

离散参数：

- 1. 序号&住宅区名：无关属性，删除
- 2. 地块信息

1	2	3	4
上海周边-花桥	松江-泗泾	浦东-周康	闸北-上海火车站
浦东-川沙	松江-松江新城	宝山-顾村	普陀-真如
浦东-临港新城	松江-洞泾	浦东-花木	黄浦-董家渡
奉贤-海湾旅游区	闵行-浦江	松江-泗泾	长宁-中山公园
上海周边-昆山	浦东-北蔡	浦东-惠南	浦东-陆家嘴
青浦-白鹤	松江-车墩	闵行-春申	徐汇-斜土路
金山-朱泾	金山-金山新城	浦东-世纪公园	黄浦-人民广场
青浦-重固	闵行-马桥	宝山-通河新村	长宁-天山
上海周边-嘉兴	青浦-青浦新城	浦东-浦东世博	徐汇-建国西路
	嘉定-马陆	松江-松江老城	浦东-梅园
	青浦-华新镇	嘉定-嘉定城区	长宁-古北
	青浦-赵巷	闵行-古美罗阳	静安-曹家渡
	宝山-杨行	浦东-南码头	徐汇-徐家汇
	杨浦-新江湾城	浦东-金桥	普陀-曹杨
	青浦-朱家角	浦东-塘桥	卢湾-打浦桥
	嘉定-安亭	普陀-长风	浦东-潍坊
	嘉定-南翔	文华苑二期	黄浦-打浦桥
	嘉定	浦江名邸	
	浦东-三林	浦东-金杨	
	闵行星河湾	闵行-七宝	
	宝山-共康	绿城黄浦湾	
	松江-松江大学城	宝山-大华	
	阳光花城	浦东-六里	
	浦东-康桥	卢湾-五里桥	
	宝山-罗店	松江-新桥	
	浦东-周浦	黄浦-蓬莱公园	
	松江-泗泾	金榜人家	
	松江-九亭	闵行-江川路	
	嘉定-江桥新城	嘉定-丰庄	
	嘉定-嘉定新城	闸北-大宁绿地	
		徐汇-康健	

- 3. 地址：无法有效离散化，删除
- 4. 户型信息：
取 q = 室数+厅数，

$q < 4$ 取 1
 $4 \leq q < 6$ 取 2
 $q \geq 6$ 取 3
其他（联排别墅，独栋别墅，叠加别墅）取 4

5. 面积：
取 $q = \text{室数} + \text{厅数}$ ，
面积 < 90 取 1
 $90 \leq q < 125$ 取 2
 $125 \leq q < 200$ 取 3
 $q \geq 200$ 取 4

6. 房源类型：
新房 取 1
二手房 取 2

7. 特殊：
普通 取 1
地区优势 取 2
精装修、学区房 取 3

8. 价格（元/m²）：
价格 < 10000 取 1
 $10000 \leq q < 20000$ 取 2
 $20000 \leq q < 60000$ 取 3
 $q \geq 60000$ 取 4

属性约简结果：

返回：11101

可知在该离散化方案下，房源类型对价格的影响较小

属性值约简结果：

地块	户型	面积	特殊	价格
-	-	1	3	3

-	1	-	3	3
4	-	1	-	3
4	1	-	-	3
2	2	1	-	3
2	-	2	-	3
2	2	3	1	4
2	-	-	3	3
-	-	-	2	3
3	3	-	3	4
-	3	3	1	3
3	-	-	1	3
2	1	-	1	2
1	-	3	-	3
-	4	-	-	3
1	-	2	1	2
1	2	-	1	2
3	-	1	-	3
3	1	-	-	3
1	-	1	-	2
2	-	4	-	3
-	2	3	3	3
1	-	-	3	3
3	-	2	-	3
3	2	-	-	3
4	-	3	3	3
4	3	-	3	3
4	3	3	-	3
4	3	-	1	4
4	-	4	-	4
2	3	-	-	3
4	-	2	1	3
4	-	3	1	4
4	-	2	3	4

测试集测试结果：

正确率：21/30 约为 70%