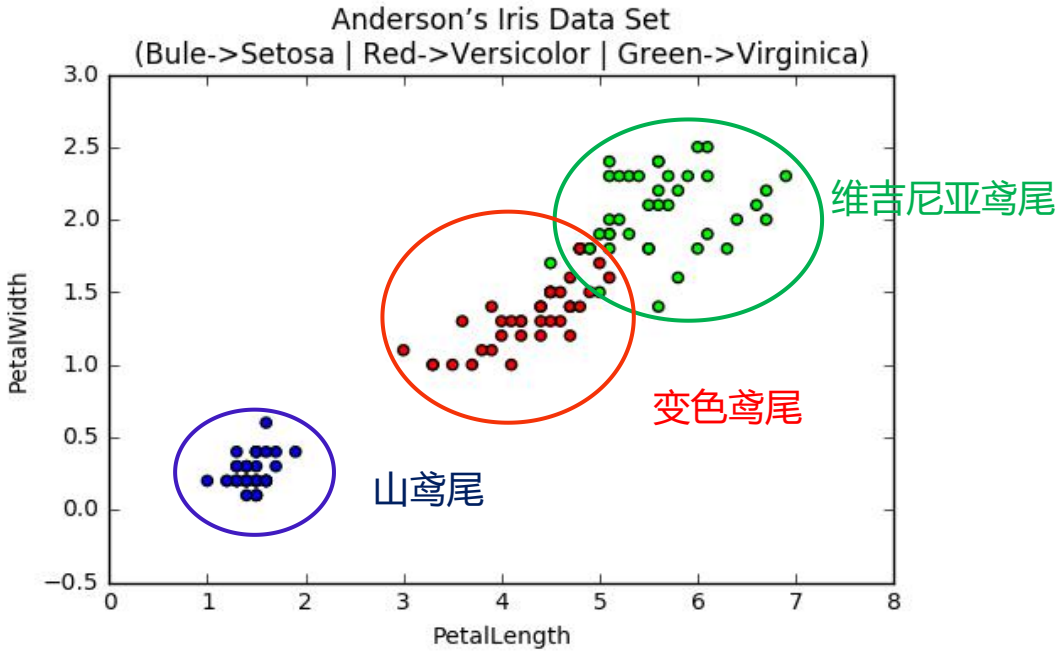
6.5.3 鸢尾花数据集可视化

鸢尾花数据散点图


Anderson's Iris Data Set
(Bule->Setosa | Red->Versicolor | Green->Virginica)
PetalWidth
PetalLength
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
7
8
维吉尼亚鸢尾
变色鸢尾
山鸢尾

- 花瓣长度

```
>>>iris[:,2]
array([5.6, 3.3, 4.5, 1.5, 1.7, 1.3, 1.5, 5.1, 4.4, 1.5, 3.9, 4.9, 1.2,
       1.7, 6.7, 4.7, 5.9, 6.6, 5.3, 1.5, 5.7, 5.6, 1.3, 5.6, 5.8, 1.5,
       4. , 5.1, 4.5, 5. , 4.4, 3. , 4.5, 5.5, 4.8, 5.7, 5.1, 5.1, 1.5,
       1.4, 6.4, 5.1, 5.2, 1.9, 1.6, 5. , 1.6, 6.9, 1. , 6. , 1.4, 4.4,
       4. , 1.2, 4.7, 4.8, 6.1, 5.1, 5.4, 3.5, 3.9, 5.6, 5. , 5.5, 4.5,
       6.3, 1.3, 6.1, 5.5, 1.5, 1.3, 4.6, 1.3, 6.1, 4.9, 1.5, 3.8, 4.2,
       4.5, 5.3, 1.5, 4.7, 4.6, 4.2, 5.6, 1.5, 4.8, 4.5, 5.1, 1.3, 5.2,
       4.7, 1.4, 1.5, 5.8, 1.4, 1.4, 6.7, 4.8, 1.6, 1.4, 3.3, 1.3, 4.1,
       1.6, 1.4, 1.5, 1.4, 3.6, 1.6, 4.9, 4.1, 1.6, 6. , 1.6, 4.4, 4.2,
       1.4, 1.4, 3.7])
```

