# Linux相关

## **2019.10.26**

### **1.Ubuntu设置字体大小**

|  |
| --- |
| (1)安装gnome-tweaks桌面配置工具，sudo apt  install    gnome-tweaks  (2)运行gnome-tweaks  (3)在字体中设置合适放大（缩小）倍数 |

### **2.Ubuntu解决WPS光标不跟随问题**

|  |
| --- |
| (1)安装qt4-qtconfig,sudo apt-get install qt4-qtconfig  (2)安装ibus-qt,sudo apt-get install ibus-qt4  (3)运行qt4-qtconfig,在q t4-qtconfig “界面”选项卡默认输入法选项中看到ibus，并选定，然后设置XIM输入风格为光标跟随风格。 |

### Ubuntu双系统下挂载Windows分区

|  |
| --- |
| 1.查看磁盘信息  (1)设备文件名称 file name  (2)挂载目录 mount point ，创建挂载目录 sudo mkdir +路径  (3)文件类型 type  (4)挂载参数 options，默认defaults  (5)是否备份 dump 0或1  (6)开机自检 pass：一般设置为2。如果不想自检，那就设置为0，挂载点为根目录的设备，设置为1，其它需要自检的设备，设置为2。  2.修改fstab文件  打开/etc.fstab文件，将磁盘或分区的文件名称和挂载信息填写：  sudo gedit /etc/fstab  例如：/dev/sdb2 /media/csl/文件 ntfs defaults 0 0  重启即可挂载 |

### Ubuntu18.04开机出现检测到系统程序出现问题

|  |
| --- |
| 打开终端，输入命令：  sudo gedit /etc/default/apport  将文件中的enabled = 1 修改为0，然后保存 |

### notepadqq不小心将工具栏删除后，无法找回

|  |
| --- |
| 解决方案：到/home/.config/Notepadqq/文件夹下，删除所有文件，然后重启notepadqq，个人习惯设置需要重新设置。 |

## 2019.10.30

### 1.Ubuntu声音太小

|  |
| --- |
| 执行：sudo usermod -a -G audio username |

# Python相关

## 2019.10.16

### **1. array 与 matrix 的区别**

|  |
| --- |
| (1)np.array( )更具通用性，np.mat()只适用于二维矩阵，维数超过2以后，np.mat()就不适用了。  (2)mat()函数与array()函数生成矩阵所需的数据格式有区别：mat()函数中数据可以为字符串以分号(；)分割，或者为列表形式以逗号（，）分割。而array()函数中数据只能为逗号。  (3) mat()函数与array()函数生成的矩阵计算方式不同  (4)利用np.mat()将数组转换为矩阵 |

### **2.np.multiply()、np.dot()和星号（\*）三种乘法运算的区别**

|  |
| --- |
| (1)np.multiply()  函数作用：数组和矩阵对应位置相乘，输出与相乘数组/矩阵的大小一致。  (2)np.dot()  函数作用：对于秩为1的数组，执行对应位置相乘，然后再相加；对于秩不为1的二维数组，执行矩阵乘法运算；超过二维的可以参考numpy库介绍  (3)星号（\*）乘法运算  作用：对数组执行对应位置相乘；对矩阵执行矩阵乘法运算 |

## **2019.10.18**

### **1.求矩阵的逆**

|  |
| --- |
| 示例代码如下：  import numpy as np  a = np.array([[1,1,1],[1,0,-1],[1,-1,0]])  array([[ 1, 1, 1],  [ 1, 0, -1],  [ 1, -1, 0]])  print(np.linalg.inv(a))  [[ 0.33333333 0.33333333 0.33333333]  [ 0.33333333 0.33333333 -0.66666667]  [ 0.33333333 -0.66666667 0.33333333]]  A = mat(a)  A.I |

