

## prerequisite

以一元三次函数建立的估算模型方程预测效果最好 (K)

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_m X_m$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

$$SST = \sum (y_i - \bar{y})^2$$

$$r^2 = \frac{SSR}{SST}$$

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2$$

NDVI越大，植被越健康

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

四个通道的波段范围:

- Blue: 400 - 600 nm, peak at 450 nm
- Green: 480 - 700 nm, peak at 550 nm
- Red: 580 - 720 nm, peak at 640 nm
- Near-infrared: 680 - 1000 nm

tiff文件可以用gdal库处理，dat+hdr文件可以用spectral库处理

## 2.Opencv提供的功能

- 图像数据操作（内存分配与释放，图像复制、设定和转换）
  - 图像/视频的输入输出（支持文件或摄像头的输入，图像/视频文件的输出）
  - 矩阵/向量数据操作及线性代数运算（矩阵乘积、矩阵方程求解、特征值、奇异值分解）
  - 支持多种动态数据结构（链表、队列、数据集、树、图）
  - 基本图像处理（去噪、边缘检测、角点检测、采样与插值、色彩变换、形态学处理、直方图、图像金字塔结构）
  - 结构分析（连通域/分支、轮廓处理、距离转换、图像矩、模板匹配、霍夫变换、多项式逼近、曲线拟合、椭圆拟合、狄劳尼三角化）
  - 摄像头定标（寻找和跟踪定标模式、参数定标、基本矩阵估计、单应矩阵估计、立体视觉匹配）
  - 运动分析（光流、动作分割、目标跟踪）
  - 目标识别（特征方法、HMM模型）
  - 基本的GUI（显示图像/视频、键盘/鼠标操作、滑动条）
  - 图像标注（直线、曲线、多边形、文本标注）
- 单变量建模：一元线性模型、幂函数模型、指数模型、乘幂模型、多项式模型等；
  - 多变量建模：多元逐步回归、主成分分析回归、神经网络回归、支持向量机回归、偏最小二乘回归、因子分析回归、多元函数回归等。

### 模型检验：

- 决定系数 $R^2$
- 均方根误差RMSE
- 相对分析误差RPD: 为预测的标准差/RMSE，一般认为：RPD>2时，模型极好；1.4<RPD<2时，模型还行，RPD<1.4时，模型凉凉

偏最小二乘回归：

多对多线性回归模型，当两组变量的个数很多且存在多重相关性时，而观测数据的数量有很小时。

偏最小二乘回归分析在建模过程中集中了主成分分析，典型相关分析和线性回归分析方法的特点，因此在分析结果中，除了可以提供一个更为合理的回归模型外，还可以同时完成一些类似于主成分分析和典型相关分析的研究内容，提供更丰富、深入的一些信息。

```
adfGeoTransform[0] /* top left x */
adfGeoTransform[1] /* w-e pixel resolution */
adfGeoTransform[2] /* 0 */
adfGeoTransform[3] /* top left y */
adfGeoTransform[4] /* 0 */
adfGeoTransform[5] /* n-s pixel resolution (negative value) */
```

gdal对栅格数据的处理是按照波段进行的，提供了元数据、块大小、颜色表和其他一些信息，这些信息都是按照波段提供的

rasterColorTable

**GPI\_GRAY:** c1表示灰度值

**GPI\_RGB:** c1表示红，c2表示绿，c3表示蓝，c4表示Alpha

**GPI\_CYMP:** c1表示青，c2表示品，c3表示黄，c4表示黑

**GPI\_HLS:** c1表示色调，c2表示亮度，c3表示饱和度

1. 开发和模型搭建工作在vscode（远程连接）上。
2. envi查看遥感数据和常规信息。
3. XFTP用于本地和服务器数据交互。（做训练的数据下载之后统一放到50服务器上个人文件夹下）。