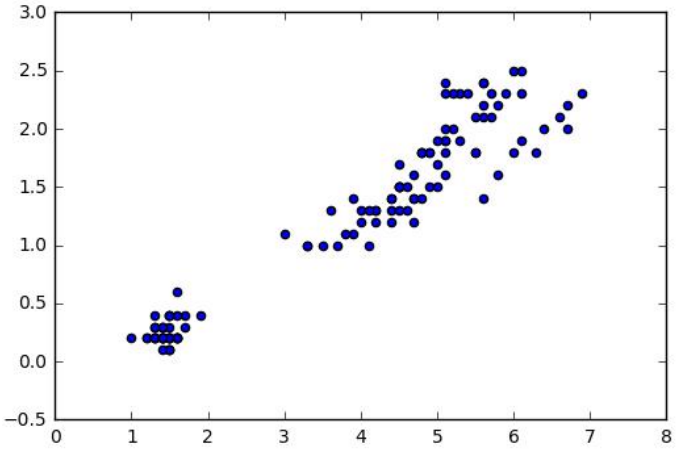
- 花瓣宽度

```
>>>iris[:,3]
array([2.2, 1. , 1.7, 0.1, 0.3, 0.2, 0.4, 2.3, 1.4, 0.4, 1.4, 1.5, 0.2,
       0.4, 2.2, 1.6, 2.3, 2.1, 2.3, 0.4, 2.1, 2.1, 0.4, 1.4, 1.6, 0.2,
       1.2, 1.8, 1.5, 1.7, 1.3, 1.1, 1.5, 2.1, 1.8, 2.3, 2. , 2.4, 0.3,
       0.3, 2. , 1.9, 2.3, 0.4, 0.2, 1.5, 0.2, 2.3, 0.2, 1.8, 0.2, 1.4,
       1.3, 0.2, 1.4, 1.8, 1.9, 1.9, 2.3, 1. , 1.1, 2.4, 1.9, 1.8, 1.5,
       1.8, 0.2, 2.5, 1.8, 0.2, 0.2, 1.3, 0.2, 2.3, 1.8, 0.1, 1.1, 1.3,
       1.5, 1.9, 0.2, 1.4, 1.5, 1.3, 2.4, 0.1, 1.4, 1.3, 1.6, 0.3, 2. ,
       1.2, 0.3, 0.2, 2.2, 0.3, 0.2, 2. , 1.8, 0.2, 0.2, 1. , 0.3, 1. ,
       0.4, 0.2, 0.2, 0.2, 1.3, 0.2, 1.8, 1.3, 0.2, 2.5, 0.6, 1.2, 1.2,
       0.2, 0.1, 1. ])
```

- 绘制散点图

```
plt.scatter(iris[:,2],iris[:,3])
plt.show()
```

运行结果：



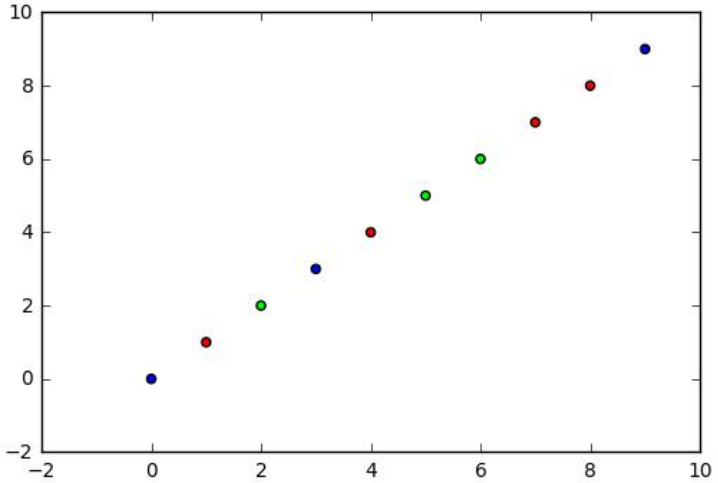
- **色彩映射**：将**参数c**指定为一个**列表或数组**，所绘制图形的颜色，可以随这个列表或数组中元素的值而**变换**，变换所对应的颜色由**参数cmap**中的颜色所提供。