## 2019.10.18

# 吴恩达机器学习笔记

## 线性回归

1.1 代价函数

|  |
| --- |
| 线性回归的代价函数： |

### 1.2特征缩放

|  |
| --- |
| 特征缩放(feature scaling)大致的思路是：梯度下降算法中，在有多个特征的情况下，如果你能确保这些不同的特征都处在一个相近的范围，这样梯度下降法就能更快地收敛。  执行特征缩放时，我们通常的目的是将特征的取值约束到−1到+1的范围内。  最简单的方法是令 ， 是平均值， 是标准差。  注意：特征缩放其实并不需要太精确，其目的只是为了让梯度下降能够运行得更快一点，让梯度下降收敛所需的循环次数更少一些而已。 |

## 2. 逻辑回归（分类问题）

### **2.1 逻辑回归模型**

|  |
| --- |
| 逻辑回归模型：  其中：X为特征向量  g 代表逻辑函数是一个常用的逻辑常用函数为S形函数，公式为：<    的作用是，对于给定的输入变量，根据选择的参数计算输出变量输出等于1的可能性，即 |

### **2.2 逻辑回归代价函数**

|  |
| --- |
| 逻辑回归的代价函数为：    简化如下：    带入代价函数得到：    即：    将代入得 |

### **2.3 求代价最小参数**

|  |
| --- |
| 得到代价函数后，可以采用梯度下降算法来求得使代价函数最小的参数。  Repeat{    (simultaneously updata all  )  }  将 求偏导后得到：  Repeat{    (simultaneously updata all  )  } |

### **2.4 特征缩放**

|  |
| --- |
| 特征缩放可以提高梯度下降的收敛速度，这个特征缩放也适用于逻辑回归。如果特征范围差距很大的话，应用特征缩放的方法，可使梯度下降收敛更快。 |

## 3. 正则化

### **3.1 过拟合问题**

|  |
| --- |
| 欠拟合：不能很好地适应训练集；  过拟合：过于强调拟合原始数据，而丢失了算法的本质：预测新数据。  解决过拟合问题的两种办法：   1. 丢弃一些不能帮助我们正确预测的特征。 2. 正则化(Regularization)。保留所有的特征，但是减少参数的大小(magnitude)。 |

### **3.2 正则化代价函数**

|  |
| --- |
| (1)线性回归：    求最小化参数：  Repeat until convergence{      }  (2)正规方程    矩阵尺寸(n+1)\*(n+1)  (3)逻辑回归方程    求最小化参数：  Repeat until convergence{      } |

## 4. 神经网络

### **4.1 神经网络模型**

|  |
| --- |
|  |

### **4.2 代价函数**

|  |
| --- |
| m ：代表样本容量总数  k ：代表K分类  l ：代表神经网络的总层数  sl ：代表神经网络l层的激活单元总数 |

# Git相关

## 添加SSH密钥

|  |
| --- |
| (1)创建密钥  ssh-keygen -t rsa -C “邮箱”  (2)在GitHub上添加密钥  (3)测试是否成功  ssh -T git@github.com  (4)ubuntu查看id\_rsa.pub文件  cat .ssh/id\_rsa.pub |

## 设置自己的名字和邮箱

|  |
| --- |
| $ git config --global user.name "Your Name"  $ git config --global user.email "email@example.com" |

## 在本地创建版本库

|  |
| --- |
| 1. cd /e/LocalGithub/   (2)变为Git管理的仓库 git init |

## 添加远程库与本地库连接

|  |
| --- |
| git remote add origin 自己的版本库链接 |

## 传输文件

|  |
| --- |
| (1)添加到本地仓库  $ git add .  (2)添加提交描述  $ git commit -m '内容'  (3)把本地仓库代码提交  $ git push -u origin master |

## 问题汇总

|  |
| --- |
| 1.合并Github和本地代码仓库命令  $ git pull --rebase origin master |

# Matplotlib

## 基本语法

|  |
| --- |
| #基本用法  代码如下：  import matplotlib.pyplot as plt  import numpy as np  x = np.linspace(-1,1,50)  y = 2\*x+1  plt.plot(x,y)  plt.show() |