## 大体思路

1. 研究一下遥感图像的组成，包括各个波段以及一些比较重要的信息  
了解ENVI和IDL  
对样本的氮磷钾三种元素和波段做相关性分析（PLSR）  
对波段可以进行组合（波段比值、波段差值等）  
提取遥感bands信息，最好输出成csv文件
2. 搭建模型，先初步做一个实验  
PLSR+BPNN（先做一个相关性分析，再选取特征波段进行分析）
3. 预测并形成结果  
预测结果输出成csv文件  
在原始遥感图像上修改对应的bands值，映射到rgb空间作为最终结果

## 数据记录

```
geo at 0,0 : (122.91558201079, 48.024294415147274)
geo at last : (123.29772533265994, 47.74509802483892)

bands count : 12
band1`s value at 0,0 -3.402823e+38
band1Data`s type <class 'numpy.ndarray'>
band1`s xsize : 2127 band1`s ysize : 1554
band1`s max value : 0.28195 band1`s min value : -3.402823e+38

geo at 0,0 : (48.024294415147274, 122.91558201079)
geo at last : (47.74509802483892, 123.29772533265994)
```

## 进展梳理

1. 标签数据（氮磷钾有机质）预处理完成。
2. 遥感图像波段值提取完成（含NDVI计算以及像素坐标与经纬度坐标的转换）。
3. 常规模型也搭建完毕，相关性分析一并做了。
4. 等吃完晚饭最新数据上传完毕后，再重新生成标签数据和波段数据，并计算以下NDVI。
5. 相关性问题解决。
6. 模型可以往GMM、提升等方向去考虑。

## 高斯过程

高斯过程对一组可以拟合的函数每个赋予一个概率值，这个概率分布符合高斯分布，概率的均值代表了这批数据最有可能的表征。

概率的方法可以将对预测的置信度结合到回归的结果中。

多元高斯函数由均值向量和协方差矩阵决定。协方差矩阵总是对称的并且是半正定。（正定矩阵一定是对称的）

高斯对应着分布，过程对应着随机过程，高斯过程定义在连续域上的无限多个高维随机变量所组成的随机过程。

高斯过程样本与一般机器学习样本的区别在于，高斯过程中样本各特征之间存在相关关系，这种相关关系是通过协方差矩阵体现的。各个维度特征之间的协方差矩阵可以通过高斯过程核来模拟，对所有随机变量均进行核函数计算便可得到协方差矩阵，有了协方差矩阵后便可以对高斯过程进行采样。

推导：

首先我们假设：

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} K(X, X) & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right). \quad (1)$$

根据多维高斯分布的条件分布性质：

$$\begin{aligned} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} &\sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^\top & B \end{bmatrix} \right) \\ \Rightarrow \mathbf{x}|\mathbf{y} &\sim \mathcal{N}(\boldsymbol{\mu}_x + CB^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), A - CB^{-1}C^\top), \quad (2) \end{aligned}$$

有

$$\begin{aligned} \mathbf{y}^*|\mathbf{y} &\sim \mathcal{N}(K(X^*, X)K(X, X)^{-1}\mathbf{y}, \\ &K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*)). \quad (3) \end{aligned}$$

由高斯分布的性质可得  $\mathbf{y}^* = K(X^*, X)K(X, X)^{-1}\mathbf{y}$  时  $p(\mathbf{y}^*|\mathbf{y})$  最大。推导完毕。

## 反演结果

### 有机质反演

选取波段3, 4进行有机质反演：

```
plsRmse : 0.29626238755110096
best params of the plsModelRes {'n_components': 1}
linearResRmse : 0.2744690448075993
svrRmse : 0.2352725659270332
```

选取波段3进行有机质反演：

```
plsRmse : 0.2929722672622635
best params of the plsModelRes {'n_components': 1}
linearResRmse : 0.29371653817720306
svrRmse : 0.3135775303553364
```

GPR:

波段3:

```
GPR RMSE: 0.2793689040874952
```

波段3, 4:

GPR RMSE: 0.22786105856658317