数据科学基础Ⅰ
（Matlab）

—— 东北大学 ——

多元线性回归

多元线性回归：一个因变量，多个自变量。

二元线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

(观察到的y)

β_0

ε_i

回归面

(x_1, x_2)

y

x_1

x_2

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
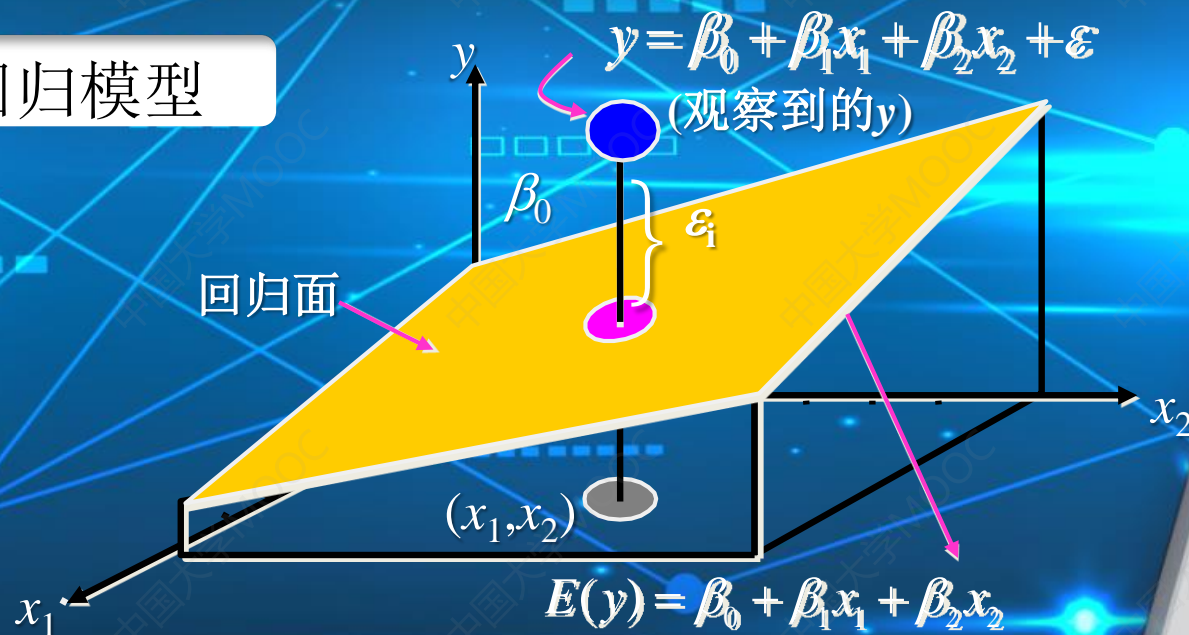
MATLAB方法

```
load carsmall; %一个汽车数据集
X = [Weight, Horsepower, Acceleration];
Mdl = fitlm(X, MPG)
```

MATLAB方法

```
Mdl =

Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                  Estimate       SE        tStat       pValue
                  _____    _____    _____    _____

    (Intercept)     47.977      3.8785       12.37    4.8957e-21
    x1           -0.0065416    0.0011274    -5.8023    9.8742e-08
    x2            -0.042943     0.024313    -1.7663       0.08078
    x3            -0.011583      0.19333   -0.059913      0.95236


Number of observations: 93, Error degrees of freedom: 89
Root Mean Squared Error: 4.09
R-squared: 0.752,  Adjusted R-Squared 0.744
F-statistic vs. constant model: 90, p-value = 7.38e-27
```
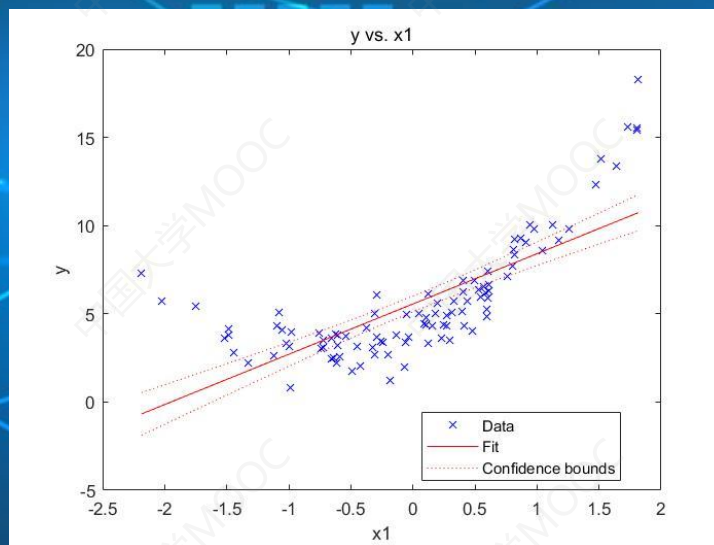
广义线性回归

