 'Treated.Posttest - Control.Posttest' is:
{}.".format(round(sample_avg,10)))
```

Q2 Part b]

(b) Do the same for all the students in the Replacement group.




```
sample average of 'Treated.Posttest - Control.Posttest' is: 7.1085714286.
```

```
In [1]: %matplotlib inline

        import pandas as pd
        import matplotlib
        import numpy as np
        import matplotlib.pyplot as plt

In [2]: df = pd.read_csv('electric.dat', sep='\s+')
        df.shape

Out[2]: (96, 7)
```

```
In [6]: rep_df = df[(df["Supplement?"] == 'R')]
        rep_df.shape

Out[6]: (35, 7)

In [7]: part_b, = plt.plot(rep_df["control.Posttest"],rep_df["treated.Posttest"],'rs')
        plt.grid(True)
        plt.ylabel("treated.Posttest")
        plt.xlabel("control.Posttest")
        plt.title("'Replacement' Group")
        plt.axis([40,120,40,125])

        # Add y=x
        plt.plot([40,130],[40,130], 'b-')
        plt.show()
        sample_avg = (rep_df["treated.Posttest"] - rep_df["control.Posttest"]).mean()
        # Print Sample Average
        print("sample average of 'Treated.Posttest - Control.Posttest' is: {}.".format(round(sample_avg,10)))
```

Code:

```
%matplotlib inline
import pandas as pd
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('electric.dat', sep='\s+')
sup_df = df[(df["Supplement?"] == 'S')]
sup_df.shape

part_b, = plt.plot(rep_df["control.Posttest"],rep_df["treated.Posttest"], 'rs')
plt.grid(True)
plt.ylabel("treated.Posttest")
plt.xlabel("control.Posttest")
plt.title("'Replacement' Group")
plt.axis([40,120,40,125])

# Add y=x
plt.plot([40,130],[40,130], 'b-')
plt.show()
sample_avg = (rep_df["treated.Posttest"] - rep_df["control.Posttest"]).mean()
# Print Sample Average
print("sample average of 'Treated.Posttest - Control.Posttest' is:
{}.".format(round(sample_avg,10)))
```

Q2 Part c]

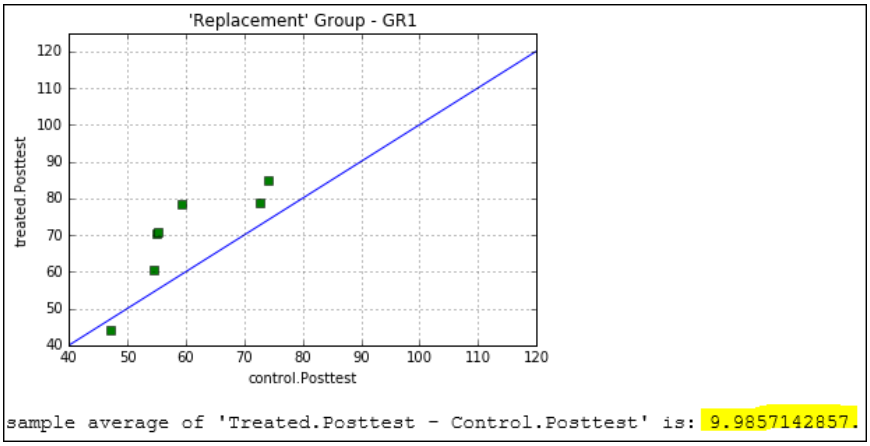
(c) Now to decide which was better: Compare the above plots. Do so by considering a line $y = x$. Are most of the points above or below that line? Using your observation of that, would you favor Supplementing or Replacing?

Visually, in part b only 4 data points (15%) appear below the y=x line and so that supports the replacement group having benefited the most. In part a, it looks like 18 data points (30%) appear below the y = x line. The second statistic is the sample average (of treated – control) is higher for the Replacement Group. So I would support Replacing.

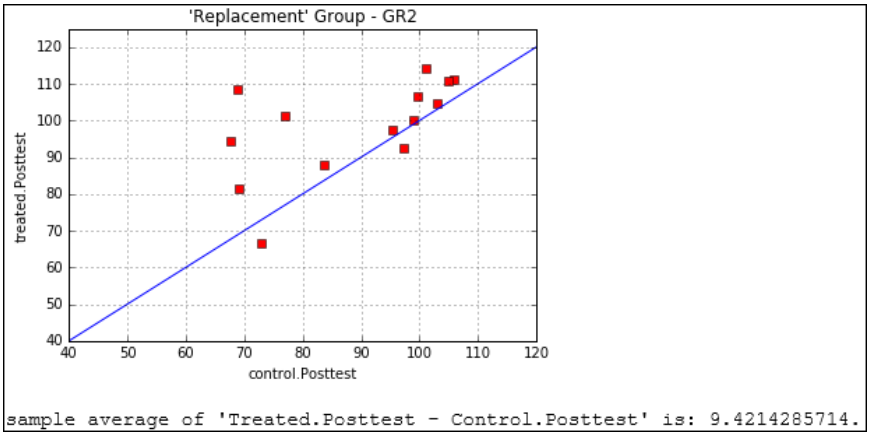
Q2 Part d]

(d) (Bonus) Are the results uniform across the student groups? Is there some group that benefited more or less?

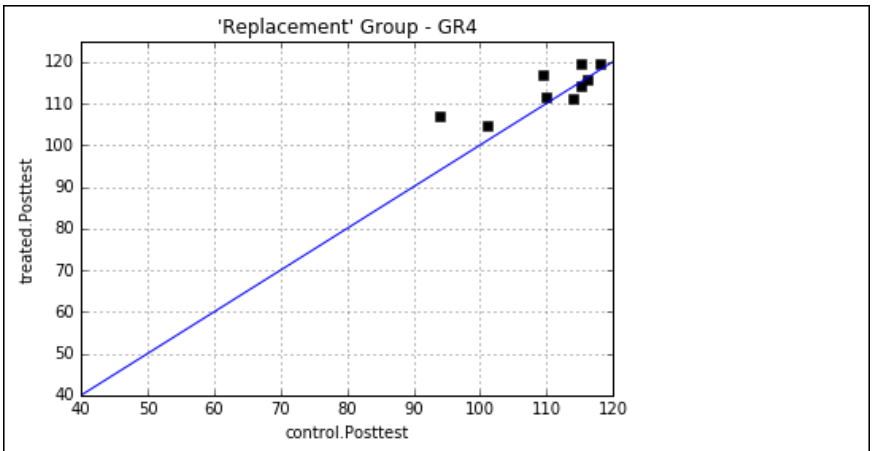
From the charts below, it looks like the Grade 1 class benefited the most. It had the most positive posttest sample average of treatment – control.



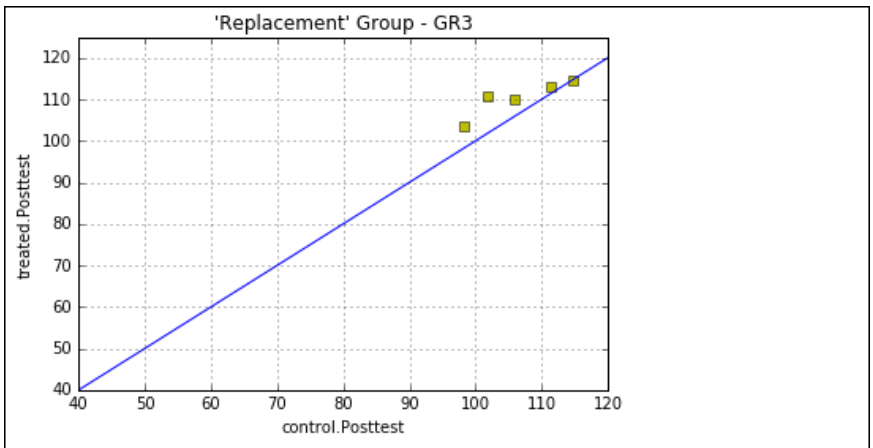
sample average of 'Treated.Posttest - Control.Posttest' is: 9.9857142857.



sample average of 'Treated.Posttest - Control.Posttest' is: 9.4214285714.




```
sample average of 'Treated.Posttest - Control.Posttest' is: 3.0.
```




```
sample average of 'Treated.Posttest - Control.Posttest' is: 4.0.
```

Q3 Part a]

(a) Plot (as a scatterplot) mortality vs the level of nitric oxides. Would you expect linear regression to provide a good fit of these data?

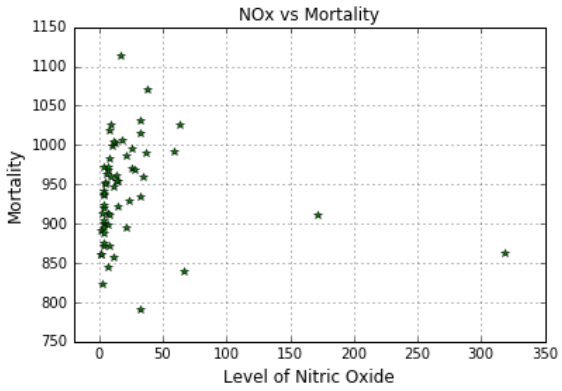
In [1]:
```
# http://www.kenbenoit.net/courses/ME104/logmodels2.pdf
# Name: Gabe Eapen
# UT EID: eapengp
# PS1 - Q3
```

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from scipy import linalg
from sklearn.preprocessing import FunctionTransformer
```

In [3]:
```
df=pd.read_stata("pollution.dta")
print(df.columns)
```

```
Index([u'prec', u'jant', u'jult', u'ovr65', u'popn', u'educ', u'hous', u'dens',
       u'nonw', u'wwdrk', u'poor', u'hc', u'nox', u'so2', u'humid', u'mort'],
      dtype='object')
```

In [4]:
```
#Part a
part_a, = plt.plot(df["nox"],df["mort"],'g*')
plt.grid(True)
plt.ylabel("Mortality",fontsize=12)
plt.xlabel("Level of Nitric Oxide",fontsize=12)
plt.title("NOx vs Mortality")
plt.xlim(-20, 350)
plt.show()
```



Q3 Part b]

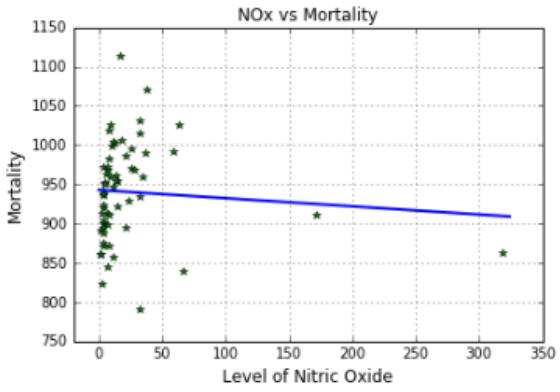
(b) Check your answer: Fit a line of the form

$$y = \beta_1 x + \beta_0$$

In [5]:
```
#Part b
# http://docs.scipy.org/doc/numpy-1.10.0/reference/generated/numpy.linalg.lstsq.html
# B1 = m; B0 = c
# y = B1x + B0
# y = Ap, where A = [[x 1]] and p = [[B1], [B0]]
x = df["nox"]
A = np.vstack([x, np.ones(len(x))]).T
#print A
y = np.array(df["mort"])
#m, c = np.linalg.lstsq(A, y)[0]
B1, B0 = np.linalg.lstsq(A, y)[0]
print("B1 is: {}.".format(round(B1,8)))
print("B0 is: {}.".format(round(B0,8)))
```

```
B1 is: -0.10388706.
B0 is: 942.71147529.
```

