简单来说网格搜索是自动化寻找最佳参数的方法。它本身没有任何神奇之处，只是把你的所有参数的可能都试一遍，每一个参数都打个分（打分方法由你来确定），返回一个分数最好的。

交叉验证是在不碰测试集的情况下，只用训练集和模型本身来检验模型表现的一种方法。恰巧这两者可以封装在一起使用，增加了学生的理解难度。不过你有哪个细节不明白，可以继续跟帖。

回答一下你的问题：

> 网格搜索是为了确定最优参数的一种全部遍历方式，这个参数是指的是什么？

大多数算法都是有几个参数可以设置的，当你不知道哪个参数最好的时候，网格搜索可以帮你验证一下。例

如 http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html#sklearn-tree-decisiontreeregressor16

max_depth : int or None, optional (default=None)

就是一个参数，默认是 None，你可以传入一个整数。

> 比如说 project 1 是指的我们表格中用哪些 features 的组合去评估最好吗？

你是指泰坦尼克哪个项目吗？选择 features 一般来说不是网格搜索做的事情。

> 还有就是用交叉验证去测试这些参数的不同组合哪些最准确，但是一定要使用交叉验证去衡量吗？

网格搜索可以不用交叉验证，交叉验证相当于把每个参数组合，都试 k 次，然后取平均分作为这个参数组合的最终得分。

http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

里面 cv_results_ 可以看到详情。

有其它方法吗？

交叉验证有很多种方法，K-fold 只是其中的一种，更多

见 https://en.wikipedia.org/wiki/Cross-validation_(statistics)[12]

你理解的是对的。GridSearchCV 帮我们解决的问题就是，在不碰测试集的前提下，不碰测试集的前提下，不碰测试集的前提下，寻找模型的最佳参数！为了找到这个最佳参数，我们就需要知道模型的表现，为了知道模型表现，我就要打分，为了打分，就要把训练集重新分割为训练集（9/10)和验证集(1/10）。我们遍历了参数，每一个可能的参数组合都在 K 份验证集上跑一遍分，取平均分最高的作为最优参数。

总结：GridSearchCV 没什么神奇的地方，就是干了体力活，只是一次性干了很多（抽象层次比较高），所以一上来不是那么容易理解。其实你把 `cv_results_` 返回的那个 dictionary 变成 csv 一看，就知道它是怎么工作的了！