一个简单例子

```
%生成测试数据
X=randn(100,1);
y=2 * X.^2 + 3 * X + 4 +
randn(100,1);
%使用一元线性回归
mdl1 = fitlm(X, y)
mdl1.plot;
```


o 拟合的程度不够，如何解决？

广义线性回归

解决方法：使用2次函数模型 $y_i = a + bx_i + cx_i^2$

```matlab
%生成测试数据
X=randn(100,1);
y=2 * X.^2 + 3 * X + 4 + randn(100,1);
%增加一维
X=[X,X.^2];
%使用线性回归
mdl2 = fitlm(X, y)
```

解决方法：使用2次函数模型

```
mdl2 =


Linear regression model:
    y ~ 1 + x1 + x2

Estimated Coefficients:
                Estimate      SE       tStat       pValue
                _____    _____    _____    _____

    (Intercept)    3.958     0.12968    30.522     1.5591e-51
    x1             2.8768    0.10834    26.553     2.7499e-46
    x2             1.9366    0.083777   23.116     3.0738e-41
```
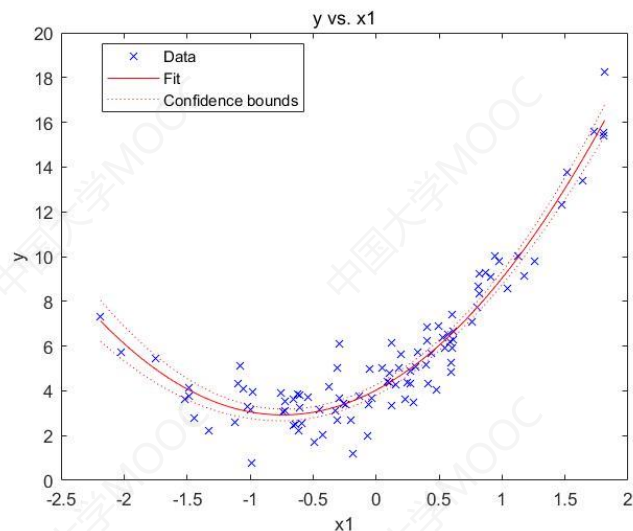
直接使用fitlm函数

```
%生成测试数据
X=randn(100,1);
y=2 * X.^2 + 3 * X + 4 +
randn(100,1);
%利用参数设定模型
mdl2 = fitlm(X, y, 'y ~ 1 + x1 +
x1^2')
mdl2.plot;
```

广义线性回归

案例：汽车数据集的多元广义线性回归

```
load carsmall;

X = [Weight, Acceleration];

Mdl = fitlm(X, MPG, 'y ~ 1 + x1*x2 + x1^2')
```

设定
模型

广义线性回归

案例：汽车数据集的多元广义线性回归

```
Mdl =


Linear regression model:
    y ~ 1 + x1*x2 + x1^2

Estimated Coefficients:
                 Estimate        SE         tStat       pValue
                 _____     _____    _____    _____

    (Intercept)     48.906       12.589      3.8847    0.00019665
    x1           -0.012781    0.0060312     -2.1192      0.036857
    x2             0.54418      0.57125      0.95261      0.34337
    x1:x2       -0.00010892   0.00017925     -0.6076        0.545
    x1^2         9.7518e-07   7.5389e-07      1.2935      0.19917
```