colormap

```
plt.scatter(x, y, c, cmap )
```

```
x=np.arange(10)
y=np.arange(10)

dot_color=[0,1,2,0,1,2,2,1,1,0]

plt.scatter(x,y,c=dot_color,cmap='brg')
plt.show()
```

0 1 2

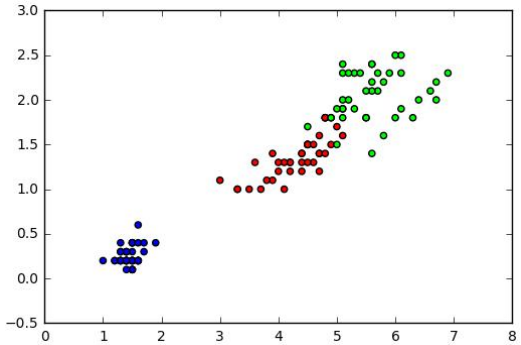


- 鸢尾花种类

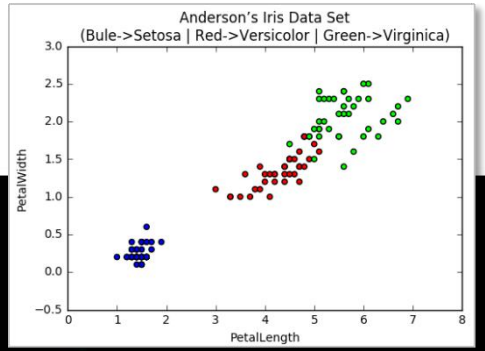
```
>>>iris_y=iris[:,4]
array([2., 1., 2., 0., 0., 0., 0., 2., 1., 0., 1., 1., 0., 0., 2., 1., 2.,
       2., 2., 0., 2., 2., 0., 2., 2., 0., 1., 2., 1., 1., 1., 1., 1., 2.,
       2., 2., 2., 2., 0., 0., 2., 2., 2., 0., 0., 2., 0., 2., 0., 2., 0.,
       1., 1., 0., 1., 2., 2., 2., 2., 1., 1., 2., 2., 2., 1., 2., 0., 2.,
       2., 0., 0., 1., 0., 2., 2., 0., 1., 1., 1., 2., 0., 1., 1., 1., 2.,
       0., 1., 1., 1., 0., 2., 1., 0., 0., 2., 0., 0., 2., 1., 0., 0., 1.,
       0., 1., 0., 0., 0., 0., 1., 0., 2., 1., 0., 2., 0., 1., 1., 0., 0.,
       1.])
```

```
plt.scatter(iris[:,2],iris[:,3],c=iris[:,4],cmap='brg')
plt.show()
```