In [6]:
```
#Part b
# Add bestfit line to ScatterPlot using B1 and B0 values
# y = B1x + B0

part_b, = plt.plot(df["nox"],df["mort"],'g*')
plt.grid(True)
plt.ylabel("Mortality",fontsize=12)
plt.xlabel("Level of Nitric Oxide",fontsize=12)
plt.title("NOx vs Mortality")
plt.xlim(-20, 350)
xVals = np.arange(0, 325)
yVals = (B1*xVals) + B0
plt.plot(xVals,yVals,'b-',linewidth=2.0)
plt.show()
```



Q3 Part c]

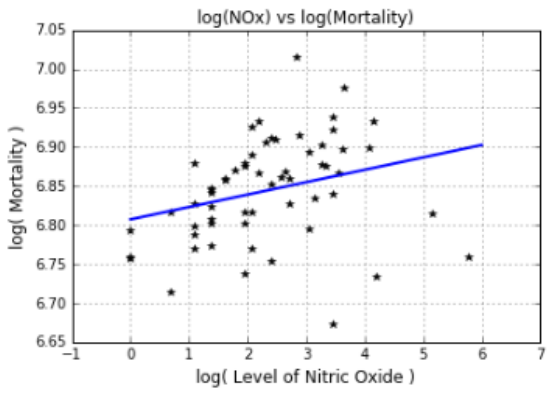
(c) Now, find a transformation of the data that gives something that is a bit closer to a linear relationship (and hence one where linear regression is a good idea). Then repeat your plotting and regression of the above.

Log-Log fitting:

In [7]:
```
#part c
# Try log-log
logX = np.log(df["nox"])
logY = np.log(df["mort"])

A = np.vstack([logX, np.ones(len(logX))]).T
#print A
y = np.array(logY)
#m, c = np.linalg.lstsq(A, y)[0]
loglogB1, loglogB0 = np.linalg.lstsq(A, y)[0]
print("B1 is: {}.".format(round(loglogB1,8)))
print("B0 is: {}.".format(round(loglogB0,8)))
```

```
B1 is: 0.01589323.
B0 is: 6.80717471.
```

In [8]:
```
#part c
# Plot log-log

part_c1, = plt.plot(logX,logY,'k*')
plt.grid(True)
plt.ylabel("log( Mortality )",fontsize=12)
plt.xlabel("log( Level of Nitric Oxide )",fontsize=12)
plt.title("log(NOx) vs log(Mortality)")
plt.xlim(-1, 7)
xVals = np.arange(0, 7)
yVals = (loglogB1*xVals) + loglogB0
plt.plot(xVals,yVals,'b-',linewidth=2.0)
plt.show()
```

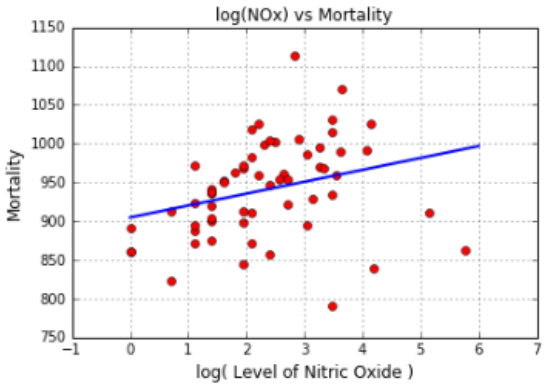


Log-Linear fitting:

In [9]:
```
#part c
# Try log-linear
logX = np.log(df["nox"])
linY = df["mort"]

A = np.vstack([logX, np.ones(len(logX))]).T
#print A
y = np.array(linY)
#m, c = np.linalg.lstsq(A, y)[0]
loglinB1, loglinB0 = np.linalg.lstsq(A, y)[0]
print("B1 is: {}.".format(round(loglinB1,8)))
print("B0 is: {}.".format(round(loglinB0,8)))
```

```
B1 is: 15.33549673.
B0 is: 904.72446381.
```

In [10]:
```
#part c
# Plot log-lin

part_c1, = plt.plot(logX,linY,'ro')
plt.grid(True)
plt.ylabel("Mortality",fontsize=12)
plt.xlabel("log( Level of Nitric Oxide )",fontsize=12)
plt.title("log(NOx) vs Mortality")
plt.xlim(-1, 7)
xVals = np.arange(0, 7)
yVals = (loglinB1*xVals) + loglinB0
plt.plot(xVals,yVals,'b-',linewidth=2.0)
plt.show()
```



Both the log-log fitting and the log-linear fitting exhibit a better linear relationship. The edge is for the log-linear fitting based on visual inspection of the plot.

Q4 Part a]

(a) Use `numpy.polyfit` to do this. If your points are generated at random, what do you observe about the degree of the polynomial you need, and the number of points you have?

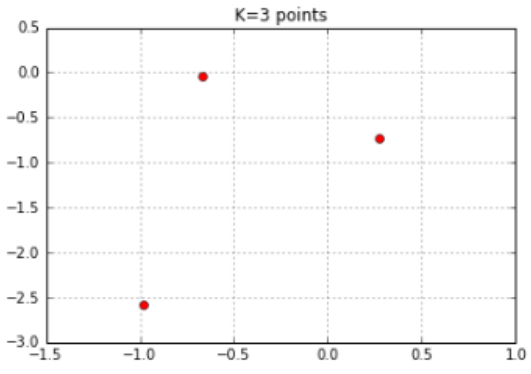
In [ ]:
```
# Name: Gabe Eapen
# UT EID: eapengp
# PS1 - Q3
```

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

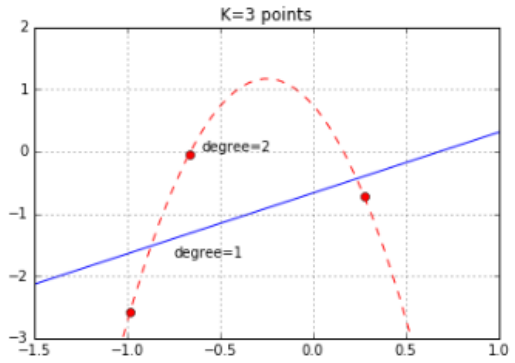
In [3]:
```
#Part a.  k=3 points
k = 3
X = np.random.randn(k)
Y = np.random.randn(k)
print X, Y
# try 2 and 1 dimensional polynomial
p2 = np.poly1d(np.polyfit(X, Y, 2))
p1 = np.poly1d(np.polyfit(X, Y, 1))
```

```
[ 0.27575856 -0.98476843 -0.66619144] [-0.72170136 -2.58437845 -0.0423013 ]
```

In [7]:
```
#plot generated Points
part_a, = plt.plot(X,Y,'ro')
plt.grid(True)
plt.title("K=3 points")
plt.xlim(-1.5, 1)
plt.ylim(-3, 0.5)
plt.show()
```



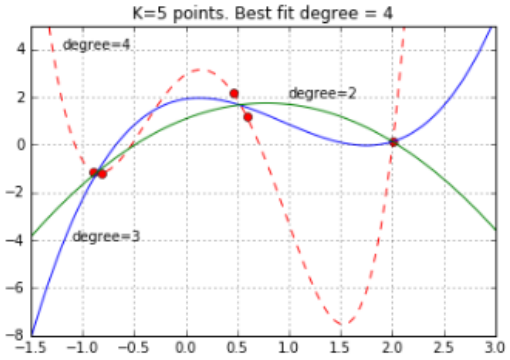
In [10]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-2, 2, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p2(xSamples), 'r--')
plt.plot(xSamples, p1(xSamples), 'b-')
plt.grid(True)
plt.title("K=3 points")
plt.xlim(-1.5, 1)
plt.ylim(-3, 2)
plt.text(-0.6,0,'degree=2')
plt.text(-0.75,-1.7,'degree=1')
plt.show()
```



In [11]:
```
#Part a.  k=5 points
k = 5
X = np.random.randn(k)
Y = np.random.randn(k)
print X, Y
# try 2 and 1 dimensional polynomial
p4 = np.poly1d(np.polyfit(X, Y, 4))
p3 = np.poly1d(np.polyfit(X, Y, 3))
p2 = np.poly1d(np.polyfit(X, Y, 2))
```

```
[-0.80807615 -0.89509325  2.00776738  0.59285527  0.45668272] [-1.17706226 -1.12493975  0.144
09225  1.17415268  2.19094668]
```

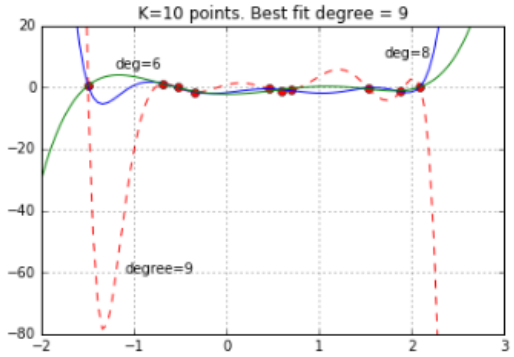
In [18]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-2, 3, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p4(xSamples), 'r--')
plt.plot(xSamples, p3(xSamples), 'b-')
plt.plot(xSamples, p2(xSamples), 'g-')
plt.grid(True)
plt.title("K=5 points. Best fit degree = 4")
plt.xlim(-1.5, 3)
plt.ylim(-8, 5)
plt.text(-1.2,4,'degree=4')
plt.text(-1.1,-4,'degree=3')
plt.text(1,2,'degree=2')
plt.show()
```



In [33]:
```
#Part a.  k=10 points
k = 10
X = np.random.randn(k)
Y = np.random.randn(k)
print X, Y
# try 2 and 1 dimensional polynomial
p9 = np.poly1d(np.polyfit(X, Y, 9))
p8 = np.poly1d(np.polyfit(X, Y, 8))
p6 = np.poly1d(np.polyfit(X, Y, 6))
```

```
[ 2.0938385   0.45364752  0.69683403 -1.49999303  1.87862826 -0.35298797
 -0.53100079  1.52833299 -0.68420982  0.59805772] [ 0.18575436 -0.43256184 -0.91854176  0.541
32357 -1.42203881 -1.55253556
 -0.10567179 -0.25694526  1.13261669 -1.14588981]
```

```
In [46]: #plot generated Points and best fit line
         xSamples = np.linspace(-2, 3, 100)
         plt.plot(X, Y, 'ro')
         plt.plot(xSamples, p9(xSamples), 'r--')
         plt.plot(xSamples, p8(xSamples), 'b-')
         plt.plot(xSamples, p6(xSamples), 'g-')
         plt.grid(True)
         plt.title("K=10 points. Best fit degree = 9")
         plt.xlim(-2, 3)
         plt.ylim(-80, 20)
         plt.text(-1.1,-60,'degree=9')
         plt.text(1.7,10,'deg=8')
         plt.text(-1.2,6,'deg=6')
         plt.show()
```



Key observation:
Best fit line (going through all points) is created when degree of polynomial is **k – 1**

Q4 Part b]

(b) Show that if you have k points and a polynomial of degree k, finding the exact fit, i.e., the degree-k polynomial that goes through the k points exactly, amounts to solving a system of linear equations: $Ax = b$.