## Figure图像

|  |
| --- |
| #figuer 使用  代码如下：  x = np.linspace(-1,1,50)  y1 = 2\*x+1  y2 = x\*x+1  plt.figure(num=1)  plt.plot(x,y1)  plt.figure(num=2,figsize=(8,5))  plt.plot(x,y1,color = 'red',linewidth = 1.0,linestyle='--')  plt.plot(x,y2)  plt.show() |

## 设置坐标1

|  |
| --- |
| # 设置坐标轴1  代码如下：  x = np.linspace(-3,3,50)  y1 = 2\*x+1  y2 = x\*x+1  plt.figure(num=2,figsize=(8,5))  plt.plot(x,y1,color = 'red',linewidth = 1.0,linestyle='--')  plt.plot(x,y2)  plt.xlim((-1,2))  plt.ylim((-2,3))  plt.xlabel("X")  plt.ylabel("Y")  new\_ticks=np.linspace(-1,2,5)  print(new\_ticks)  plt.xticks(new\_ticks)  plt.yticks([-2,-1.8,-1,1.22,3],  [r'$really\ bad\ \alpha$',r'$bad$',r'$normal$',r'$good$',  r'$really\ good$'])  plt.show() |

## 设置坐标2

|  |
| --- |
| #设置坐标轴2  代码如下：  x = np.linspace(-3,3,50)  y1 = 2\*x  y2 = x\*x  plt.figure(num=2,figsize=(8,5))  plt.plot(x,y1,color = 'red',linewidth = 1.0,linestyle='--')  plt.plot(x,y2)  plt.xlim((-1,2))  plt.ylim((-2,3))  plt.xlabel("X")  plt.ylabel("Y")  new\_ticks=np.linspace(-1,2,5)  print(new\_ticks)  plt.xticks(new\_ticks)  plt.yticks([-2,-1.8,-1,1.22,3],  [r'$really\ bad\ \alpha$',r'$bad$',r'$normal$',r'$good$',  r'$really\ good$'])  #gca = 'get current axis'  ax = plt.gca()  ax.spines['right'].set\_color('none')  ax.spines['top'].set\_color('none')  ax.xaxis.set\_ticks\_position('bottom')  ax.yaxis.set\_ticks\_position('left')  ax.spines['bottom'].set\_position(('data',0))#outward,axes  ax.spines['left'].set\_position(('data',0))  plt.show() |

## Legend图例

|  |
| --- |
| #Legend 图例  代码如下：  x = np.linspace(-3,3,50)  y1 = 2\*x  y2 = x\*x  plt.figure(num=2,figsize=(8,5))  l1, = plt.plot(x,y1,color = 'red',linewidth = 1.0,linestyle='--',label='down')  l2, = plt.plot(x,y2,label='up')  plt.legend(handles=[l1,l2,],labels=['aaa','bbb'],loc='upper right')  plt.xlim((-1,2))  plt.ylim((-2,3))  plt.xlabel("X")  plt.ylabel("Y")  new\_ticks=np.linspace(-1,2,5)  print(new\_ticks)  plt.xticks(new\_ticks)  plt.yticks([-2,-1.8,-1,1.22,3],  [r'$really\ bad\ \alpha$',r'$bad$',r'$normal$',r'$good$',  r'$really\ good$'])  #gca = 'get current axis'  ax = plt.gca()  ax.spines['right'].set\_color('none')  ax.spines['top'].set\_color('none')  ax.xaxis.set\_ticks\_position('bottom')  ax.yaxis.set\_ticks\_position('left')  ax.spines['bottom'].set\_position(('data',0))#outward,axes  ax.spines['left'].set\_position(('data',0))  plt.show() |

