

本次作业需要用到数据文件：E3E5\_data/

- 训练集：train\_data\_E3E5\_2genes.txt  
train\_data\_E3E5\_10genes.txt
- 测试集：test\_data\_E3E5\_2genes.txt  
test\_data\_E3E5\_10genes.txt

## 作业 2018-10-22. 贝叶斯分类器与线性分类器的实验

- 背景：同作业 20180929。  
训练集上十折交叉验证，  
用训练集训练，验证集验证  
用训练集训练，训练集验证
  - 1. 分别用 2 个基因的数据和 10 个基因的数据进行线性 SVM 分类实验和高斯核函数 SVM 分类实验，参照第一次作业的内容自行设计需要考查和分析的实验结果。（目的是学习 SVM 软件包的使用，体会用 SVM 及不同核函数进行分类实验的特点。）
  - 2. 分别用 2 个基因的数据和 10 个基因的数据进行多层感知器神经网络分类实验，参照第一次作业的内容和第一题自行设计需要考查和分析的实验结果。（目的是学习神经网络软件包的使用，体会 MLP 结构和训练参数对训练过程和结果的影响。部分同学可以练习自己编写多层感知器 BP 算法程序。）
- 交作业日期：2018 年 11 月 11 日（周日）前打包提交到网络学堂。
  - 如发现抄袭、杜撰或未经说明的引用，或发现捏造结果，本次作业将记-10 到-20 分。如出现雷同报告但无法区分谁是原作者，则都按抄袭论处。

### 作业要求：

#### 1、提交内容：

- a) 实验报告：PDF 文件，需适当排版，正文用 5 号字仿宋体单倍行距，不超过 3 页。正文无法包括的内容可以放在附件中，但正文必须保证在不阅读附件的情况下的完整性。作业成绩主要依据正文。
- b) 附件：因篇幅限制无法放入正文但又需要介绍和讨论的细节内容可以用附件 PDF 文件提交，包括实验设计中的细节和所用程序的出处及参数设置等等，需提交实验结果的数据文件，如使用非开源程序则需提供程序源代码。附件所提供材料需保证实验内容能够完全重复。

#### 2、关于编程和讨论：

鼓励自己写程序（用任何语言），也允许使用工具包，不禁止使用他人程序。在实验报告及程序报告中须明确写明程序出处和作者，所提供细节必须足以完全重复实验结果。

欢迎同学间就作业相关内容进行讨论，但实验和报告必须独立完成，并在实验报告中对所参与的讨论进行说明（参加人及讨论内容）。

如发现抄袭、杜撰或未经说明的引用，或发现捏造结果，本次作业将记-10 到-20 分。如出现雷同报告但无法区分谁是原作者，则都按抄袭论处。