From part a we have experimentally seen that if we have **k** points, you need a polynomial of degree k-1.

The general form of the equation of the polynomial of degree k-1 is:

$$f(x) = a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \cdots + a_2x^2 + a_1x^1 + a_0x^0$$
$$f(x) = a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \cdots + a_2x^2 + a_1x + a_0$$

Where you have k unknowns (0 … k-1 = k unknowns)
So if you have k points, we can create k linear equations by substituting each point (x_j, y_j) where $j = 0 \ldots k-1$ as follows:

$$y_j = a_jx_j^j + a_{j-1}x_j^{j-1} + \cdots + a_2x_j^2 + a_1x_j + a_0$$
$$y_{j-1} = a_jx_{j-1}^j + a_{j-1}x_{j-1}^{j-1} + \cdots + a_2x_{j-1}^2 + a_1x_{j-1} + a_0$$
.
.
$$y_1 = a_jx_1^j + a_{j-1}x_1^{j-1} + \cdots + a_2x_1^2 + a_1x_1 + a_0$$
$$y_0 = a_jx_0^j + a_{j-1}x_0^{j-1} + \cdots + a_2x_0^2 + a_1x_{j0} + a_0$$

So we have **k** (j = 0 .. k-1) linear equations to solve with **k** unknowns

Q4 Part c]

(c) (Bonus) As we did in class, experiment with overfitting vs amount of data. Generate points according to a line, but then add a little noise. Then try fitting a line (using least squares). Generate more and more points according to that same relationship (always adding noise) and notice that the line that you find, very quickly stabilizes and does not change. Now try fitting a much higher degree, say, 10 or 15, polynomial. Note that it takes many more points before the produced solution stops changing. In particular, note how much the solution changes between the 10 and 11 (or so...) points that you generate and fit.

In [1]:
```
# Name: Gabe Eapen
# UT EID: eapengp
# PS1 - Q4c
```

In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from scipy import linalg
# Y = 3x
def f_noise(x):
    """f(x)+noise"""
    return x + (0.2 * np.random.randn())

def f(x,m,c):
    """f(x) = mx + c"""
    return m*x + c
```

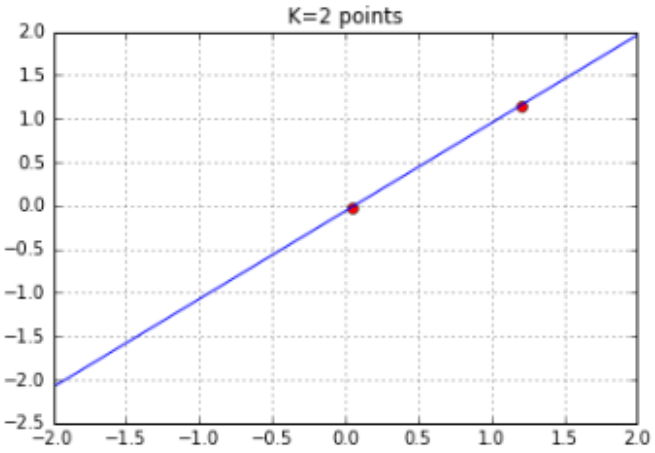
In [3]:
```
#k=2 points
k = 2
X = np.random.randn(k)
Y = np.array([f_noise(x) for x in X])
```

In [4]:
```
# http://docs.scipy.org/doc/numpy-1.10.0/reference/generated/numpy.linalg.ls
#X = np.array([0, 1, 2, 3])
#Y = np.array([0, 1, 2, 3])
A = np.vstack([X, np.ones(len(X))]).T
m, c = np.linalg.lstsq(A, Y)[0]
print m, c
```

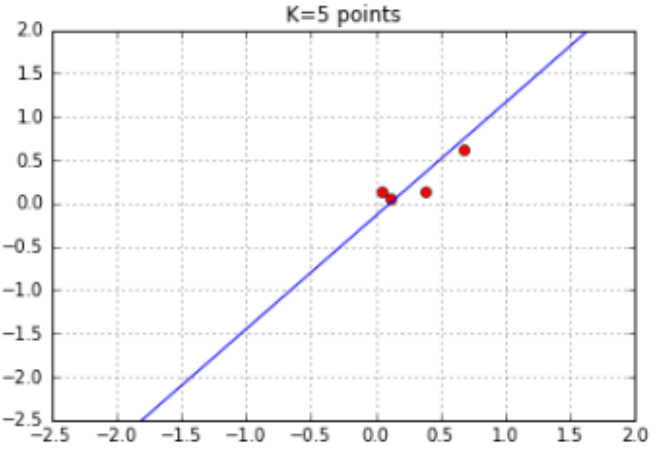
```
1.01117019406 -0.0635816577271
```

In [5]:
```
xSamples = np.linspace(-2, 2, 100)
#print type(xSamples)
Y_Lsqr = np.array([f(x,m,c) for x in xSamples])
#print type(Y_Lsqr)
```

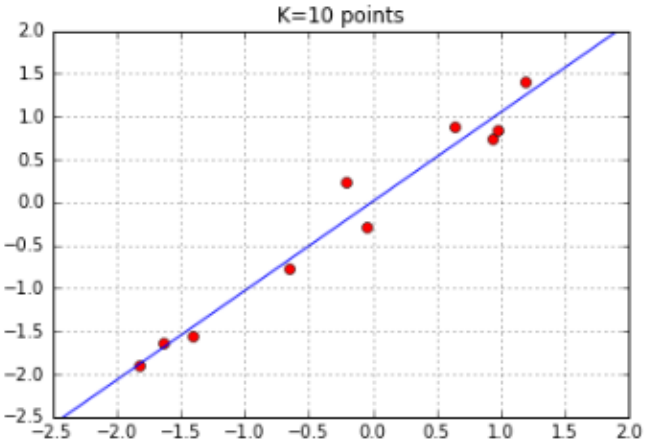
In [6]:
```
#plot generated Points and best fit line using Least Squares K=2
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=2 points")
plt.show()
```



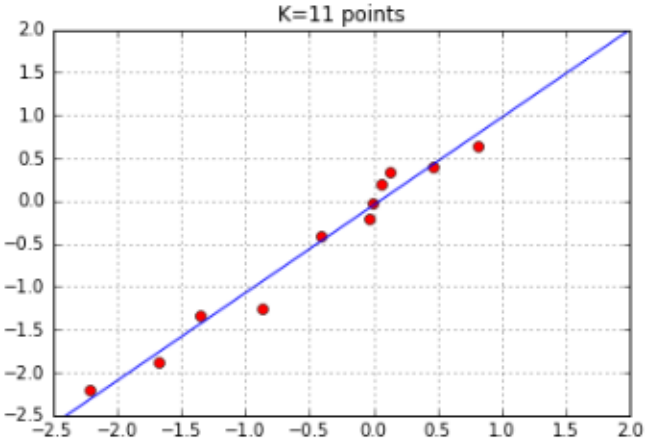
In [9]:
```
#plot generated Points and best fit line using Least Squares K=5
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=5 points")
plt.xlim(-2.5, 2)
plt.ylim(-2.5, 2)
plt.show()
```



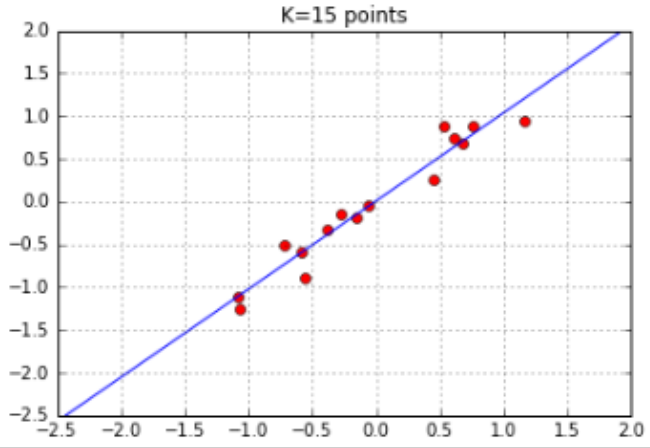
In [12]:
```
#plot generated Points and best fit line using Least Squares K=10
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=10 points")
plt.xlim(-2.5, 2)
plt.ylim(-2.5, 2)
plt.show()
```



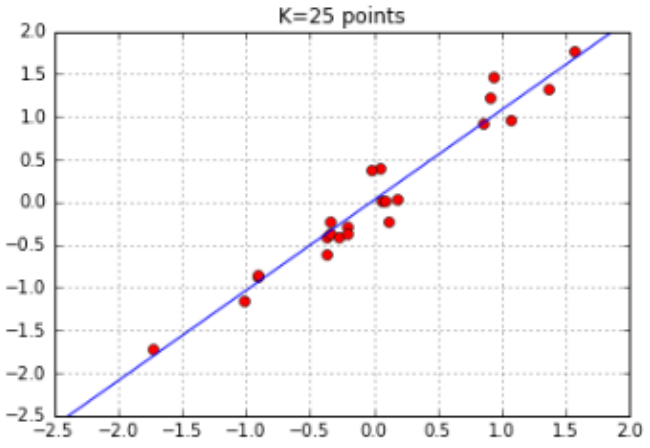
In [15]:
```
#plot generated Points and best fit line using Least Squares K=11
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=11 points")
plt.xlim(-2.5, 2)
plt.ylim(-2.5, 2)
plt.show()
```



In [18]:
```
#plot generated Points and best fit line using Least Squares K=15
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=15 points")
plt.xlim(-2.5, 2)
plt.ylim(-2.5, 2)
plt.show()
```



In [21]:
```
#plot generated Points and best fit line using Least Squares K=25
plt.plot(X, Y, 'ro')
plt.plot(xSamples, Y_Lsqr, 'b-')
plt.grid(True)
plt.title("K=25 points")
plt.xlim(-2.5, 2)
plt.ylim(-2.5, 2)
plt.show()
```



It looks like the line fit is pretty stable whether two or 25 points. We can get a line fit with fewer points.

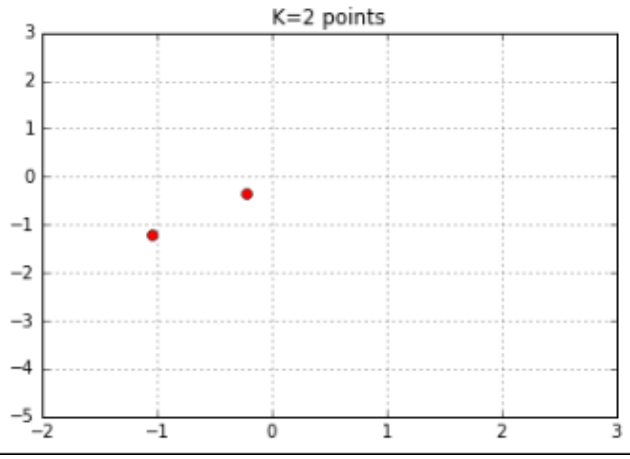
Next let us look at Polynomial fit on degree 10 and 11 and similar number of points

In [22]:
```
# Now fit a polynomial
# k = 2
k = 2
X = np.random.randn(k)
Y = np.array([f_noise(x) for x in X])
#print X, Y
# try 2 and 1 dimensional polynomial
p10 = np.poly1d(np.polyfit(X, Y, 10))
p11 = np.poly1d(np.polyfit(X, Y, 11))
```

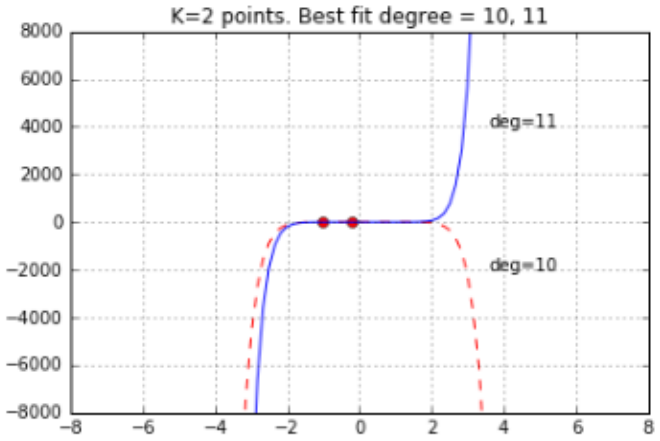
```
C:\Anaconda2\lib\site-packages\numpy\lib\polynomial.py:594: RankWarning:
Polyfit may be poorly conditioned
  warnings.warn(msg, RankWarning)
C:\Anaconda2\lib\site-packages\numpy\lib\polynomial.py:594: RankWarning:
Polyfit may be poorly conditioned
  warnings.warn(msg, RankWarning)
```

* Warning due to insufficient points

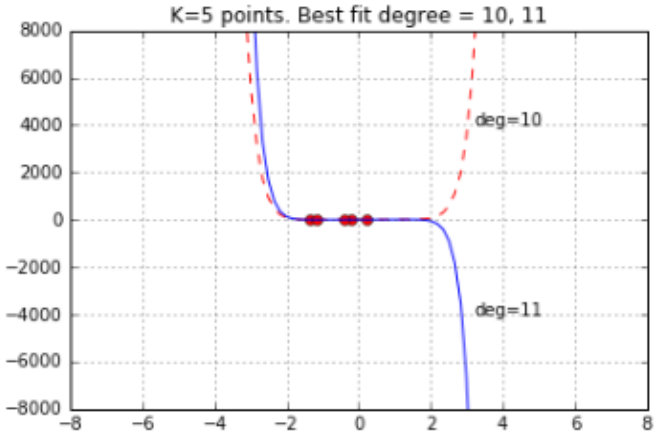
In [23]:
```
#plot generated Points (k=2)
part_a, = plt.plot(X,Y,'ro')
plt.grid(True)
plt.title("K=2 points")
plt.xlim(-2, 3)
plt.ylim(-5, 3)
plt.show()
```



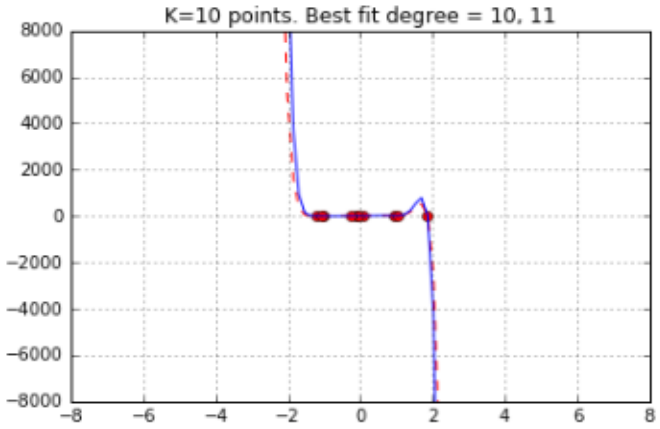
In [25]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=2 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
plt.text(3.6,-2000,'deg=10')
plt.text(3.6,4000,'deg=11')
plt.show()
```



