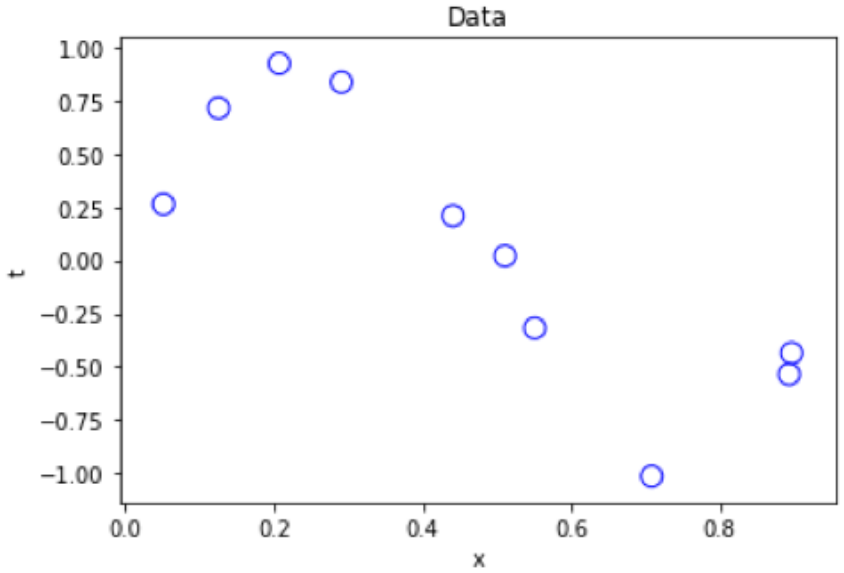
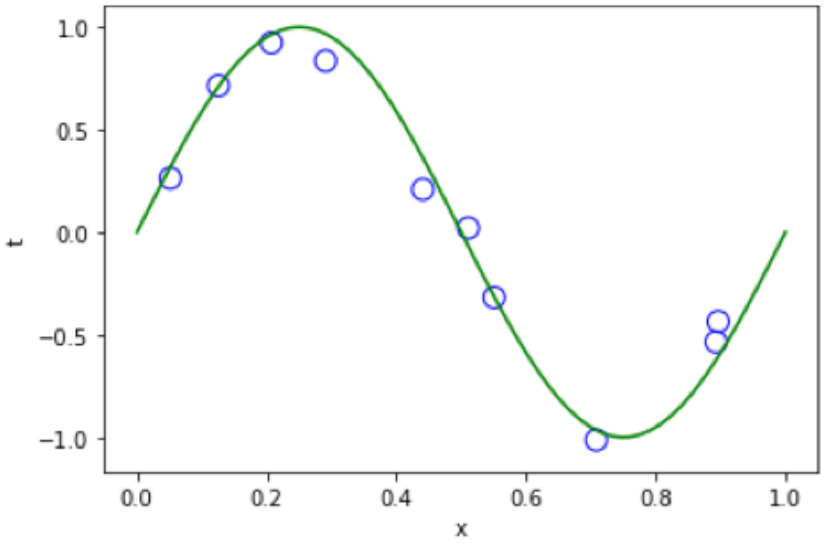
模型评估和模型选择

2019年8月10日

多项式回归

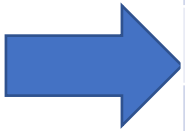




Input	Output
$x^{(1)} = 0.03$	$y^{(1)} = 0.19$
$x^{(2)} = 0.78$	$y^{(2)} = -0.91$
$x^{(3)} = 0.25$	$y^{(3)} = 0.88$
…	…
$x^{(15)} = 0.64$	$y^{(15)} = -0.91$

多项式回归数据特征

Input	Output
$x^{(1)} = 0.03$	$y^{(1)} = 0.19$
$x^{(2)} = 0.78$	$y^{(2)} = -0.91$
$x^{(3)} = 0.25$	$y^{(3)} = 0.88$
…	…
$x^{(15)} = 0.64$	$y^{(15)} = -0.91$