## Annotation标注

|  |
| --- |
| #scastter画点图  代码如下：  #Annotation标注  x = np.linspace(-3,3,50)  y = 2\*x+1  plt.figure(num=2,figsize=(8,5))  plt.plot(x,y)  #gca = 'get current axis'  ax = plt.gca()  ax.spines['right'].set\_color('none')  ax.spines['top'].set\_color('none')  ax.xaxis.set\_ticks\_position('bottom')  ax.yaxis.set\_ticks\_position('left')  ax.spines['bottom'].set\_position(('data',0))#outward,axes  ax.spines['left'].set\_position(('data',0))  x0 = 1  y0 = 2\*x0+1  plt.scatter(x0,y0,s=50,color = 'b')  plt.plot([x0,x0],[y0,0],'k--',linewidth=2.5)  #method 1  ###################  plt.annotate(r'$2x+1=%s$'%y0,xy=(x0,y0),xycoords='data',xytext=(+30,-30),textcoords='offset points', fontsize=16,  arrowprops=dict(arrowstyle='->',connectionstyle='arc3,rad=.2'))  #method 2  ###################  plt.text(-3.7,3,r'$This\ is\ the\ some\ text.\sigma\_i\ \alpha\_t$',  fontdict={'size':16,'color':'red'})  plt.show() |

## Tick能见度

|  |
| --- |
| #代码如下  #Annotation标注  x = np.linspace(-3,3,50)  y = 0.1\*x  plt.figure(num=2,figsize=(8,5))  plt.plot(x,y,linewidth=30)  #gca = 'get current axis'  ax = plt.gca()  ax.spines['right'].set\_color('none')  ax.spines['top'].set\_color('none')  ax.xaxis.set\_ticks\_position('bottom')  ax.yaxis.set\_ticks\_position('left')  ax.spines['bottom'].set\_position(('data',0))#outward,axes  ax.spines['left'].set\_position(('data',0))  for label in ax.get\_xticklabels()+ax.get\_yticklabels():  label.set\_fontsize(15)  label.set\_bbox(dict(facecolor='white',edgecolor='red',alpha=0.9))  plt.show() |

## scatter散点图

|  |
| --- |
| 代码如下：  #scatter散点图  import numpy as np  import matplotlib.pyplot as plt  n = 1024  X = np.random.normal(0,1,n)  Y = np.random.normal(0,1,n)  T = np.arctan2(Y,X)#for color value  plt.scatter(X,Y,s=100,c=T,alpha=0.5)  plt.xlim((-1.5,1.5))  plt.ylim((-1.5,1.5))  plt.xticks(())  plt.yticks(())  plt.show() |

## Bar柱状图

|  |
| --- |
| 代码如下：  #Bar柱状图  import numpy as np  import matplotlib.pyplot as plt  n = 12  X = np.arange(n)  Y1 = (1-X/float(n))\*np.random.uniform(0.5,1.0,n)  Y2 = (1-X/float(n))\*np.random.uniform(0.5,1.0,n)  plt.bar(X,+Y1,facecolor='#9999ff',edgecolor='white')  plt.bar(X,-Y2)  for x,y in zip(X,Y1):  #ha:horizontal alignment  plt.text(x+0.04,y+0.05,'%.2f'%y,ha='center',va='bottom')  for x,y in zip(X,-Y2):  #ha:horizontal alignment  plt.text(x+0.04,y-0.15,'%.2f'%y,ha='center',va='top')  plt.xlim((-.5,n))  plt.ylim((-1.25,1.25))  plt.xticks(())  plt.yticks(())  plt.show() |

## 10.contours等高线图

|  |
| --- |
| 代码如下：  #等高线图  import matplotlib.pyplot as plt  import numpy as np  def f(x,y):  #the height function  return (1-x/2+x\*\*5+y\*\*3)\*np.exp(-x\*\*2-y\*\*2)  n = 256  x = np.linspace(-3,3,n)  y = np.linspace(-3,3,n)  X,Y = np.meshgrid(x,y)  #use plt.contour to filling contours  #X,Y and value for (x,y)point  plt.contourf(X,Y,f(X,Y),8,alpha = 0.75,cmap=plt.cm.hot)  #use plt.contour to add contour lines  C=plt.contour(X,Y,f(X,Y),8,colors='black',linewidths=0.5)  #adding label  plt.clabel(C,inline=True,fontsize=10)  plt.xticks(())  plt.yticks(())  plt.show() |