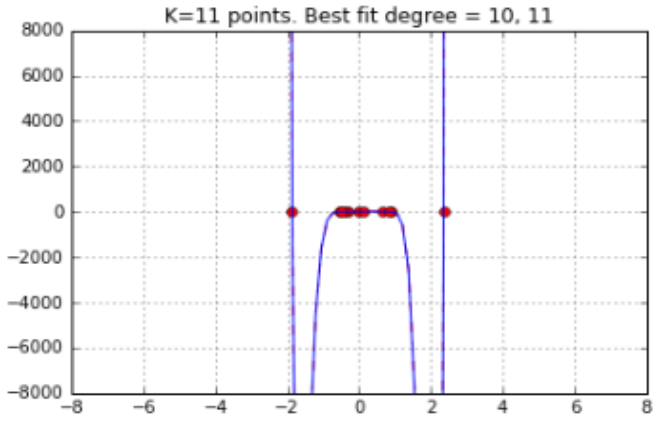
In [30]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=5 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
plt.text(3.2,4000,'deg=10')
plt.text(3.2,-4000,'deg=11')
plt.show()
```



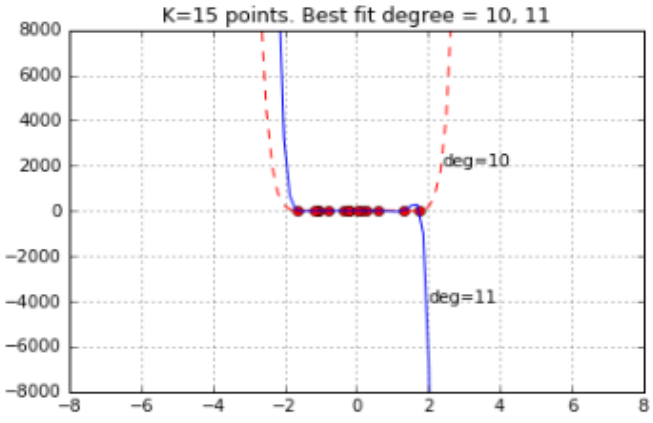
In [34]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=10 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
#plt.text(-5,-3000,'deg=10')
#plt.text(-5,3000,'deg=11')
plt.show()
```



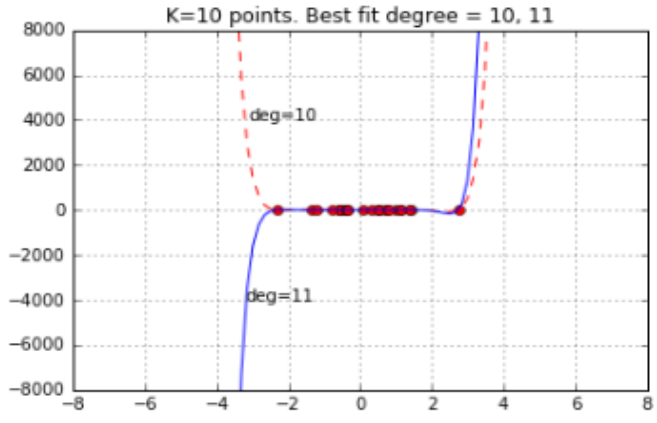
In [38]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=11 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
#plt.text(-5,-3000,'deg=10')
#plt.text(-5,3000,'deg=11')
plt.show()
```



In [42]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=15 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
plt.text(2.4,2000,'deg=10')
plt.text(2,-4000,'deg=11')
plt.show()
```



In [46]:
```
#plot generated Points and best fit line
xSamples = np.linspace(-8, 8, 100)
plt.plot(X, Y, 'ro')
plt.plot(xSamples, p10(xSamples), 'r--')
plt.plot(xSamples, p11(xSamples), 'b-')
plt.grid(True)
plt.title("K=10 points. Best fit degree = 10, 11")
plt.xlim(-8, 8)
plt.ylim(-8000, 8000)
plt.text(-3.1,4000,'deg=10')
plt.text(-3.2,-4000,'deg=11')
plt.show()
```



Q5 Part a]

(a) Plot the *unit ball* with respect to the p-norm for $p = 1, 2, 10, \infty$, in two dimensions, $n = 2$:

$$B_{\ell p} = \{\boldsymbol{x} = (x_1, x_2) \,:\, \|\boldsymbol{x}\|_p \leq 1\}.$$

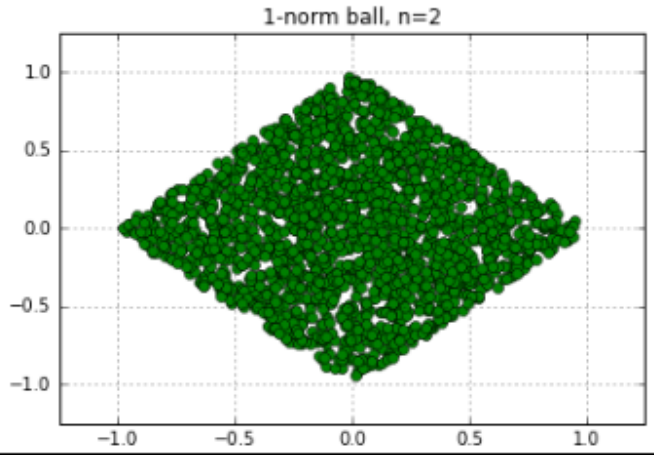
One way to do this is by generating random points in the square $[-1, 1]^n$ and plotting those that have norm less than 1. You can use the `numpy` command `linalg.norm`.

