# 2018/9/17

* 一. Python学习
  + python 中numpy 包用法，具体学习dataframe部分
* 一. 论文阅读
  + 基于支持向量机的赖氨酸翻译后修饰位点预测方法研究
    - 特征提取与特征选择：
      * k-spaced氨基酸对组成（Composition of K.spaced Amino Acid Pairs，CKSAAP）
      * 即计算给定肽段的k-spaced氨基酸对出现的频率，有效地反映位点周围的序列模式信息
      * 即用20个氨基酸和一个虚拟的氨基酸X构成样本序列
      * 2-spaced氨基酸对共有441种可能：AxxA,AxxC,AxxD,……,XxxY,XxxX
      * 编码为441维向量，每一维表示为出现的次数
      * 取k=0，1，2，3，4，将这些k-spaced氨基酸对特征结合起来组成一个2205维的特征向量
      * 因为CKSAAP编码会产生高纬的特征向量，可以使用F-score 剔除高位特征的冗余特征
* F−Score=(1+β2)⋅Precision⋅Recall / β2⋅Precision+Recall
* 三. 明日计划：
  + 尝试完成CKSAAP编码代码
  + 尝试完成进行F-score打分矩阵