运行结果：



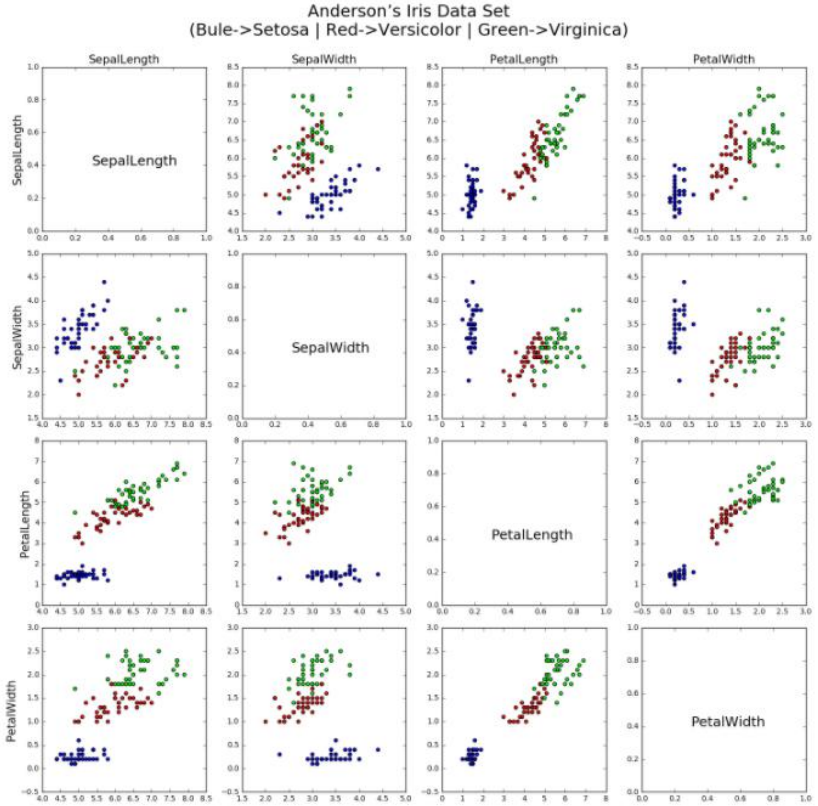
鸢尾花数据散点图 —— 2个属性



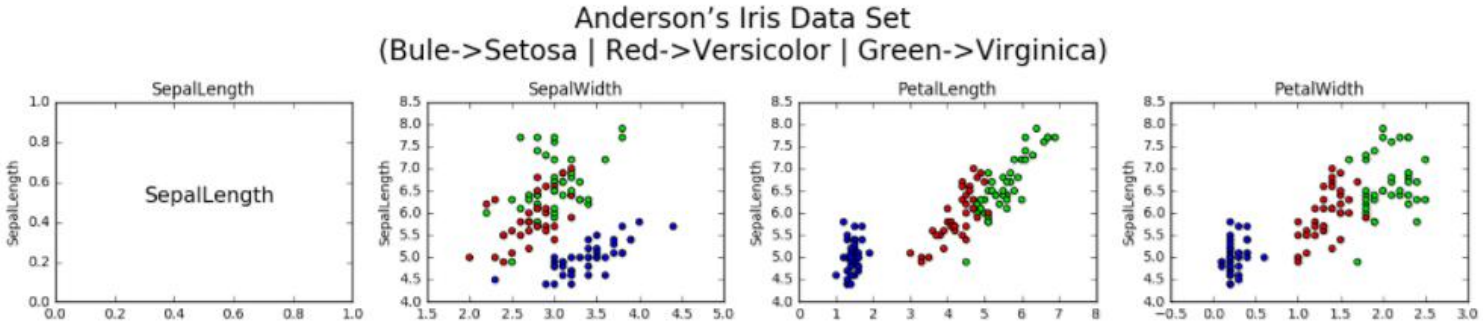

```
import tensorflow as tf
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"
train_path = tf.keras.utils.get_file(TRAIN_URL.split('/')[-1], TRAIN_URL)

COLUMN_NAMES = ['SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth', 'Species']
df_iris = pd.read_csv(train_path, names=COLUMN_NAMES, header=0)

iris=np.array(df_iris)

plt.scatter(iris[:,2],iris[:,3],c=iris[:,4],cmap='brg')
plt.title("Anderson's Iris Data Set\n(Bule->Setosa | Red->Versicolor | Green->Virginica)")
plt.xlabel(COLUMN_NAMES[2])
plt.ylabel(COLUMN_NAMES[3])
plt.show()
```

