



数据科学基础 I (Matlab)

—— 东北大学 ——





非线性回归



非线性回归是指回归函数为非线性函数

```
load carbig
tbl = table(Horsepower,Weight,MPG);
% 设定回归模型
modelfun = @(b,x)b(1) + b(2)*x(:,1).^b(3) +
b(4)*x(:,2).^b(5);
% 设定参数初值
beta0 = [-50 500 -1 500 -1];
mdl = fitnlm(tbl,modelfun,beta0)
```




非线性回归

mdl =

Nonlinear regression model:

$$\text{MPG} \sim b1 + b2 * \text{Horsepower}^{b3} + b4 * \text{Weight}^{b5}$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
b1	-49.383	119.97	-0.41164	0.68083
b2	376.43	567.05	0.66384	0.50719
b3	-0.78193	0.47168	-1.6578	0.098177
b4	422.37	776.02	0.54428	0.58656
b5	-0.24127	0.48325	-0.49926	0.61788



非线性回归



案例：分析婴幼儿年龄和头围的关系



数据

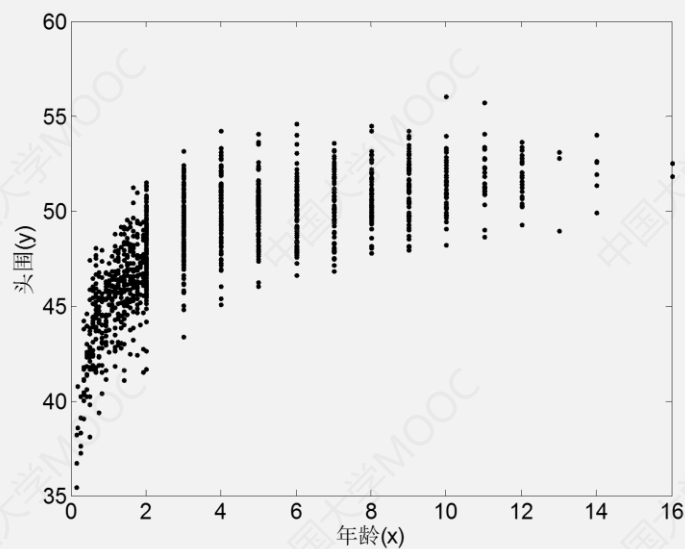
年龄	头围
11	50.90952
1.666667	50.4282
10	51.35181
3	50.27417
3	48.52064
5	50.30917
4	52.34006
8	49.05821
8	51.17692
0.916667	46.21404



非线性回归



案例：分析婴幼儿年龄和头围的关系





非线性回归



模型选择：根据数据的特点选择模型



负指数函数

$$y = \beta_1 e^{\frac{\beta_2}{x + \beta_3}}$$



双曲线函数

$$y = \frac{x + \beta_1}{\beta_2 x + \beta_3}$$



幂函数

$$y = \beta_1 (x + \beta_2)^{\beta_3}$$



非线性回归



模型选择：根据数据的特点选择模型



Logistic 曲线函数

$$y = \frac{\beta_1}{1 + \beta_2 e^{-(x + \beta_3)}}$$



对数函数


$$y = \beta_1 + \beta_2 \ln(x + \beta_3)$$



非线性回归

```
HeadData = xlsread('headcir.xls');  
x = HeadData(:, 1);  
y = HeadData(:, 2);  
% 理论回归方程对应的匿名函数  
HeadCir = @(beta,  
x)beta(1)*exp(beta(2)./(x+beta(3)));  
% 参数初值  
beta0 = [50, 0.1, 0.1];  
Mdl = fitnlm(x, y, HeadCir, beta0)  
% 绘制结果
```

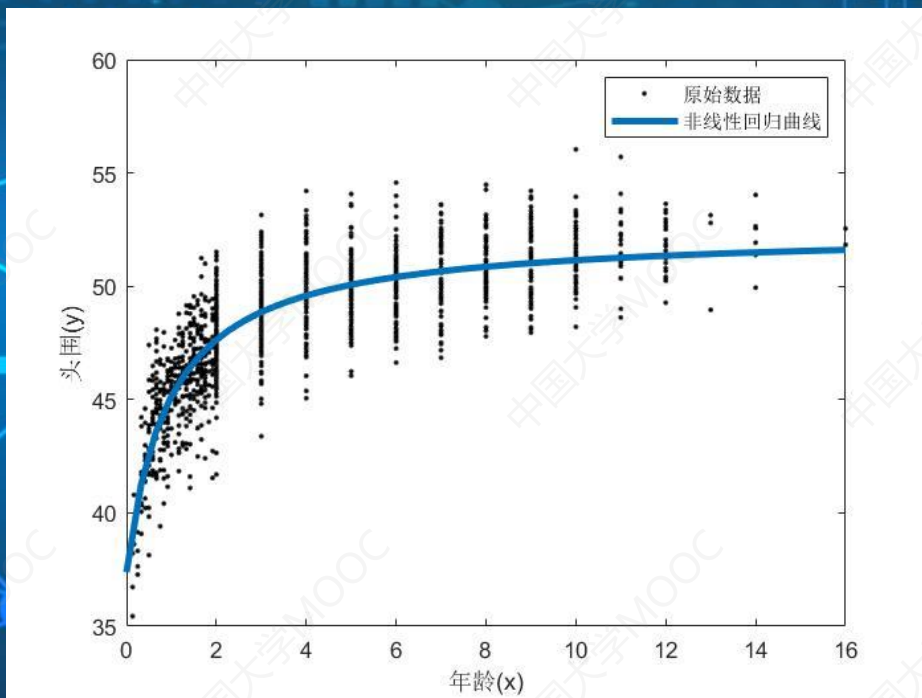
```
xnew = linspace(0, 16, 50)';  
ynew = predict(Mdl, xnew);  
figure;  
plot(x, y, 'k.');
```



```
hold on;  
plot(xnew, ynew, 'linewidth', 3);  
xlabel('年龄(x)');  
ylabel('头围(y)');  
legend('原始数据','非线性回归曲线');
```




非线性回归





非线性回归



讨论：

- 模型的选择需要通过对数据的观察来确定，适合的模型并不一定是唯一的。
- 非线性回归中的未知参数选取是难点，同样需要通过对数据的观察来确定。
- 对数据进行探索性分析应该是选择模型前的重要工作。



模型选择



欠拟合vs过拟合

```
x=randn(9,1);  
y=2*x.^2 + 3*x + 4 + randn(9,1);  
figure;  
mdl1 = fitlm(x, y);  
subplot(1,3,1);  
mdl1.plot;  
mdl2 = fitlm(x, y, 'y ~ 1 + x1 + x1^2');  
subplot(1,3,2);  
mdl2.plot;  
mdl3 = fitlm(x, y, 'y ~ 1 + x1 + x1^2 + x1^3 + x1^4 + x1^5 + x1^6 + x1^7 + x1^8 + x1^9');  
subplot(1,3,3);  
mdl3.plot;
```



模型选择



欠拟合



过拟合

