2015-2016 课程作业安排

1. 作业提交方式：

* 利用目前较为流行的版本管理体系Github：作业在网络学堂布置，学生提交代码到Github；
* 制定简单的R和Python语言基本代码规范，使其养成按照规范编写代码的习惯；
* 代码需满足我们布置时所必需的接口，以方便我们进行代码检测；
* 作业涉及抄袭，一旦确认，双方期末成绩最多给通过；
* 建议利用vim/Emacs 进行代码编写；
* 建议利用Shell中常见的命令进行文本处理；
* 建议利用latex进行文档编写。

1. 习题课互动形式：
2. 考虑不分组。
3. 第三，四，五次习题课，给大家30分钟时间，找几个同学，进行文献阅读报告及作业展示。
4. 作业内容安排：3 次平时作业 + 1次课程结束的报告。
   1. 第一节课的预作业(不算在3次平时作业内)

建立Github 帐号并熟悉Github 平台；

熟悉R或Python语言，并了解基本代码规范。

* 1. 第一次作业 (第二次习题课结束后进行布置) ： **序列比对**

1. 给定几种DNA序列（是否含有外显子，内含子）以及密码子列表，将其翻译为相应的蛋白质序列；

2. 给定两段较长的DNA序列，编写S-W比对算法，进行序列比对，并给出比对分数的显著性；

3. 给定某一蛋白质，利用blastp，完成多物种间同一蛋白质的序列比对，从而构建物种进化树；

4. 文献阅读报告(指定几篇相关文献，选择一篇就行阅读，并完成阅读总结，建议利用英文)。

* 1. 第二次作业 （第三次习题课结束后，进行布置）： **聚类分析**

1. 编写主成分分析的程序，在给定的一组病人中，利用其mRNA表达数据，甲基化数据，miRNA数据，进行聚类分析。

2. 在上述聚类中，尝试不同的聚类策略，不同的Kernel-PCA降维方法，对结果的影响。

3. 给出聚类图谱 (利用ggplot等绘图)。

4. 文献阅读报告（同第一次作业要求，但更换文献主题）。

* 1. 第三次作业 (在第四次习题课结束后，进行布置) ： 网络分析

1. 提供药物与蛋白质结合的数据，药物相似性（几种相似性水平）数据，蛋白质相似性数据（几种相似性水平），使其构建类似DrugCIPHER中的网络。

2. 在上述网络中，利用随机游走，DrugCIPHER，最大流等方法，进行网络分析，预测药物与蛋白质的结合可能。

3. 文献阅读报告（同第一次作业要求，但更换文献主题）。

* 1. 课程报告 （第五次习题课结束后，进行布置）

1. 利用不同的特征选择方法，及回归方法，构建病人phenotye与mRNA 表达，miRNA数据的关系，了解是否能从基因组层次的数据预测指定表型。
2. 文献阅读报告 （同第一次作业要求，但更换文献主题）。
3. 课程总结与建议报告。