6.5.3 鸢尾花数据集可视化

	Sepal Length	Sepal Width	Petal Length	Petal Width
Sepal Length	X	O	O	O
Sepal Width	O	X	O	O
Petal Length	O	O	X	O
Petal Width	O	O	O	X



6.5.3 鸢尾花数据集可视化



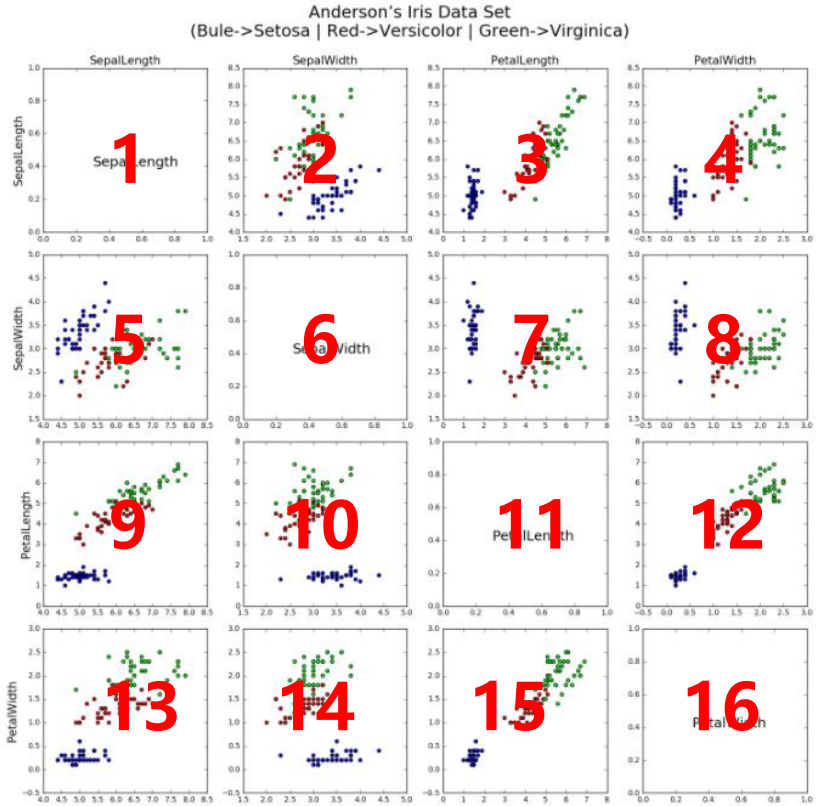

```
for i in range(4):
    plt.subplot(1, 4, i + 1)

    if(i==0):
        plt.text(0.3,0.5,COLUMN_NAMES[0],fontsize=15)
    else:
        plt.scatter(iris[:,i], iris[:,0], c=iris[:,4], cmap='brg')

    plt.title(COLUMN_NAMES[i])
    plt.ylabel(COLUMN_NAMES[0])
```

```
fig = plt.figure('Iris Data', figsize=(15, 3))

fig.suptitle("Anderson's Iris Data Set\n(Bule->Setosa | Red->Versicolor | Green->Virginica)")

for i in range(4):
    plt.subplot(1, 4, i + 1)
    if(i==0):
        plt.text(0.3,0.5,COLUMN_NAMES[0],fontsize=15)
    else:
        plt.scatter(iris[:,i], iris[:,0], c=iris[:,4], cmap='brg')

    plt.title(COLUMN_NAMES[i])
    plt.ylabel(COLUMN_NAMES[0])

plt.tight_layout(rect=[0,0,1,0.9])

plt.show()
```

