



## Bagging 算法:

我们先将 60% 的数据划为训练集，可以想象成一个装数据的大袋子。

然后我们从大袋子里有放回的随机取样本选形成一个个小袋子， $D_1, D_2, D_3, D_4$ ，用这些小袋子里的数据来训练模型。

假设有  $m$  个小袋子，每个小袋子中球的个数是  $n'$ ，大袋子里的球数是  $n$ ， $n'$  一般小于  $n \cdot 60\%$

每个小袋子中的数据都是从大袋子中有放回的随机取  $n'$  个数据形成的。