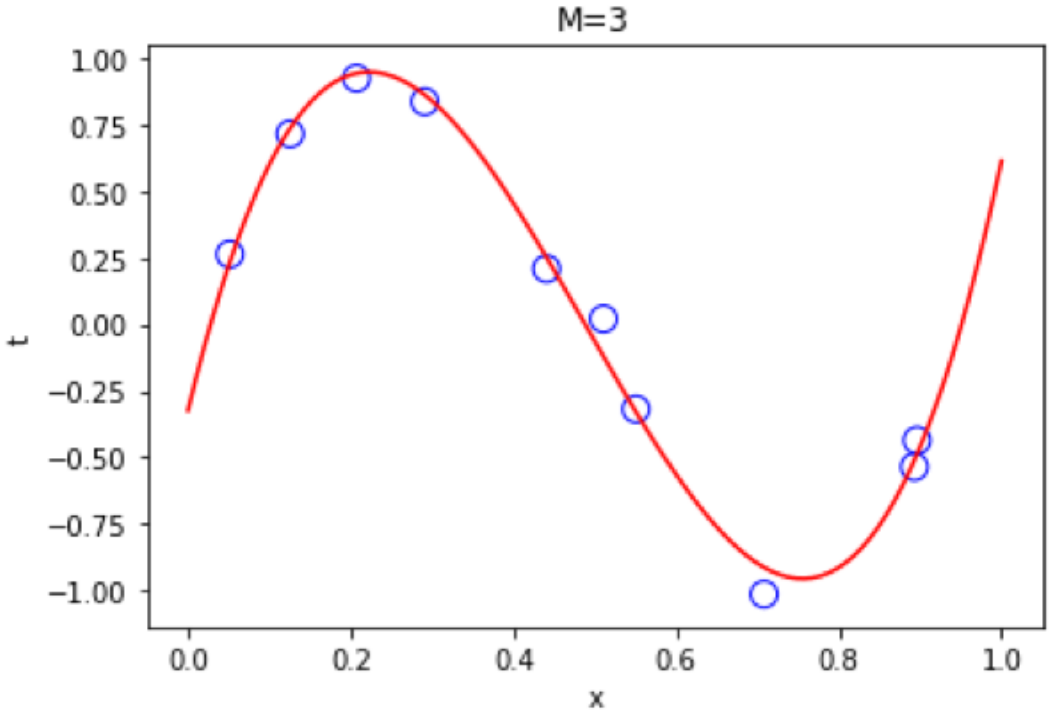
Input			Output
$x_1^{(1)} = 0.03$	$x_2^{(1)} = 0.001$	$x_3^{(1)} = 0.00005$	$y^{(1)} = 0.1907$
$x_1^{(2)} = 0.78$	$x_2^{(2)} = 0.608$	$x_3^{(2)} = 0.47455$	$y^{(2)} = -0.9118$
$x_1^{(3)} = 0.25$	$x_2^{(3)} = 0.0625$	$x_3^{(3)} = 0.01562$	$y^{(3)} = 0.8866$
…	…	…	…
$x_1^{(15)} = 0.64$	$x_2^{(15)} = 0.409$	$x_3^{(15)} = 0.26214$	$y^{(15)} = -0.9104$

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

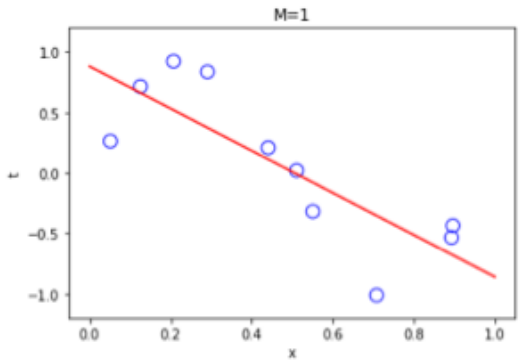
M如何选取?