In [1]:
```
# Name: Gabe Eapen
# UT EID: eapengp
# PS1 - Q5
```

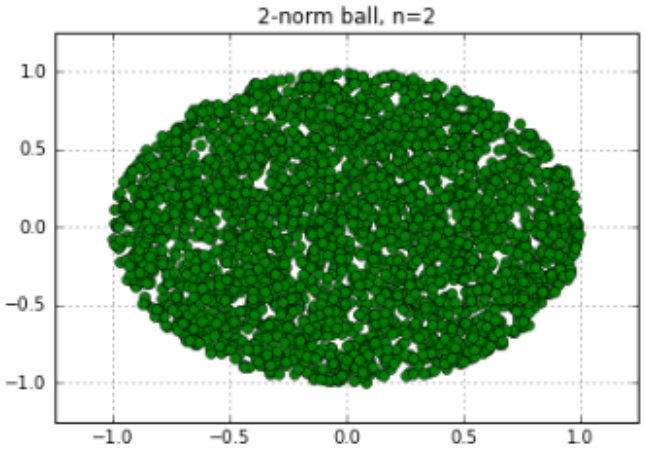
In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:
```
#Part a
def plotUnitBall(p,title):
    for i in range(4000):
        #rand returs in range 0 - 1.  We need in range -1 to 1
        xAxis = np.array([np.random.rand()*2-1,np.random.rand()*2-1])
        if np.linalg.norm(xAxis,p) < 1:
            plt.plot(xAxis[0],xAxis[1],'go')
    plt.grid(True)
    plt.title(title)
    plt.xlim(-1.25, 1.25)
    plt.ylim(-1.25, 1.25)
    #plt.axis("equal")
    plt.show()
```

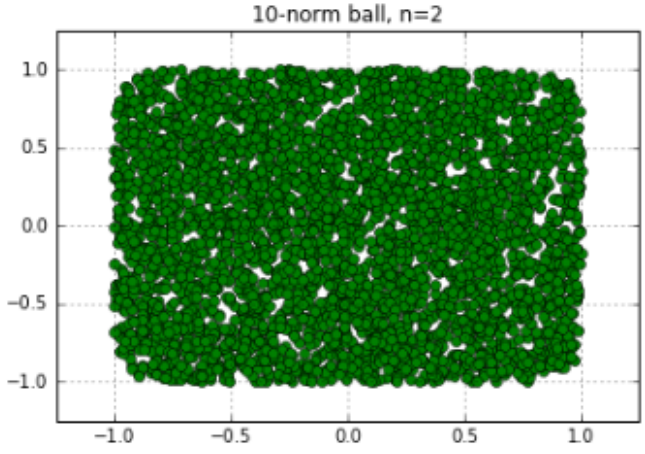
In [4]:
```
# p=1 Norm, n = 2
plotUnitBall(1,"1-norm ball, n=2")
```



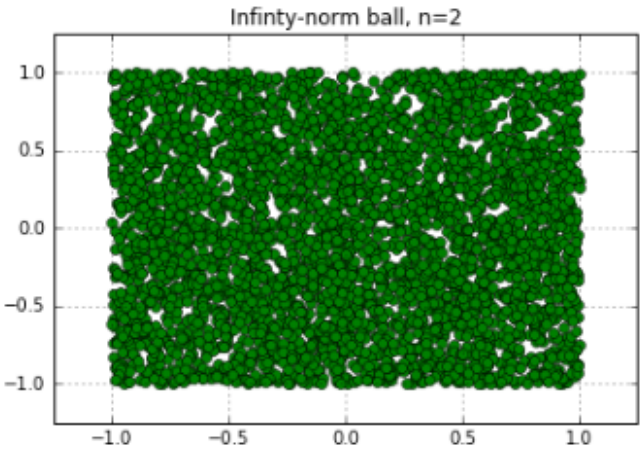
In [5]:
```
# p=2 Norm, n = 2
plotUnitBall(2,"2-norm ball, n=2")
```



In [6]:
```
# p=10 Norm, n = 2
plotUnitBall(10,"10-norm ball, n=2")
```



In [7]:
```
# p=inf Norm, n = 2
plotUnitBall(np.inf,"Infinty-norm ball, n=2")
```

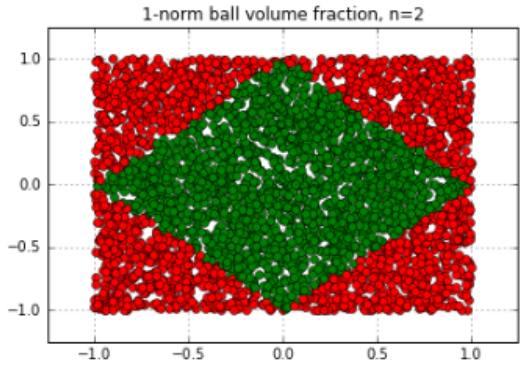


Q5 Part b]

(b) We can use the *Monte Carlo* approach to estimating the probability of events, or, equivalently, the volume of a set. The idea is: generate randomly distributed points from a set whose volume you know, that contains the set whose volume you'd like to compute. Then, count the fraction of the random points that land inside your set. Use this approach to approximate the volume of the unit ball with respect to the ℓ^p norm, for $p = 1, 2, \infty$.

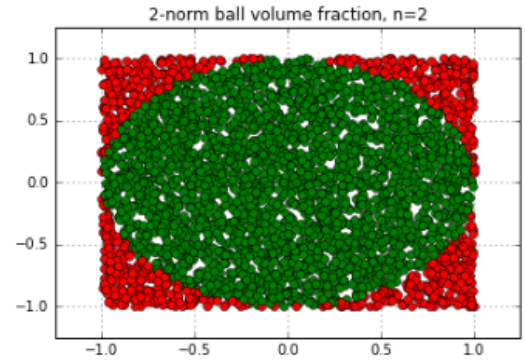
In [8]:
```
# part b - Plot/compute using Monte Carlo method
def plotUsingMonteCarlo(p,title):
    num_Pts = 4000
    inPts = []
    for i in range(num_Pts):
        #rand returs in range 0 - 1.  We need in range -1 to 1
        xAxis = np.array([np.random.rand()*2-1,np.random.rand()*2-1])
        if np.linalg.norm(xAxis,p) < 1:
            inPts.append([xAxis[0],xAxis[1]])
            plt.plot(xAxis[0],xAxis[1],'go')
        else:
            plt.plot(xAxis[0],xAxis[1],'ro')
    plt.grid(True)
    plt.title(title)
    plt.xlim(-1.25, 1.25)
    plt.ylim(-1.25, 1.25)
    #plt.axis("equal")
    plt.show()
    print "Number of points inside: ",len(inPts)
    print "Ratio of InPts to OutPts (Vol Approx): ",len(inPts)/float(num_Pts)
```

In [9]:
```
# p=1 Norm, n = 2
plotUsingMonteCarlo(1,"1-norm ball volume fraction, n=2")
```



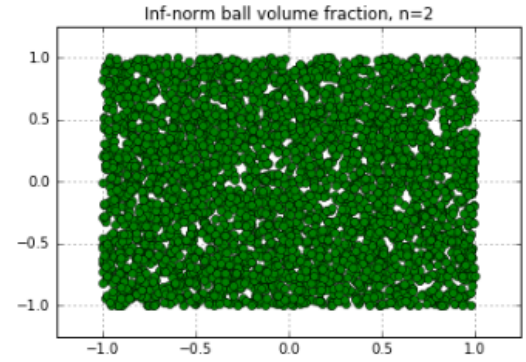

```
Number of points inside:  2009
Ratio of InPts to OutPts (Vol Approx):  0.50225
```

In [10]:
```
# p=2 Norm, n = 2
plotUsingMonteCarlo(2,"2-norm ball volume fraction, n=2")
```




```
Number of points inside:  3101
Ratio of InPts to OutPts (Vol Approx):  0.77525
```

In [11]:
```
# p=inf Norm, n = 2
plotUsingMonteCarlo(np.inf,"Inf-norm ball volume fraction, n=2")
```




```
Number of points inside:  4000
Ratio of InPts to OutPts (Vol Approx):  1.0
```

Q5 Part c]

(c) (Bonus) Now, try to use this Monte Carlo approach to estimate the probability of the ℓ^p ball, for $p = 2$, in higher dimensions. Try 5, 10 and then 100 dimensions. Do you succeed? Note that you can just look up the answer to check if your Monte Carlo result is correct, or even close to correct. Comment on the success and/or ease of your method.

The Monte Carlo plotting and Volume approximation method I used in part B only works in 2 dimensions. I do not know how to plot multi-dimensions (5, 10, 100). I suspect if we calculated the volume approximation without plotting, we may be able to get some approximation to the volume of the unit ball as we did in 2 dimensions but cannon comment on how accurate it might be.

Q6 Part a]

(a) To warm up, write a gradient descent routine to solve the 1-D least squares regression problem that you wrote down in class on Sunday. Recall that the gradient descent algorithm initializes with some x_0, and then updates via

$$x^{(k+1)} = x^{(k)} - \eta f'(x^{(k)}).$$

Choose $\eta < f''(x)$ and $\eta > f''(x)$, and comment on what you observe, ideally with a graphical explanation of why there is such a big difference between the two cases.

Loss Function from class for

$f(\beta) = \frac{5}{2}\beta^2 + 4.975\beta + 2.48$	$f'(\beta) = 5\beta + 4.975$	$f''(\beta) = 5$

Solving for minimum using $f'(\beta) = 0$ (closed form)

$5\beta + 4.975 = 0$

$5\beta = -4.975$

$\beta = -0.995$

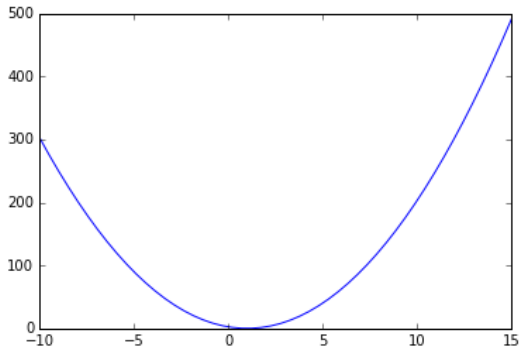
Now using Python and Iterative Gradient Descent

```
In [1]: # 6a
        #To warm up, write a gradient descent routine to solve the 1-D least squares regression
        #problem that you wrote down in class on Sunday. Recall that the gradient descent
        #algorithm initializes with some x0, and then updates via
        #x(k+1) = x(k) □ •f0(x(k)):
```

```
In [2]: import numpy as np
        import matplotlib.pyplot as plt
        %matplotlib inline

        def fB(x):
            return 2.5*B**2 - 4.975*B + 2.48

        def fB_derivative(x):
            return 5*x + 4.975
```

```
In [3]: B = np.linspace(-10,15,num=50)
        y = fB(B)
        fig, ax = plt.subplots()
        ax.plot(B, y)
```

```
Out[3]: [<matplotlib.lines.Line2D at 0x9f1c390>]
```



```
In [4]: # Using Gradient Descent
x_k = x_k_plus_1 = 0
theta = 0.1
iterations = 20
min_vals = []

for x in xrange(iterations):
    x_k = x_k_plus_1
    # x(k+1) = x(k) - theta*f'(x(k))
    x_k_plus_1 = x_k - (theta * fB_derivative(x_k))
    min_vals.append(x_k)

print("Minimum from Gradient Descent occurs at {}.".format(round(x_k_plus_1,3)))
print min_vals
```

```
Minimum from Gradient Descent occurs at -0.995.
[0, -0.4975, -0.74625, -0.870625, -0.9328124999999999, -0.9639062499999999, -0.97945312499999
99, -0.9872265625, -0.99111328125, -0.993056640625, -0.9940283203125, -0.9945141601562499,
-0.9947570800781249, -0.9948785400390624, -0.9949392700195311, -0.9949696350097655, -0.994984
8175048828, -0.9949924087524413, -0.9949962043762206, -0.9949981021881102]
```