- 鸢尾花数据集可视化

子图序号：

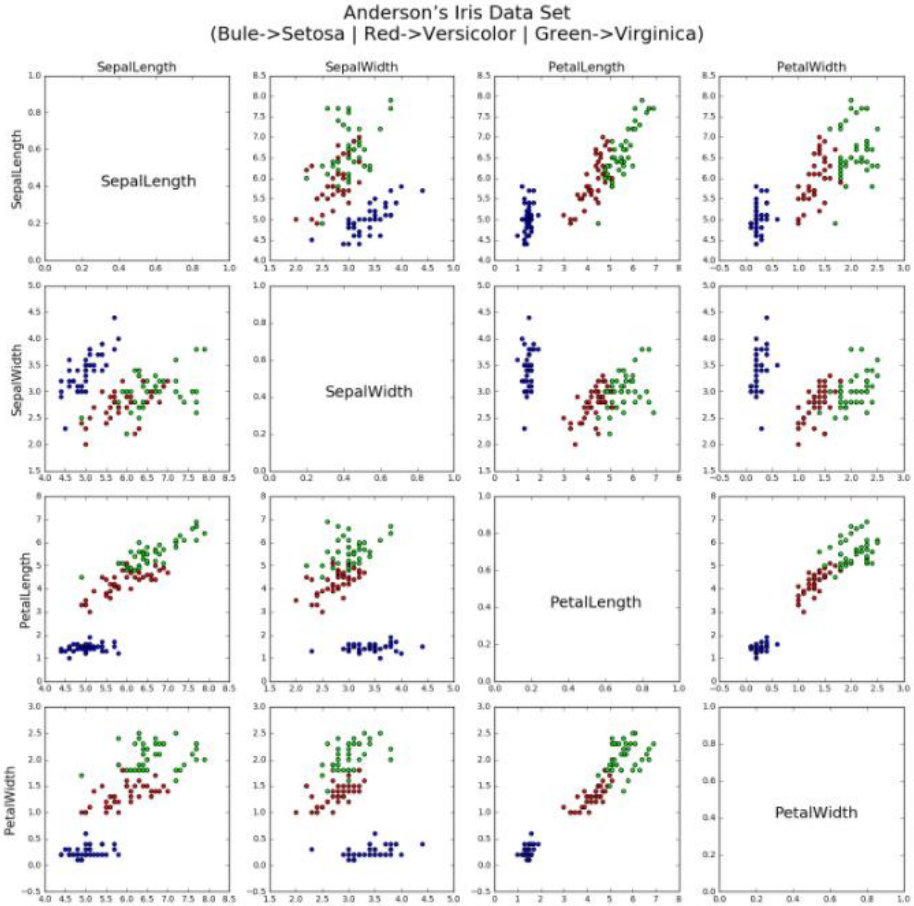
4×i + (j+1)



6.5.3 鸢尾花数据集可视化


```
import tensorflow as tf
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

TRAIN_URL = "http://download.tensorflow.org/data/iris_training.csv"
train_path = tf.keras.utils.get_file(TRAIN_URL.split('/')[-1], TRAIN_URL)

COLUMN_NAMES = ['SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth', 'Species']
df_iris = pd.read_csv(train_path, names=COLUMN_NAMES, header=0)
iris=np.array(df_iris)

fig = plt.figure('Iris Data', figsize=(15, 15))
fig.suptitle("Anderson's Iris Data Set\n(Bule->Setosa | Red->Versicolor | Green->Virginica)",fontsize=20)

for i in range(4):
    for j in range(4):
        plt.subplot(4, 4, 4*i + (j + 1))
        if(i==j):
            plt.text(0.3,0.4,COLUMN_NAMES[i],fontsize=15)
        else:
            plt.scatter(iris[:,j], iris[:,i], c=iris[:,4], cmap='brg')
        if(i == 0):
            plt.title(COLUMN_NAMES[j])
        if(j == 0):
            plt.ylabel(COLUMN_NAMES[i])

plt.show()
```

- 鸢尾花数据集可视化


Anderson's Iris Data Set
(Bule->Setosa | Red->Versicolor | Green->Virginica)
SepalLength
SepalWidth
PetalLength
PetalWidth