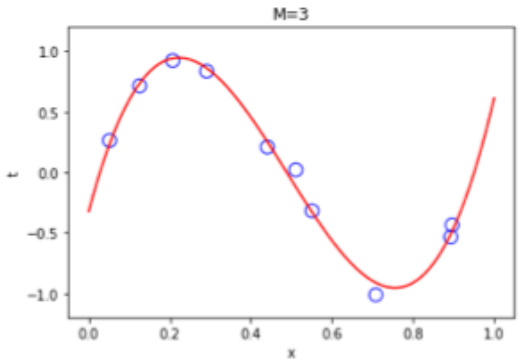
多项式回归

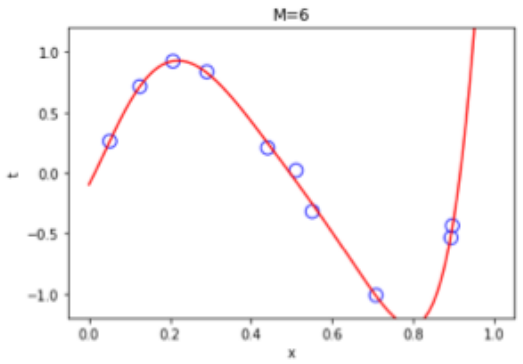
- 多项式阶数$M = 3$

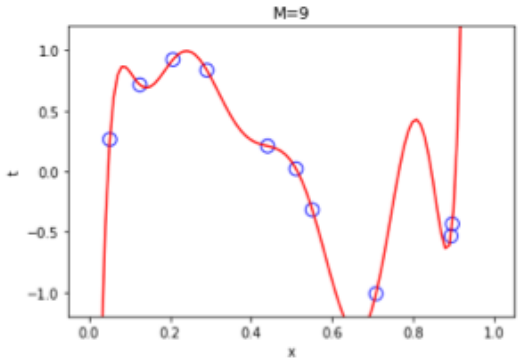


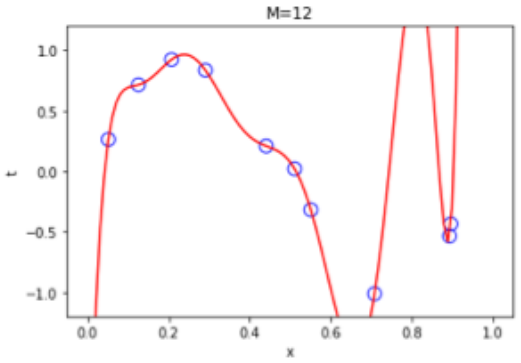
多项式回归

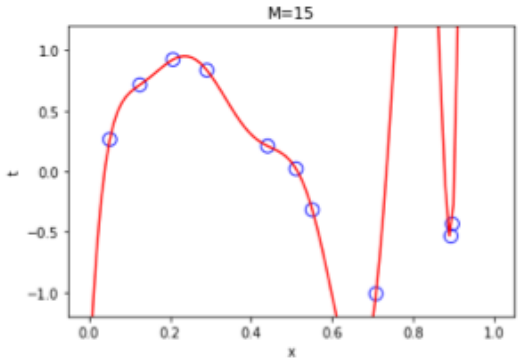