Q6 Part b]

Here, $\nabla f(\boldsymbol{x})$ denotes the *gradient* of the function f at the point $\boldsymbol{x}$. If f is a function that takes n arguments, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, then its gradient is a $n \times 1$ vector:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}.$$

Compute the gradient for the following functions:

- $$f(\boldsymbol{x}) = 2x_1 - x_2 + 5x_3.$$
- $$f(\boldsymbol{x}) = x_1^2 x_2 + x_2 \sin x_3 + 2x_3.$$

$f(x) = 2x_1 - x_2 + 5x_3$

$\frac{\partial f}{\partial x_1}(x) = 2$ $\quad$ $\frac{\partial f}{\partial x_2}(x) = -1$ $\quad$ $\frac{\partial f}{\partial x_3}(x) = 5$

$$\nabla f(x) = \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}$$

$f(x) = {x_1}^2 x_2 + x_2 \sin x_3 + 2x_3$

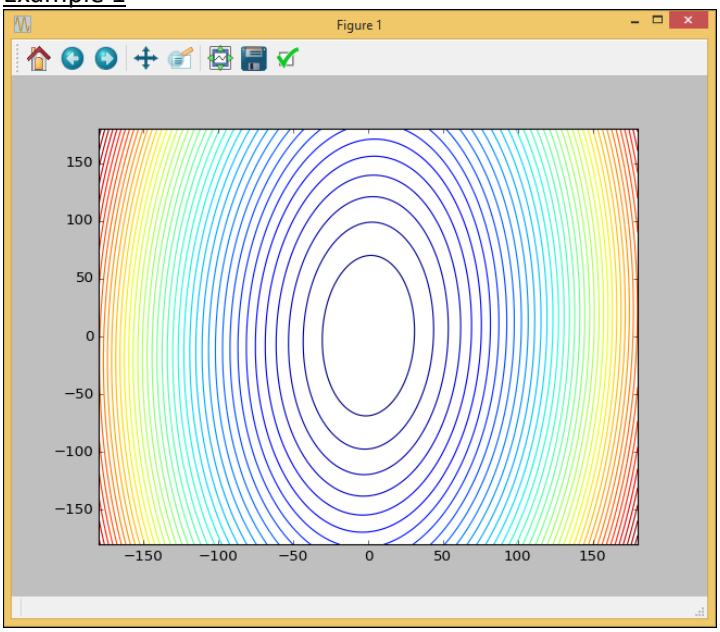
$\frac{\partial f}{\partial x_1}(x) = 2\, x_1 x_2$ $\quad$ $\frac{\partial f}{\partial x_2}(x) = {x_1}^2 + \sin x_3$ $\quad$ $\frac{\partial f}{\partial x_3}(x) = x_2 \cos x_3 + 2$

$$\nabla f(x) = \begin{pmatrix} 2\, x_1 x_2 \\ {x_1}^2 + \sin x_3 \\ x_2 \cos x_3 + 2 \end{pmatrix}$$

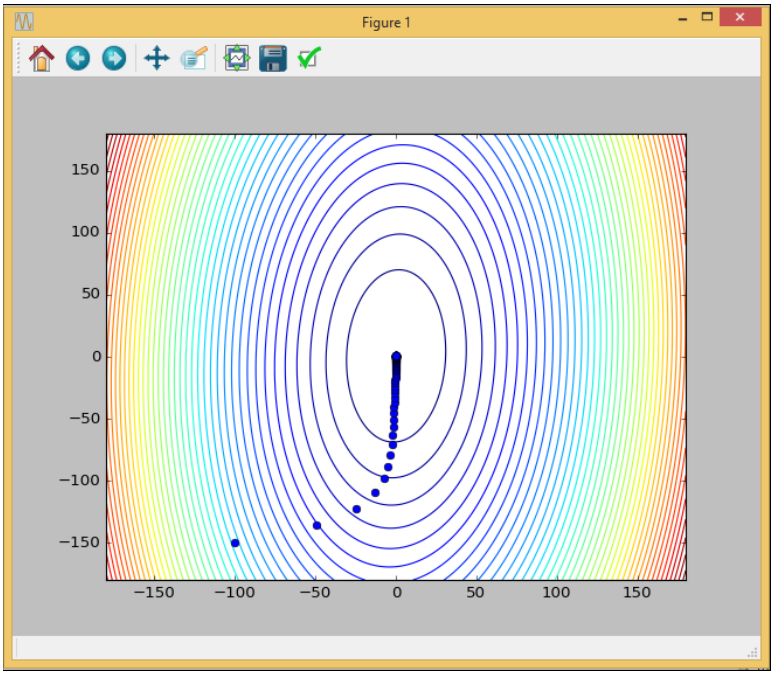
Q6 Part c]

(c) Have a look at the IPython notebook `2D Quadratic Gradient Descent`. Generate three (3) random examples, and while keeping the example fixed, try to find values of η, the step size, for which gradient descent converges, and for which it diverges.

Example 1

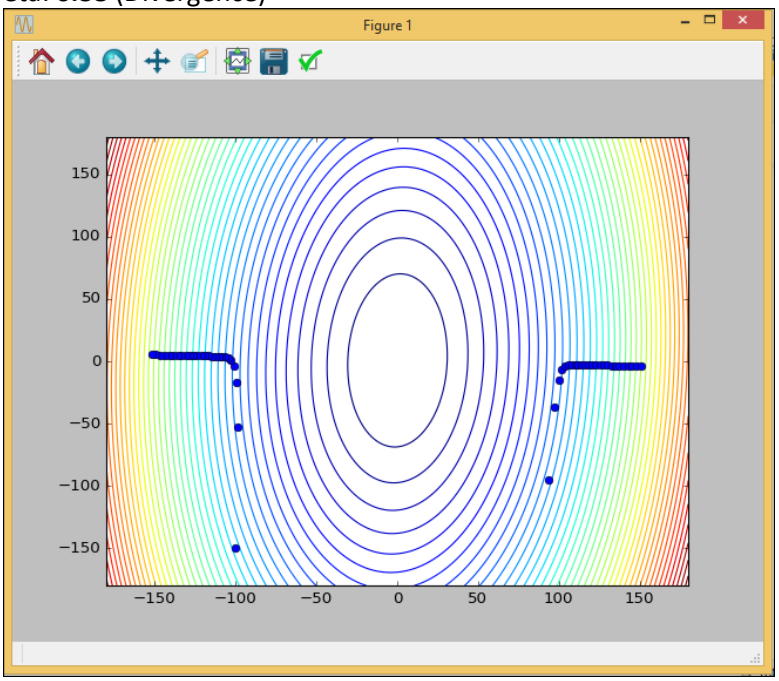


Iterations: 100
Start from (-100,-150)
eta: 0.1 (Convergence)

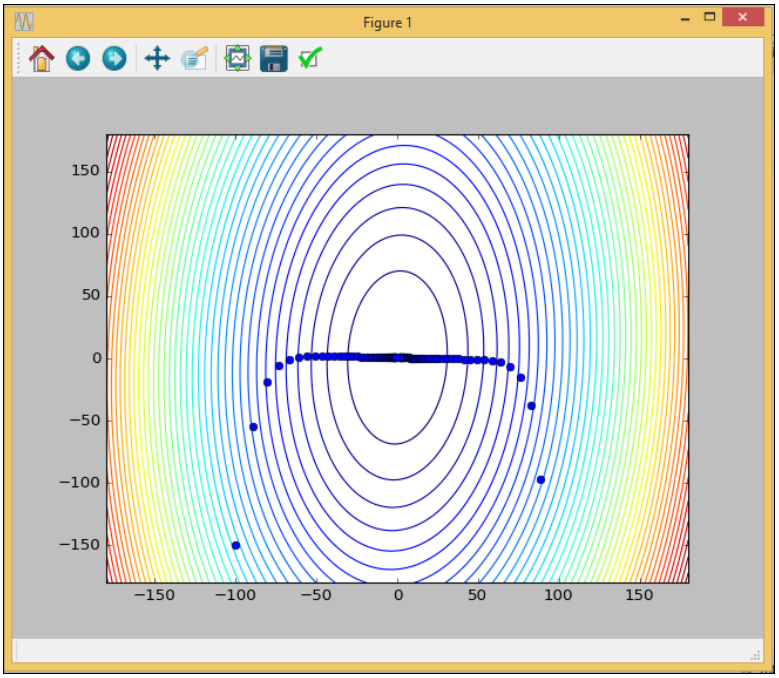


Iterations: 50
Start from (-100,-150)

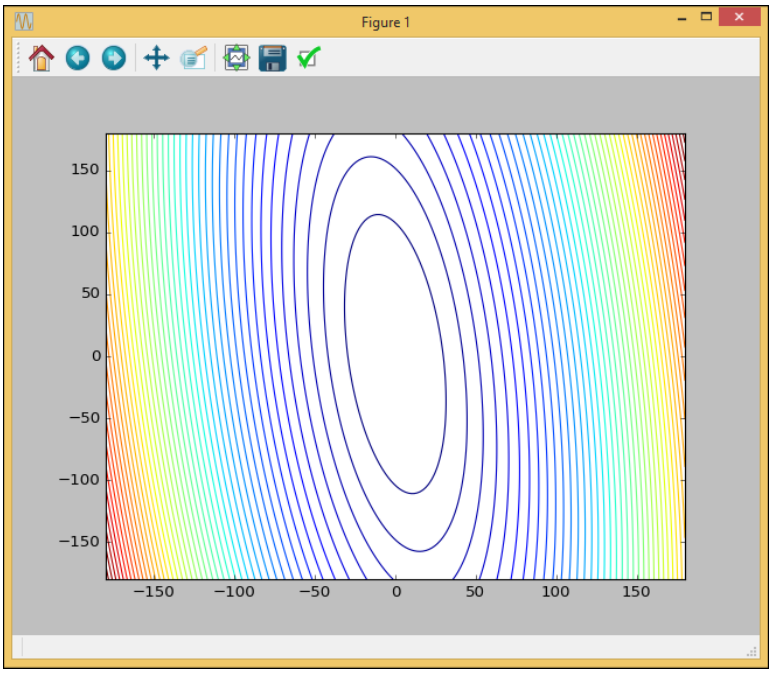
eta: **0.38** (Divergence)



