

ridge

April 4, 2014

To view the description of this assignment see <http://www.cs.ubc.ca/~nando/540-2013/lectures/homework1.pdf>

```
In [1]: %pylab inline
```

Populating the interactive namespace from numpy and matplotlib

```
In [1]: file_path = 'datasets/prostate.data'
X = loadtxt(file_path, skiprows=1)
with open(file_path, 'r') as myfile:
    label_names = myfile.readline().split()
y = X[:, -1]
X = X[:, 0:-1]
```

NameError

Traceback (most recent call last)

```
<ipython-input-1-93b2337044cc> in <module>()
    1 file_path = 'datasets/prostate.data'
----> 2 X = loadtxt(file_path, skiprows=1)
    3 with open(file_path, 'r') as myfile:
    4     label_names = myfile.readline().split()
    5 y = X[:, -1]
```

NameError: name 'loadtxt' is not defined

```
In [3]: ytrain, ytest = y[0:50], y[50:]
Xtrain, Xtest = X[0:50], X[50:]
```

```
In [4]: Xbar = mean(Xtrain, axis=0)
Xstd = std(Xtrain, axis=0)
ybar = mean(ytrain)
ytrain = ytrain - ybar
Xtrain = (Xtrain - Xbar) / Xstd
```

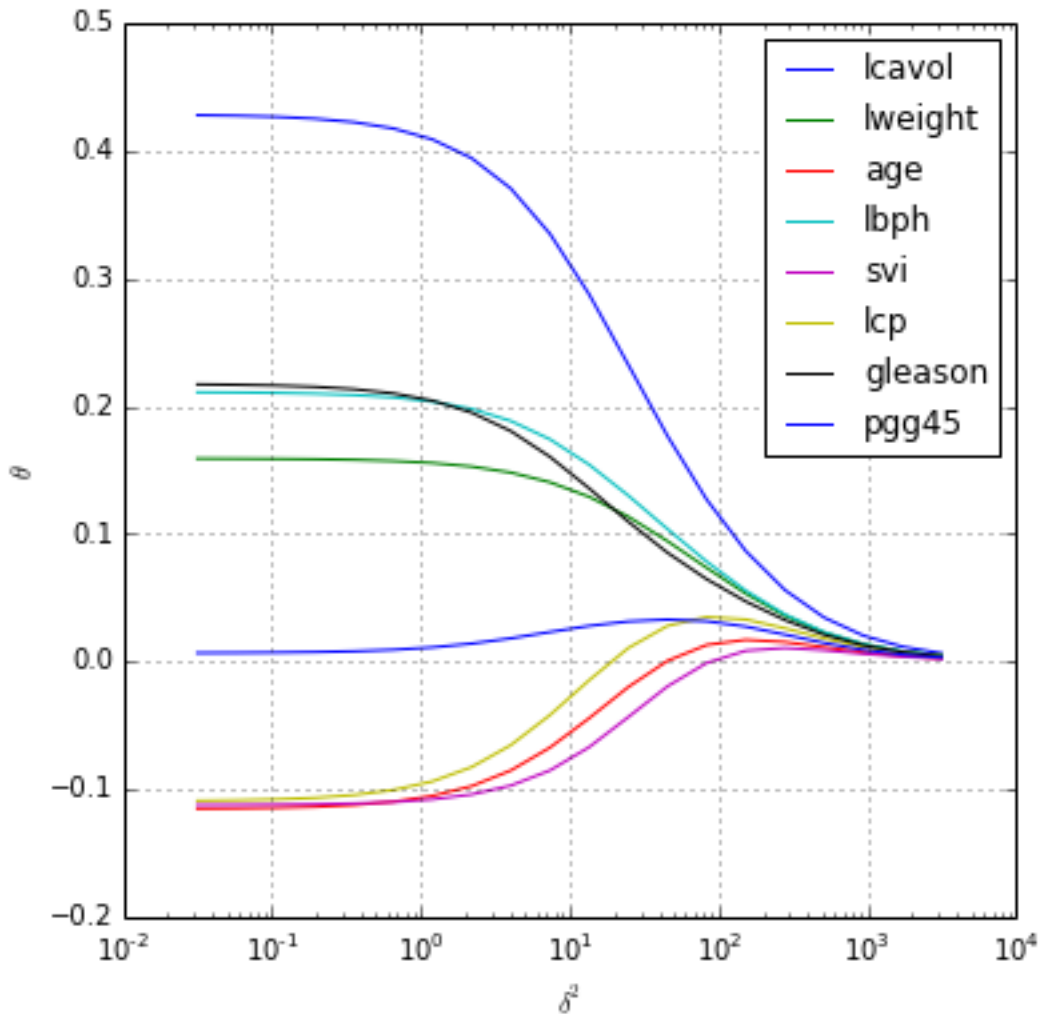
```
In [5]: def ridge(X, y, d2):
    return dot(dot(inv(dot(X.T, X) + d2*eye(X.shape[1])), X.T), y)
```

```
In [6]: d2range = logspace(-1.5, 3.5, num=20)
thetas = array([ridge(Xtrain, ytrain, d2) for d2 in d2range])
figure(figsize=(6,6))
xscale('log')
```

```

grid()
xlabel(r'$\delta^2$'); ylabel(r'$\theta$')
plot(d2range, thetas)
legend(label_names)
show()

```



```

In [7]: testerror = []
        trainerror = []
        min_err = None
        for i, theta in enumerate(thetas):
            yhatstest = ybar + dot((Xtest - Xbar) / Xstd, theta)
            yhatstrain = ybar + dot(Xtrain, theta)
            trainerror.append(norm((ytrain + ybar) - yhatstrain, ord=2) / norm(ytrain + ybar, ord=2))
            testerror.append(norm(ytest - yhatstest, ord=2) / norm(ytest, ord=2))
            max_err = max((trainerror[-1], testerror[-1]))
            if min_err == None or max_err < min_err:
                min_err = max_err
                best_delta = d2range[i]

```

```

print("best delta:", best_delta)
figure(figsize=(7,7))
xscale('log')
grid()
xlabel(r'$\delta^2$'); ylabel('error')
plot(d2range, trainerror, '-bo', linewidth=2)
plot(d2range, testerror, '-g^', linewidth=2)
legend(["Train", "Test"])
show()

```

('best delta:', 13.538761800225432)