## 11.image图片

|  |
| --- |
| 代码如下：  #image图  import numpy as np  import matplotlib.pyplot as plt  a = np.array([2,160,150,255,56,89  ,12,33,109]).reshape(3,3)  plt.imshow(a,interpolation='nearest',cmap='bone',origin='upper')  plt.colorbar(shrink=0.8)#示例缩放比  plt.xticks(())  plt.yticks(())  plt.show() |

## 12.3D数据

|  |
| --- |
| 代码如下：  import numpy as np  import matplotlib.pyplot as plt  from mpl\_toolkits.mplot3d import Axes3D  fg = plt.figure()  ax = Axes3D(fg)  X = np.arange(-4,4,0.25)  Y = np.arange(-4,4,0.25)  X,Y = np.meshgrid(X,Y)  R = np.sqrt(X\*\*2+Y\*\*2)  #height value  Z = np.sin(R)  ax.plot\_surface(X,Y,Z,rstride=1,cstride = 1,cmap='rainbow')  ax.contourf(X,Y,Z,zdir='z',offset=-1,cmap='rainbow') |

## 13.subplot多合一显示图

|  |
| --- |
| 代码如下：  #subplot多合一显示  import matplotlib.pyplot as plt  plt.subplot(2,1,1)  plt.plot([0,1],[0,1])  plt.subplot(2,3,4)  plt.plot([0,1],[0,1])  plt.subplot(2,3,5)  plt.plot([0,1],[0,1])  plt.subplot(2,3,6)  plt.plot([0,1],[0,1])  plt.show() |

## 14.subplot分格显示

|  |
| --- |
| 代码如下：  #subplot 分格显示  import matplotlib.pyplot as plt  import matplotlib.gridspec as gridspec  '''  #nethod1:subplot2grid  #######################  plt.figure()  ax1 = plt.subplot2grid((3,3),(0,0),colspan=3,rowspan=1)  ax1.plot([1,2],[1,2])  ax1.set\_title('ax1\_title')  ax2 = plt.subplot2grid((3,3),(1,0),colspan=2,rowspan=1)  ax2.plot([1,2],[1,2])  ax3 = plt.subplot2grid((3,3),(1,2),colspan=1,rowspan=2)  ax3.plot([1,2],[1,2])  ax4 = plt.subplot2grid((3,3),(2,0),colspan=1,rowspan=1)  ax4.plot([1,2],[1,2])  ax5 = plt.subplot2grid((3,3),(2,1),colspan=1,rowspan=1)  ax5.plot([1,2],[1,2])  #nethod2:gridspec  #######################  plt.figure()  gs = gridspec.GridSpec(3,3)  ax1 = plt.subplot(gs[0,:])  ax2 = plt.subplot(gs[1,:2])  ax3 = plt.subplot(gs[1:,2])  ax4 = plt.subplot(gs[-1,0])  ax4 = plt.subplot(gs[-1,1])  '''  #nethod3:easy to define structure  #######################  f,((ax11,ax12),(ax21,ax22))=plt.subplots(2,2,sharex=True,sharey=True)  plt.tight\_layout()  plt.show() |

## 15.图中图

|  |
| --- |
| 代码如下：  #图中图  import matplotlib.pyplot as plt  fig = plt.figure()  x = np.arange(1,7)  y = np.arange(1,7)  left,bottom,width,height = 0.1,0.1,0.8,0.8  ax1=fig.add\_axes([left,bottom,width,height])  ax1.plot(x,y,'r')  left,bottom,width,height = 0.2,0.6,0.25,0.25  ax2=fig.add\_axes([left,bottom,width,height])  ax2.plot(x,y,'b')  plt.axes([0.6,0.2,0.25,0.25])  plt.plot(x,y,'g')  plt.show() |