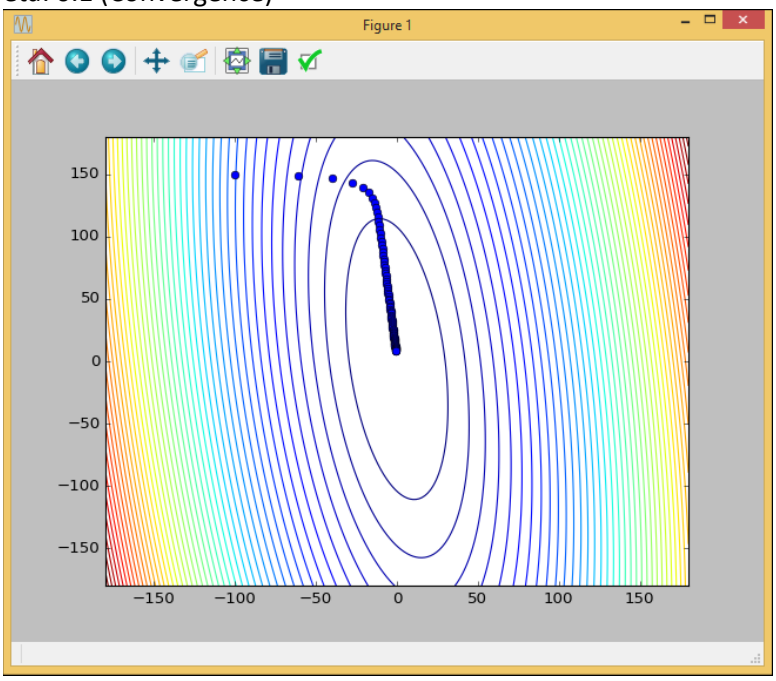
Iterations: 100
Start from (-100,-150)
eta: **0.37** (Overshoots but Convergence)



Example 2

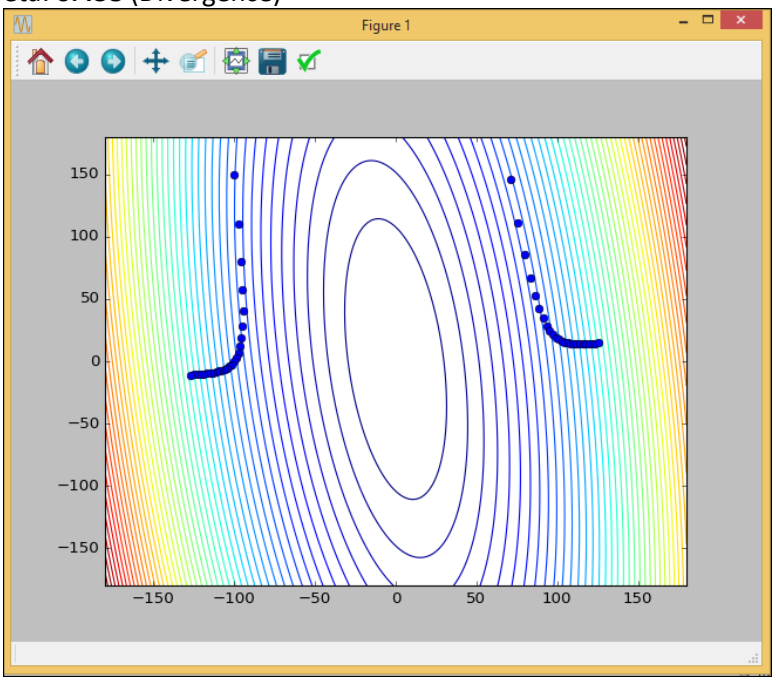


Iterations: 100
Start from (-100,150)
eta: 0.1 (Convergence)

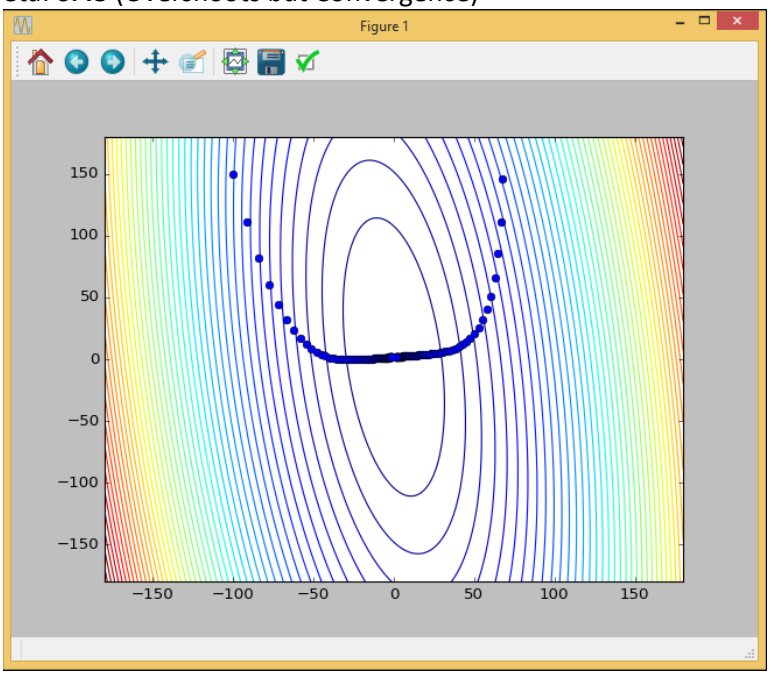


Iterations: 50
Start from (-100,150)

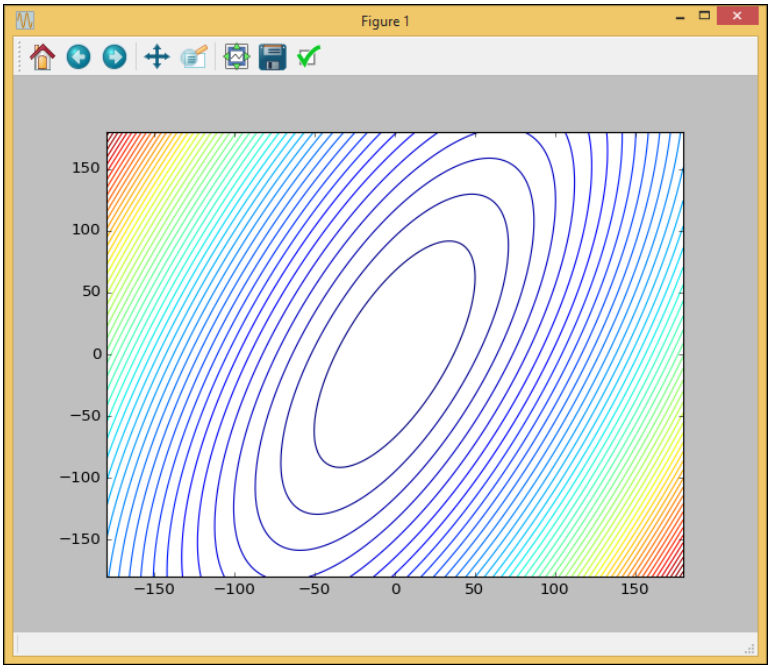
eta: **0.438** (Divergence)



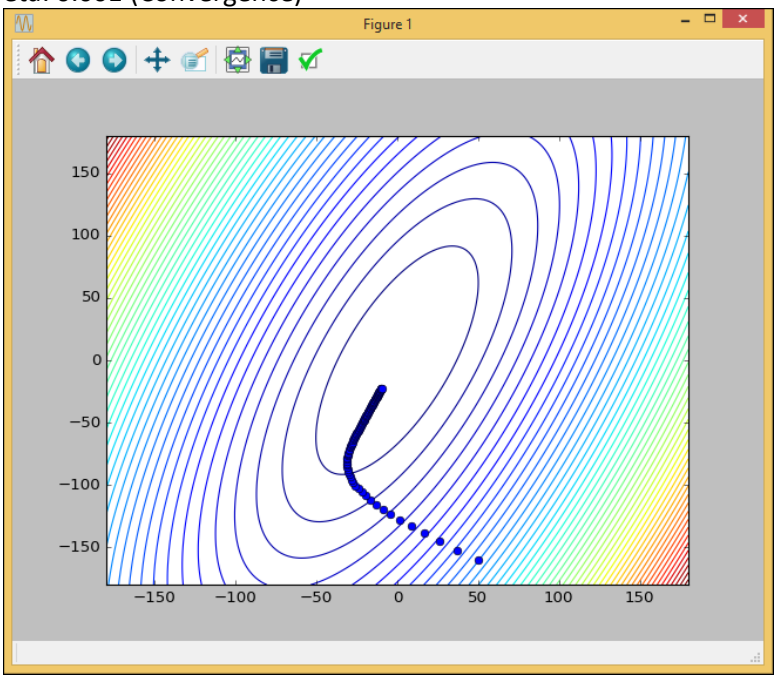
Iterations: 150
Start from (-100,150)
eta: **0.43** (Overshoots but Convergence)



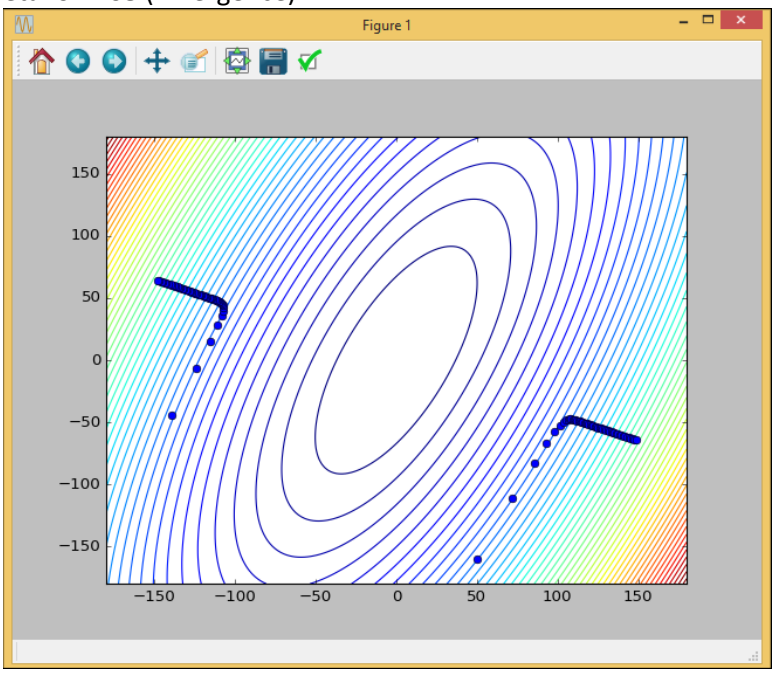
Example 3



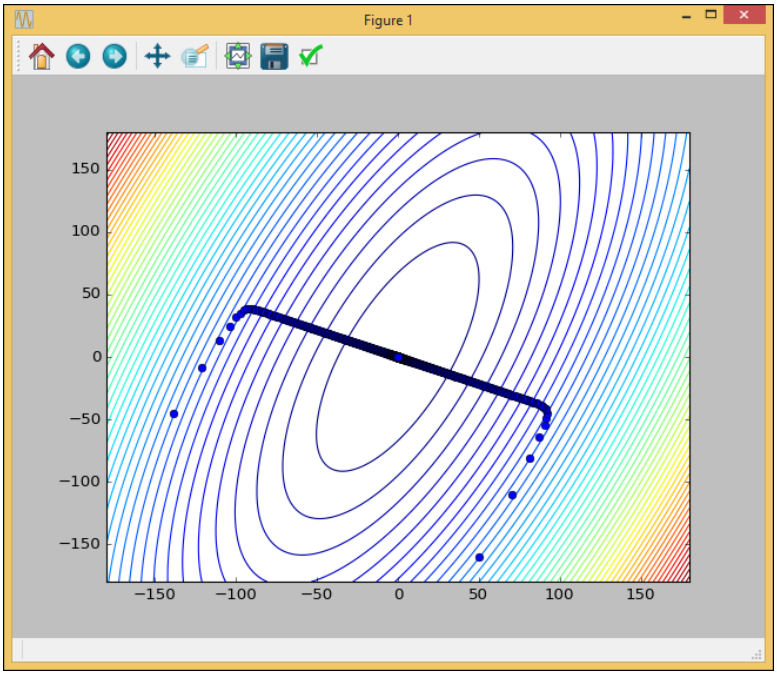
Iterations: 100
Start from (50,-160)
eta: 0.001 (Convergence)



