决策树；

采用10 fold cross validation进行分类结果的验证。

由于感兴趣的是最终是否订阅，而购买（36548）远少于未购买（4641），我们分别用灵敏-有效性和精度-召回率两种方式比较。

剪枝；

108f2fe60

结果正确：

age：连续值；

job：离散值；

month显然和day重叠，而且day几乎没有意义，去掉

duration：连续值，打电话时间

分析unknown的数量和比例

计划：

连续值：age，duration，campaign，pdays，previous，emp.var.rate，cons.price.idx，，cons.conf.idx，，euribor3m，nr.employed

始终维持在这里：

right:586;wrong:604

贝叶斯分类：

**"poutcome"**:True,  
**"emp.var.rate"**:False,  
**"cons.price.idx"**:False,  
**"cons.conf.idx"**:False,  
**"euribor3m"**:False,  
**"nr.employed"**:False,

去掉这些属性，正确率居然提升了。

poutcome

"no" 32144, 395, 3423,

"yes" 2939, 606, 493,

default

"no" 27816, 3, 8143,

"yes" 3605, 1, 433,

day\_of\_week

"no" 7454, 7501, 7039, 6877, 7091,

"yes" 894, 753, 839, 741, 811,

marital

"no" 4077, 9676, 22142, 67,

"yes" 409, 1390, 2227, 12,

job

"no" 848, 8580, 938, 1323, 532, 8886, 3605, 1213, 5930, 2565, 289, 1253,

"yes" 122, 606, 92, 117, 215, 1161, 292, 345, 623, 292, 31, 142,

contact

"no" 21879, 14083,

"yes" 3301, 737,

month

"no" 4755, 12883, 2093, 270, 5410, 292, 93, 6390, 161, 3615,

"yes" 554, 886, 539, 276, 535, 222, 89, 473, 106, 358,

loan

"no" 29607, 5485, 870,

"yes" 3359, 594, 85,

education

"no" 14, 10273, 8348, 3689, 1437, 2096, 5536, 4569,

"yes" 4, 1443, 900, 369, 209, 177, 438, 498,

housing

"no" 16351, 18741, 870,

"yes" 1764, 2189, 85,