## 16.次坐标轴

|  |
| --- |
| 代码如下：  #次坐标轴  import matplotlib.pyplot as plt  import numpy as np  x = np.arange(0,10,0.1)  y1 = 0.005\*x\*\*2  y2 = -1\*y1  fig,ax1= plt.subplots()  ax2 = ax1.twinx()  ax1.plot(x,y1,'g')  ax2.plot(x,y2,'r')  ax1.set\_xlabel('X data')  ax1.set\_ylabel('Y1',color='g')  ax2.set\_ylabel('Y2',color='r')  plt.show() |

## 17.Animation动画

|  |
| --- |
| 代码如下：  #Animation动画  import numpy as np  import matplotlib.pyplot as plt  from matplotlib import animation  fig,ax = plt.subplots()  x= np.arange(0,2\*np.pi,0.01)  line,=ax.plot(x,np.sin(x))  def animate(i):  line.set\_ydata(np.sin(x+i/10))  return line,  def init():  line.set\_ydata(np.sin(x))  return line,  ani=animation.FuncAnimation(fig=fig,func=animate,frames=100,init\_func=init,interval=20,blit=True)  plt.show() |

# Numpy & Pandas

## numpy & pandas有什么用

|  |
| --- |
| 应用于数据分析，如tensorflow等  计算速度快  Pandas是numpy的升级版 |

## numpy&pandas安装

|  |
| --- |
| (1)Anaconda集成  (2)官网下载 |

## numpy属性

|  |
| --- |
| #numpy 属性  import numpy as np  array = np.array([[1,2,3],  [2,3,4]])  print(array)  print('number of dim:',array.ndim)#纬度  print('shape',array.shape)  print('size',array.size) |

## numpy的创建array

|  |
| --- |
| #numpy的创建array  import numpy as np  a = np.array([[2,3,4],  [4,5,6]],  dtype=np.int)  print(a)  b = np.zeros((3,3),dtype=np.int)  print(b)  c = np.ones((3,3),dtype=np.int)  print(c)  d = np.eye(3)  print(d)  e = np.arange(10,22,2).reshape(2,3)  print(e)  f = np.linspace(1,10,12).reshape((3,4))  print(f) |

## numpy的基础运算

|  |
| --- |
| #numpy 基本运算  import numpy as np  a = np.array([10,20,30,40])  b = np.arange(4)  print(a,b)  c = a+b  print(c)  d = b\*\*2  print(d)  e = np.sin(a)  print(e)  print(b<3)  f = np.array([[1,1],[0,1]])  g = np.arange(4).reshape(2,2)  h = f\*g  i = np.dot(f,g)  print(h)  print(i)  j = np.random.random((2,4))  print(j)  print(np.sum(j))  print(np.sum(j,axis=0))  print(np.sum(j,axis=1)) |

## numpy的基础运算2

|  |
| --- |
| import numpy as np  A = np.arange(14,2,-1).reshape(3,4)  print(A)  print(np.argmin(A))#最小值索引  print(np.argmax(A))#最大值索引  print(np.mean(A))#求平均值  print(np.mean(A,axis=0))#对列进行计算  print(np.mean(A,axis=1))#对行进行计算  print(np.median(A))#中位数  print(np.cumsum(A))#逐项累加  print(np.diff(A))#累差  print(np.nonzero(A))#非零数的位置  print(np.sort(A))#逐行排序  print(np.transpose(A))#求A的转置  print(A.T)#求A的转置  print(np.clip(A,5,9))#小于5等于5，大于9等于9，中间不变 |

## numpy的索引

|  |
| --- |
| #numpy的索引  import numpy as np  A = np.arange(3,15).reshape((3,4))  print(A)  print(A[2])#索引第二行  print(A[2,:])#索引第二行  print(A[:,1])#索引第二列  print(A[1][1])#精确位置索引  print(A[1,1])#精确位置索引  print(A[:,0:2])#索引第一列和第二列 |