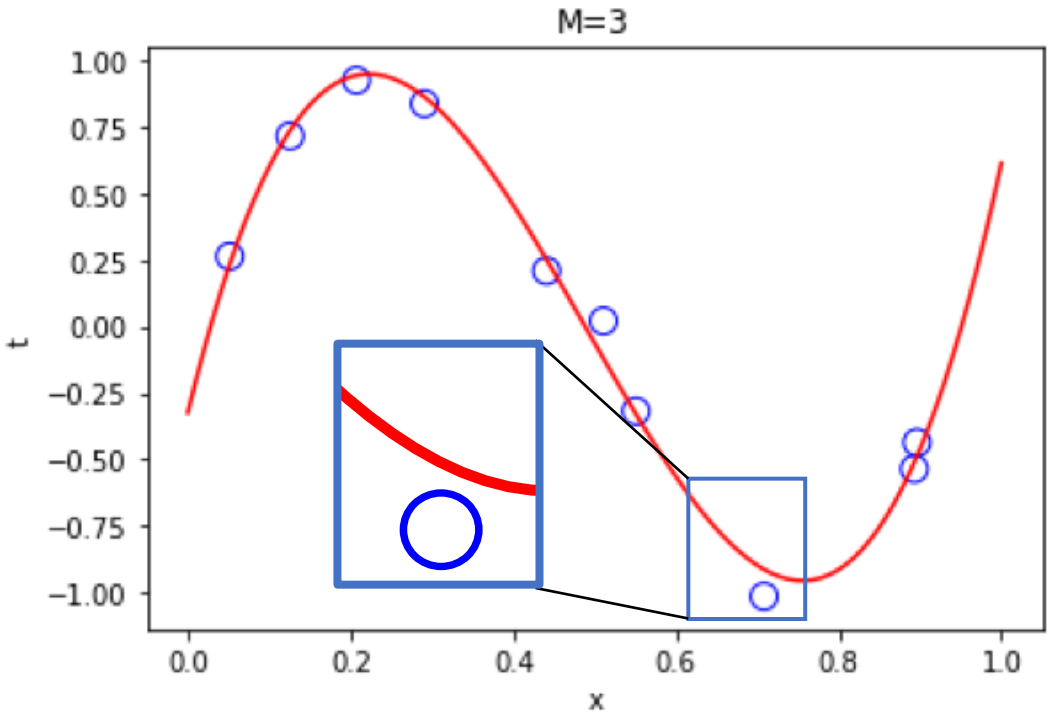




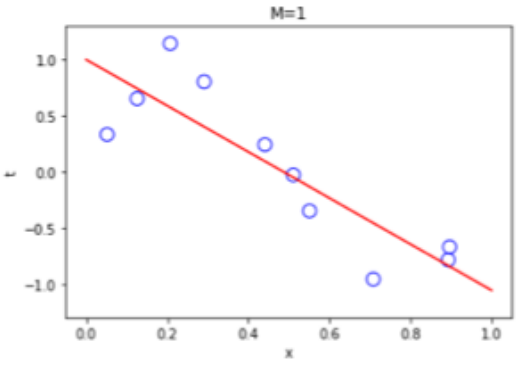


训练误差

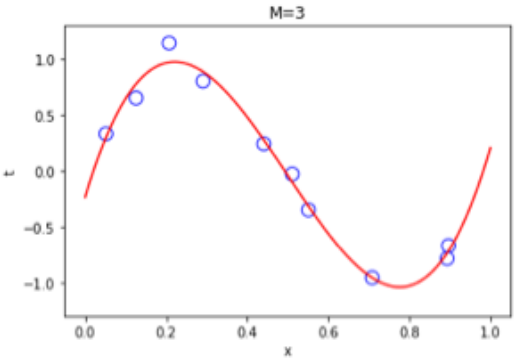
- 多项式阶数$M = 3$
- 训练误差$J_{M=3} = 0.98$



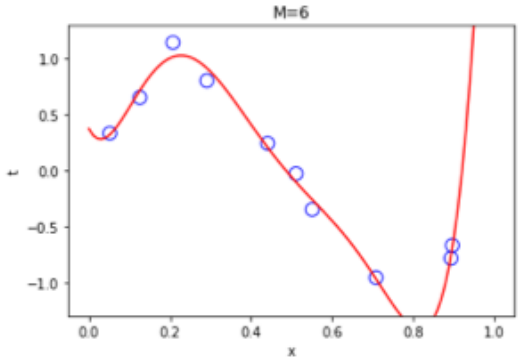
多项式回归



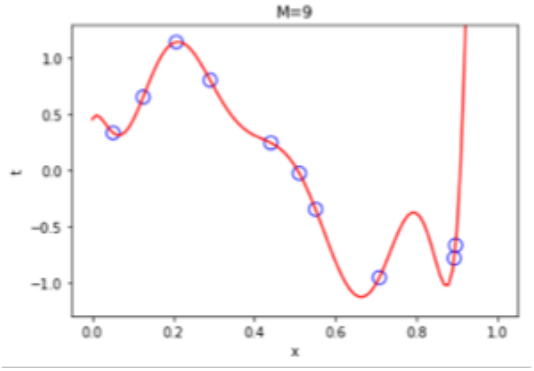
