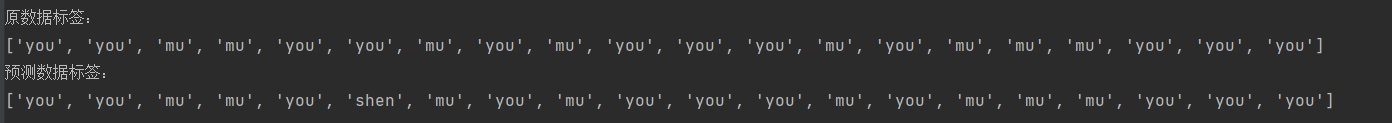
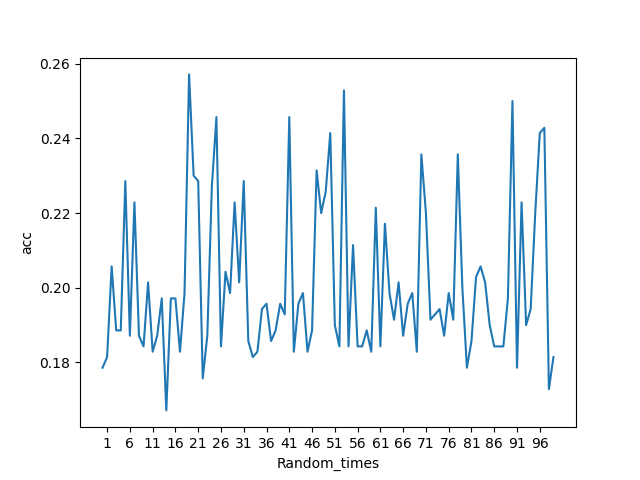
**结果分析**

1. KNN算法的执行时间与哪些因素有关？

KNN在数据处理阶段不参与实质性的模型训练，但需要跟每一个样本做计算，所以这一步运算时间与数据量的多少有很大关系。在我们的数据集测试下，KNN算法最耗时的步骤依然是模型训练部分，这一部分需要测试不同K值下数据的拟合效果，寻找最合适的K值以用于汉字识别。

1. KNN算法中K值的选取对识别结果的影响

选择较小的K值，相当于用较小范围的训练实例进行预测，“学习”近似误差会减小，只有与输入实例较近的训练实例才对预测结果起作用，但是估计误差会增大，容易发生过拟合；选择较大的K值，相当于用较大范围的训练实例进行预测，优点是可以减少学习的估计误差，但是近似误差会增大。这时，与输入实例较远的训练实例也会对预测起作用，使预测发生错误，且K值增大就意味着模型变得简单。我们使用Python对不同K值进行了测试，自动筛选出合适的K值供使用。

1. K-means算法的执行时间与哪些因素有关？

在我们的测试下，K-means算法最耗时间的部分是在寻找更高的拟合精度这一步骤，这一步需要很多次循环遍历来进行，导致程序运行变慢。

1. K-means算法结果精度的分析

相较KNN算法来说，我们测试的K-means算法汉字识别得出的结果准确率较低，经过不断的测试，我们初步推断原因出在以下几点：

1. K-means算法对于数据数量和分布的要求较高，而KNN算法更适用于我们测试用的数据集。
2. K-means函数的参数输入有问题，需要调整参数。

这个问题的解决方法还需留待日后发现。