## numpy的array合并

|  |
| --- |
| #numpy的array合并  import numpy as np  A = np.array([1,1,1])  B = np.array([2,2,2])  C = np.vstack((A,B))#vertical stack  print(C)  print(A.shape,C.shape)  D = np.hstack((A,B))#horizontal stack  print(D)  print(D.shape)  print(A[np.newaxis,:])  print(A[:,np.newaxis])  A = A[:,np.newaxis]  B = B[:,np.newaxis]  print(np.hstack((A,B)))  print(np.concatenate((A,B,B,A),axis=0))  print(np.concatenate((A,B,B,A),axis=1)) |

## numpy的array分割

|  |
| --- |
| #numpy的array分割  import numpy as np  A = np.arange(12).reshape(3,4)  print(A)  #横向分割  print(np.split(A,2,axis= 1))  #纵向分割  print(np.split(A,3,axis = 0))  print(np.vsplit(A,3))  print(np.hsplit(A,2)) |

## 10.numpy的copy & deep copy

|  |
| --- |
| #numpy的copy和deep copy  import numpy as np  #copy同一个地址  a = np.arange(4)  b = a  c = a  d = a  a[0] = 11  print(a)  print(b)  print(c)  print(d)  print(a is b,a is c,a is d)#判断a,b是否相同  #深拷贝  b = a.copy()#deep copy  a[3] = 112  print(a)  print(b)  print(b is a) |

## 11.pandas基本介绍

|  |
| --- |
| #pandas基本介绍  '''  numpy 和 pandas的区别  numpy：序列化的排列  pandas：字典型的numpy  '''  import numpy as np  import pandas as pd  s = pd.Series([1,3,6,np.nan,44,1])  print(s)  dates = pd.date\_range('20191103',periods=6)  print(dates)  df = pd.DataFrame(np.random.randn(6,4),index = dates,columns=['a','b','c','d'])  print(df)  df1 = pd.DataFrame(np.random.randn(6,4))  print(df1)  df2 = pd.DataFrame()  print(df.dtypes)  print(df.index)  print(df.columns)  print(df.values)  print(df.describe())  print(df.T)  print(df.sort\_index(axis=1,ascending=False))  print(df.sort\_index(axis=0,ascending=False))  print(df.sort\_values(by='b')) |

## 12.pandas选择数据

|  |
| --- |
| #pandas选择数据  import numpy as np  import pandas as pd  dates = pd.date\_range('20191103',periods=6)  print(dates)  df = pd.DataFrame(np.arange(24).reshape((6,4)),index = dates,columns=['A','B','C','D'])  #print(df['A'],df.A)  #print(df[0:3],df['2019-11-03': '2019-11-06'])  #select by label:loc  #print(df.loc['20191104'])  #print(df.loc[:,['A','B']])  #print(df.loc['20191105',['A','B']])  #select by position:iloc  print(df.iloc[3,:])  #mixed selection:ix  #print(df.ix[:3,['A','C']])  #Boolean indexing  print(df[df.A > 8]) |

## 13.pandas设置值

|  |
| --- |
| #pandas设置值  import numpy as np  import pandas as pd  dates = pd.date\_range('20191103',periods=6)  df = pd.DataFrame(np.arange(24).reshape((6,4)),index = dates,columns=['A','B','C','D'])  df.iloc[2,2]=1111  print(df)  df.A[df.A>4]=0  print(df)  df['F'] = np.nan  df['E'] = pd.Series([1,2,3,4,5,6],index=pd.date\_range('20191103',periods=6))  print(df) |