Iterations: 100
Start from (50,-160)
eta: **0.1468** (Divergence)



Iterations: 1000
Start from (50,-160)
eta: **0.1461** (Overshoots but Convergence)



Example 4

Q6 Part d]

(d) (Commentary... nothing to do...) The key to realize, is that convergence rate depends, just like in the 1 dimensional case, on the curvature of the quadratic. The challenge, conceptually, is that unlike quadratics in 1-D whose curvature is given by the coefficient in front of the squared term, in two dimensions, quadratics do not have just one degree of curvature. Quadratics in two dimensions can be characterized by two curvature numbers. Basically, imagine sitting on a bowl that opens up slowly in one direction, and quickly in the perpendicular direction. It turns out that one need only consider two directions in two dimensions. The right directions are the major and minor axes of the ellipses that make up the *level sets* of the function.

In the IPython notebook near the top, you'll see commands for plotting the level sets, and you will notice that in every single case, you will get ellipsoidal level sets.

So, how do we find these directions, and the steepness in each?

Yes, our results in part c validate that the major axis of the ellipses dictate the least curvature and the direction in which the gradient descent tries to approach the minimum point so long as the step size is appropriate. The gradient of the curvature determines the steepness similar to the 1-D case. This was very interesting to observe in each example from part C. I was looking at https://www.math.ucla.edu/~jteran/270c.1.11s/notes_wk4.pdf and it looks like if the initial guess is an eigen vector, then the search direction will align perfectly with the solution and steepest descent will converge exactly in the one iteration.

Q6 Part e]

(e) Go back to your work in the second part of this problem, where you found the threshold value of η, where above that gradient descent diverges, and below it converges. Now use the `numpy.linalg.eig` command to find the eigenvalues and eigenvectors of the matrix Q. The command will return two numbers – these are the two eigenvalues of Q – and then two 2-dimensional vectors – these are the eigenvectors of Q. How do the eigenvalues compare to the threshold values of η that you found above?

In each of the examples in part c, it seems that the threshold eta value that converged was less than the largest eigenvalue. Anything higher than the largest eigen value caused divergence

Example 1:
Eigen Values:

```
In [64]: np.linalg.eig(Q)[0]
Out[64]: array([ 2.64390233,  0.51801378])
```

Example 2:
Eigen Values:

```
In [71]: np.linalg.eig(Q)[0]
Out[71]: array([ 2.29256455,  0.156297  ])
```

Example 3:
Eigen Values:

```
In [113]: np.linalg.eig(Q)[0]
Out[113]: array([ 6.82536906,  0.82166975])
```

Results:

Example #	Smallest EV	Largest EV	1 / L.EV	Observed Threshold (Divergence)
1	0.51801378	2.64390233	0.378228798	**0.38**
2	0.156297	2.29256455	0.436192734	**0.438**
3	0.82166975	6.82536906	0.146512224	**0.1468**

Q6 Part f]

(f) Playing with eigenvalues and eigenvectors: if the eig command returns eigenvalues λ_1 and λ_2, and eigenvectors v_1 and v_2, compute Qv_1 and Qv_2. Notice anything? The relationship you notice is precisely the defining equation of eigenvalues and eigenvectors.

```
In [1]: import matplotlib.pyplot as plt
        import numpy as np

In [2]: Q = np.array([[ 1.,  2.], [ 4.,  3.]])

In [3]: eig_val, eig_vec = np.linalg.eig(Q)
        v1 = eig_vec[:,0]
        v2 = eig_vec[:,1]
        lambda1 = eig_val[0]
        lambda2 = eig_val[1]

In [4]: print eig_val
        print eig_vec

        [-1.  5.]
        [[-0.70710678 -0.4472136 ]
         [ 0.70710678 -0.89442719]]
```

v1 v2

```
In [5]: # Does Q.v1 = lambda1(v1)? - YES
        print Q.dot(v1)
        print lambda1*v1

        [ 0.70710678 -0.70710678]
        [ 0.70710678 -0.70710678]

In [6]: # Does Q.v2 = lambda2(v2)? - YES
        print Q.dot(v2)
        print lambda2*v2

        [-2.23606798 -4.47213595]
        [-2.23606798 -4.47213595]

In [7]: # Does determinant |lambda1*Identity2x2 - Q | = 0?
        print np.linalg.det(lambda1*np.eye(2) - Q)

        0.0

In [8]: # Does determinant |lambda2*Identity2x2 - Q | = 0?
        print np.linalg.det(lambda1*np.eye(2) - Q)

        0.0
```

The following observations can be made:

$Q\overline{v} = \lambda\overline{v}$

$Q\overline{v} - \lambda\overline{v} = \overline{0}$

when λ is an eigenvalue, $\left|(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - Q)\right| = 0$

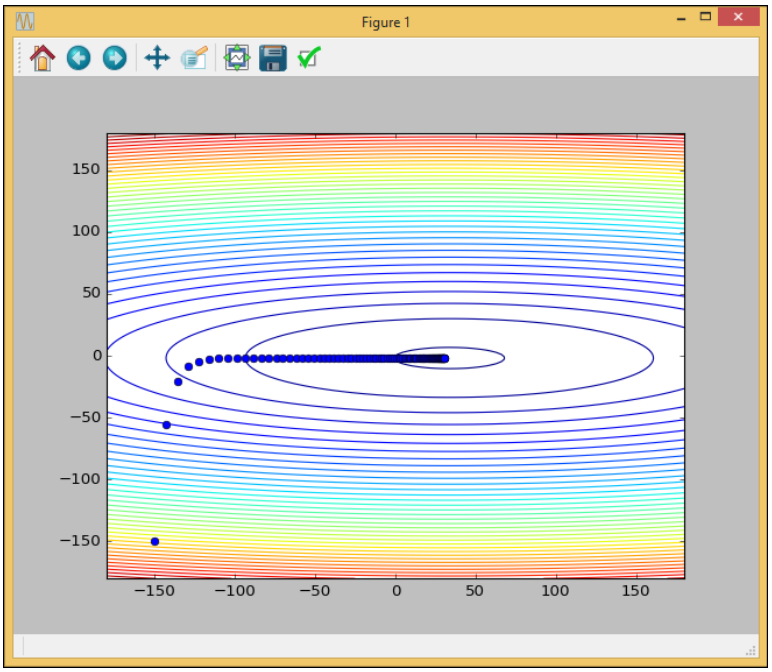
Q6 Part g]

(g) (Bonus) The above discussion and exercises show that the *largest eigenvalue* governs how large our step size should be. The larger the eigenvalue, the more the curvature and hence the smaller the stepsize. What role does the smallest eigenvalue play? It turns out that the ratio of largest to smallest, controls how quickly we converge.

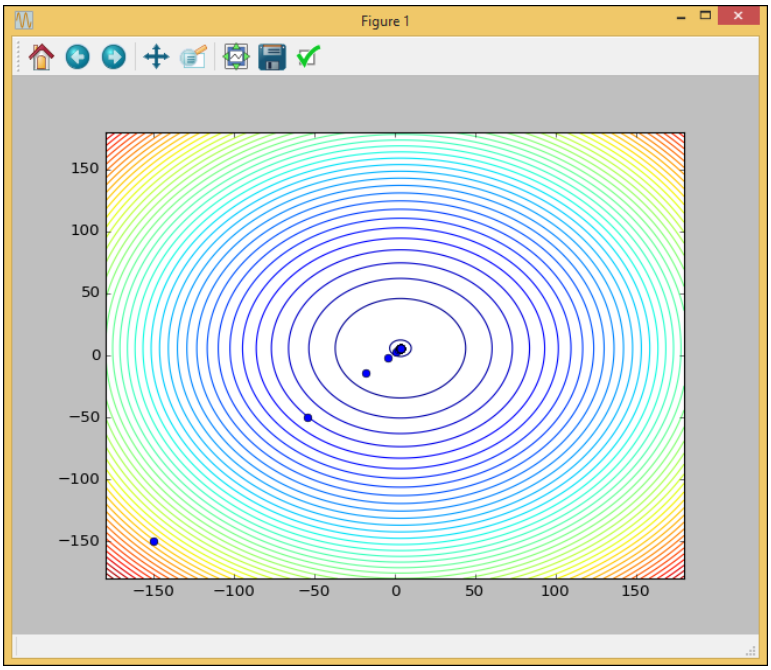
Generate examples where the largest eigenvalue is fixed (and hence you are using the same stepsize), and vary the second eigenvalue from very small (close to zero, but positive), to very big – almost as big as the largest eigenvalue. Plot the convergence of gradient descent. Can you discern a relationship?

Largest EigenValue fixed at **0.16**
Iterations: 500
Start from (-150,-150)
eta: **2** (Convergence)

```
In [69]: np.linalg.eig(Q)[0]
Out[69]: array([ 0.01,  0.16])
```



```
In [95]: np.linalg.eig(Q)[0]
Out[95]: array([ 0.156025,  0.16    ])
```



When the eigenvalues are very close to each other, the elliptical level sets are closer to circular which makes sense. But convergence happens a lot faster and in very few iterations (distance between blue dots). I was able to test convergence in 5 iterations.

```
# now we can compute the sequence of updates
for i in np.arange(5)+1:
    # we have to choose a step size
    eta = 2
    x_new, y_new = np.array([x_step[i-1],y_step[i-1]]) - eta*gradient(x_step[-1],y_step[i-1],Q,c
    x_new = np.array([x_new])
    y_new = np.array([y_new])
    x_step = np.concatenate((x_step,x_new))
    y_step = np.concatenate((y_step,y_new))
```