$J_{M=1} = 0.1166$



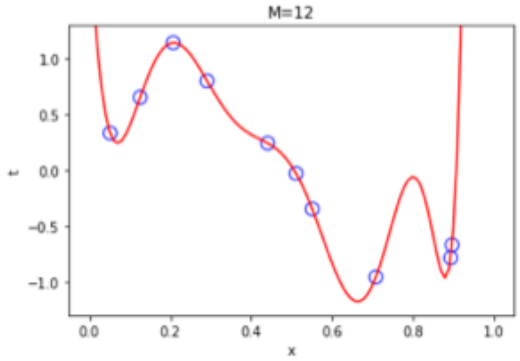
$J_{M=3} = 0.0065$



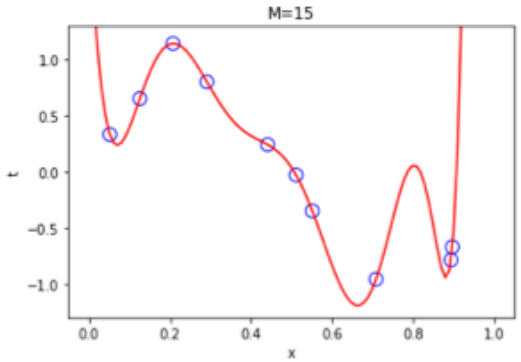
$J_{M=6} = 0.0049$



$J_{M=9} = 0$

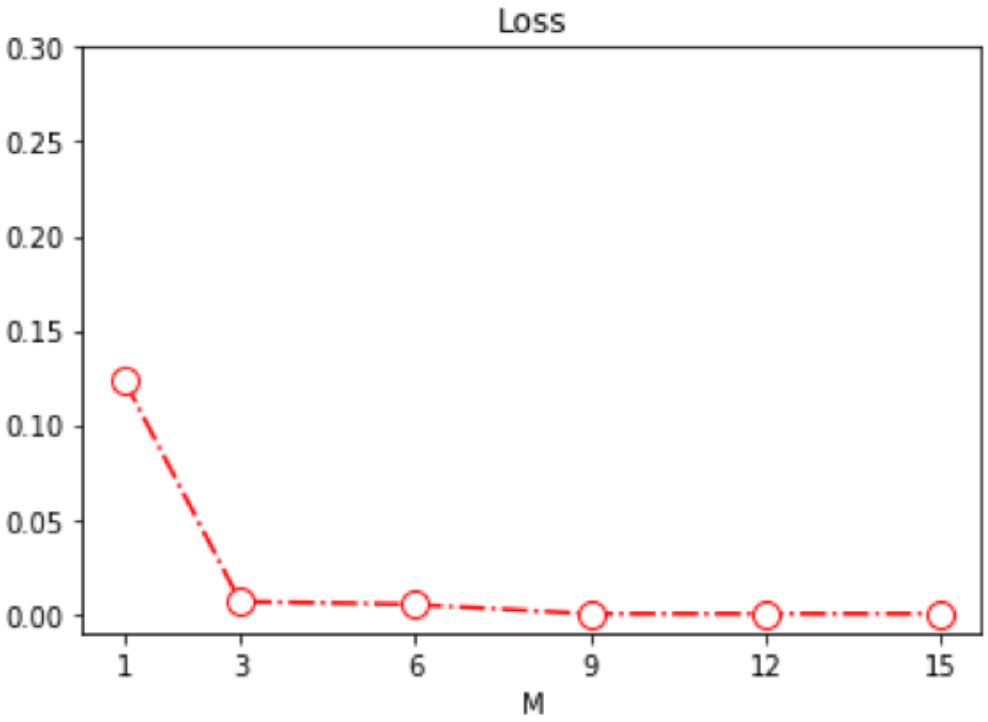


$J_{M=12} = 0$