## 14.pandas处理丢失数据

|  |
| --- |
| #pandas处理数据  import numpy as np  import pandas as pd  dates = pd.date\_range('20191103',periods=6)  df = pd.DataFrame(np.arange(24).reshape((6,4)),index = dates,columns=['A','B','C','D'])  df.iloc[0,1] = np.nan  df.iloc[1,2] = np.nan  print(df)  print(df.dropna(axis = 0,how='any'))#how = {'any','all'}  print(df.fillna(value=0))  print(df.isnull()) |

## 15.pandas导入导出

|  |
| --- |
| #pandas导入导出  import pandas as pd  data = pd.read\_csv('path',name=[])  data.to\_excel('path') |

## 16.pandas合并concat

|  |
| --- |
| #pandas合并concat  import pandas as pd  import numpy as np  '''  #concatenating  df1 = pd.DataFrame(np.ones((3,4))\*0,columns=['a','b','c','d'])  df2 = pd.DataFrame(np.ones((3,4))\*1,columns=['a','b','c','d'])  df3 = pd.DataFrame(np.ones((3,4))\*2,columns=['a','b','c','d'])  res = pd.concat([df1,df2,df3],axis=0,ignore\_index=True)  print(res)  res = pd.concat([df1,df2,df3],axis=1,ignore\_index=True)  print(res)  #join,[inner,outer]  df1 = pd.DataFrame(np.ones((3,4))\*0,columns=['a','b','c','d'],index=[1,2,3])  df2 = pd.DataFrame(np.ones((3,4))\*1,columns=['b','c','d','e'],index=[2,3,4])  res=pd.concat([df1,df2],join='inner',ignore\_index=True)  print(res)  #join\_axes  df1 = pd.DataFrame(np.ones((3,4))\*0,columns=['a','b','c','d'],index=[1,2,3])  df2 = pd.DataFrame(np.ones((3,4))\*1,columns=['b','c','d','e'],index=[2,3,4])  res=pd.concat([df1,df2],axis=1,join\_axes=[df1.index])  print(res)  '''  #append  df1 = pd.DataFrame(np.ones((3,4))\*0,columns=['a','b','c','d'])  df2 = pd.DataFrame(np.ones((3,4))\*1,columns=['b','c','d','e'])  df3 = pd.DataFrame(np.ones((3,4))\*1,columns=['b','c','d','e'],index=[2,3,4])  res = df1.append(df2,ignore\_index=True)  print(res)  res = df1.append([df2,df3])  print(res)  s1 = pd.Series([1,2,3,4],index=['a','b','c','d'])  res = df1.append(s1,ignore\_index=True)  print(res) |

## 17.pandas合并merge

|  |
| --- |
| #pandas合并merge  import pandas as pd  '''  left = pd.DataFrame({'key':['K0','K1','K2','K3'],  'A':['A0','A1','A2','A3',],  'B':['B0','B1','B2','B3',]})  right = pd.DataFrame({'key':['K0','K1','K2','K3'],  'C':['C0','C1','C2','C3',],  'D':['D0','D1','D2','D3',]})  print(left)  print(right)  res = pd.merge(left,right,on='key')  print(res)  '''  #two consider  left = pd.DataFrame({'key1':['K0','K0','K1','K2'],  'key2':['K0','K1','K0','K1'],  'A':['A0','A1','A2','A3',],  'B':['B0','B1','B2','B3',]})  right = pd.DataFrame({'key1':['K0','K1','K1','K2'],  'key2':['K0','K0','K0','K0'],  'C':['C0','C1','C2','C3',],  'D':['D0','D1','D2','D3',]})  print(left)  print(right)  res = pd.merge(left,right,on=['key1','key2'],how='inner')  print(res) |

## 18.pandas plot画图

|  |
| --- |
| #pandas plot画图  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  #plot data  #Series  data = pd.Series(np.random.randn(1000),index=np.arange(1000))  data = data.cumsum()  #DataFrame  data = pd.DataFrame(np.random.randn(1000,4),index=np.arange(1000),columns=list("ABCD"))  data = data.cumsum()  print(data)  data.plot()  plt.show() |