

# 先端データ解析論 (杉山将先生・本多淳也先生)

## 第2回レポート

ashiato45

2017年5月2日

### 宿題1

プログラムは付録に記す。標本の分割は、標本点を5つおきに飛ばすことで5グループ作成した。パラメタとして  $h = 0.2, \lambda = 0.3$  で実行し、図1の結果を得た。テスト誤差の平均は、0.668であった。

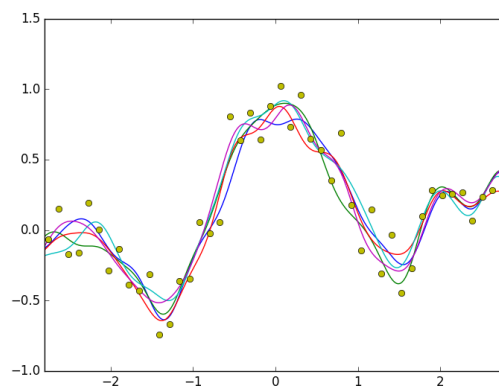


図1

### 付録

```
import sys
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(42)

h = float(sys.argv[1])
lam = float(sys.argv[2])

datanum = 50
datasamples = np.linspace(-3, 3, datanum)
datasamples = datasamples.reshape((len(datasamples), 1))
y = np.sin(np.pi*datasamples)/(np.pi*datasamples) + 0.1*datasamples
y += np.random.randn(datanum, 1)*0.2 # normal distribution
# print(y)

def gk(x, c):
    return np.exp(-(x-c)**2/(2*(h**2)))

totaler = 0
for i in range(5):
    group = []
    for j in range(50) if j % 5 != i]:
        yy = y[group, :]
        ds = datasamples[group, :]
        phi = np.fromfunction(lambda i, j: gk(ds[i, 0], ds[j, 0]), (len(ds), len(ds)), dtype=int)

        theta = np.linalg.solve(np.dot(phi.transpose(), phi) + lam*np.eye(len(ds), dtype=int), np.dot(phi.transpose(), yy))
        graphnum = 5000
        graphsamples = np.linspace(-3, 3, graphnum)
        graphsamples = graphsamples.reshape(graphnum, 1)
        m = np.fromfunction(lambda i, j: gk(graphsamples[i, 0], ds[j, 0]), (len(graphsamples), len(ds)), dtype=int)
        ans = np.dot(m, theta)
```

```

plt.plot(graphsamples, ans, '-')
em = np.fromfunction(lambda i,j: gk(ds[i, 0], ds[j, 0]), (len(ds), len(ds)), dtype=int)
eans = np.dot(em, theta)
er = (np.linalg.norm(eans - yy)**2)
print("error" + str(i) + ":" + str(er))
totaler += er

print(theta)
print(np.sum(abs(theta) < 1e-8))
print(totaler/5.0)

plt.axis([-2.8, 2.8, -1, 1.5])

plt.plot(datasamples, y, 'o')
# plt.plot(datasamples, eout, 'o')
plt.show()

```