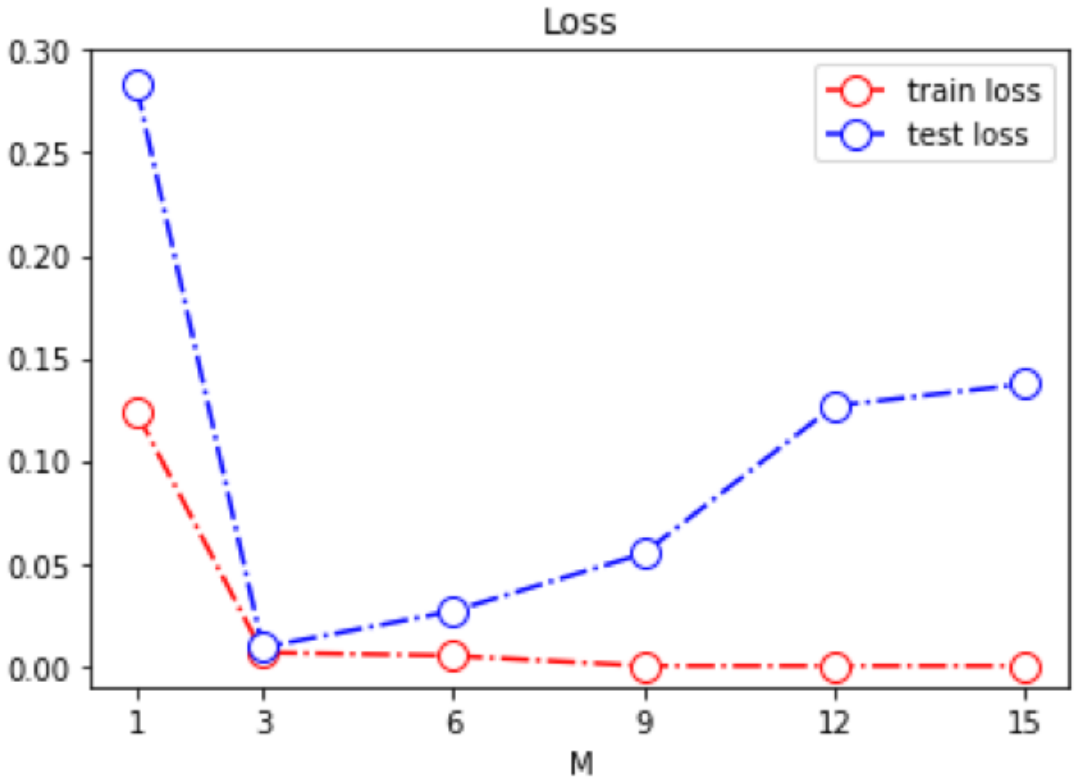
$J_{M=15} = 0$

训练误差曲线



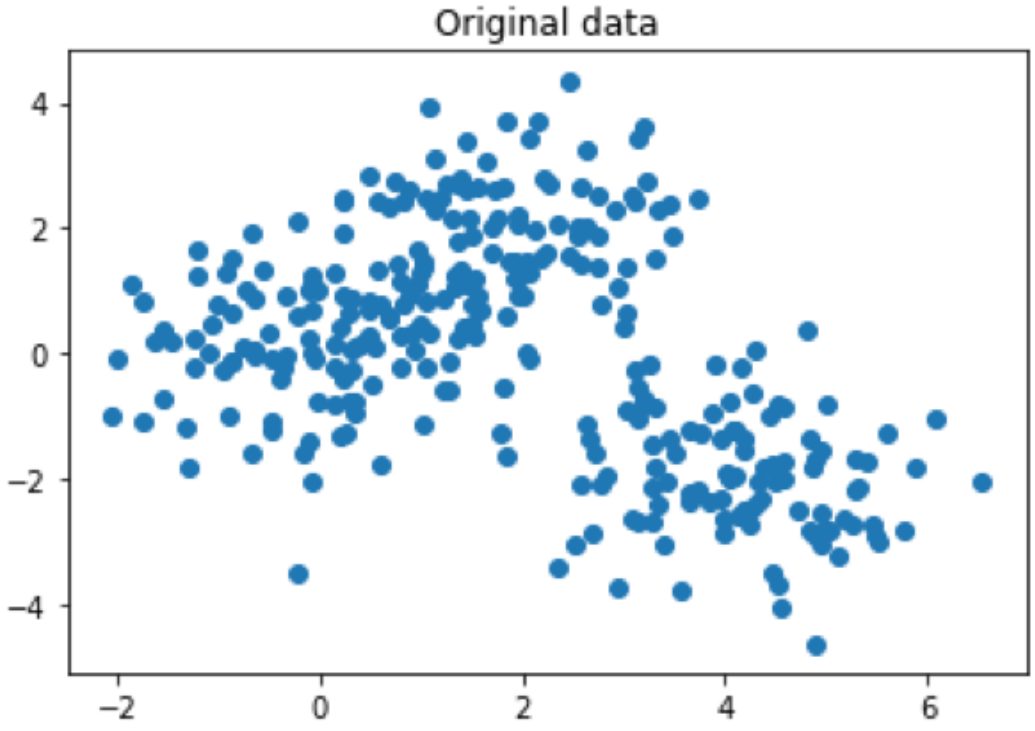
训练和测试数据集

训练和训练误差曲线

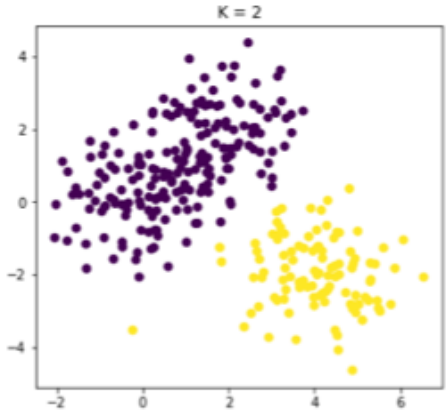

Loss
train loss
test loss
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
3
6
9
12
15
M

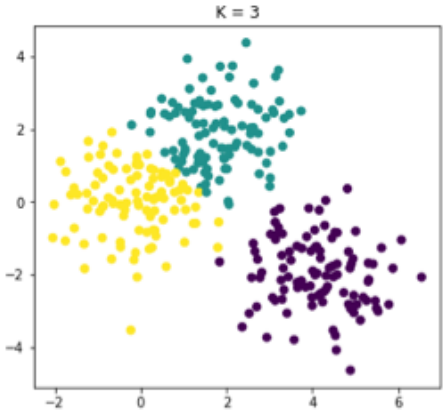
K均值聚类

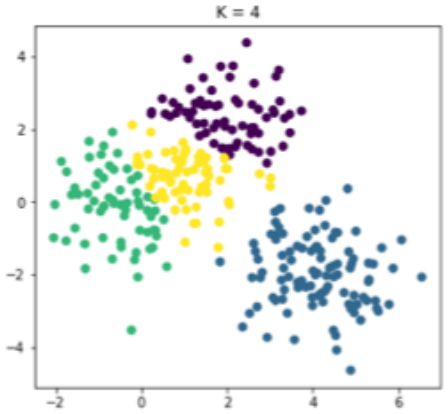
- 原始数据

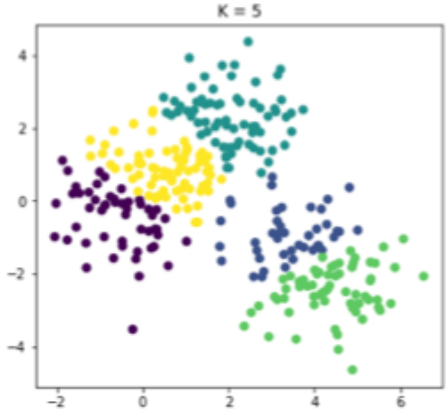


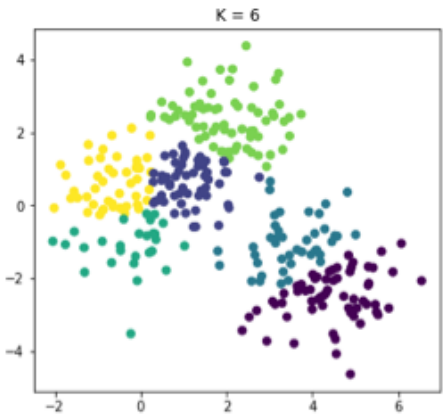
不同K值聚类结果

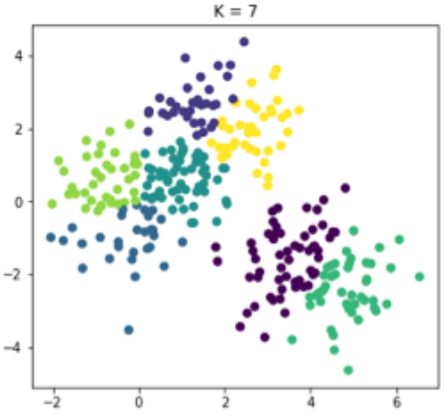












手肘法