

14.2 コミッティ

平成 28 年 9 月 11 日

概 要

PRML の「14.2 コミッティ」についての実装と考察

目 次

1	問題設定	2
2	アルゴリズム	2
3	コード	2
4	結果	3
5	まとめ	4

1 問題設定

コミッティを用いて回帰問題に対する予測性能を向上させる.

2 アルゴリズム

まず, ブーストラップデータ集合を作る.

—— ブーストラップ ——

1. N 個のデータ点 $X = x_1, \dots, x_N$ から N 回復元抽出することによって新たなデータ集合を作る.

このデータ集合を用いて, ブートストラップ集約を行う.

—— ブートストラップ集約 (バギング) ——

1. M 個のブーストラップデータ集合を作る.
2. それぞれのデータ集合に対して M 個の予測モデル $y_m(x)$ を作る.
3. M 個の予測モデル $y_m(x)$ の平均がコミッティの予測である.

3 コード

ブートストラップ集約のコード (committee.py).

```
L=10
S_N=np.zeros((L,M,M))
m_N=np.zeros((L,M))
for l in range(L):
    #ブートストラップ
    x_l,t_l=[],[]
    for n in range(N):
        num=rd.randint(N)
        x_l.append(x[num])
        t_l.append(t[num])

    for n in range(N):
        for m in range(M):
            P[n,m]=gauss_basis(x_l[n],m,mu[m],s)

    alpha,beta=0,1.0/(0.3)**2
    delta=0.5
    lam=eigvals(np.dot(P.T,P))

    while delta>10**-6:
        tmp_alpha,tmp_beta=alpha,beta
        S_N[l,:,:]=inv(alpha*I+beta*np.dot(P.T,P))
        m_N[l,:]=beta*np.dot(S_N[l,:,:],np.dot(P.T,t_l))
        new_lam=lam*beta
        gamma=0
        for m in range(M):
            gamma+=new_lam[m]/(alpha+new_lam[m])
        alpha=gamma/np.dot(m_N[l,:],m_N[l,:])
        tmp=0
        for n in range(N):
            tmp+=(t_l[n]-np.dot(m_N[l,:],P[n,:]))**2
```

```

        beta=(N-gamma)/tmp
        delta=abs(alpha-tmp_alpha)+abs(beta-tmp_beta)

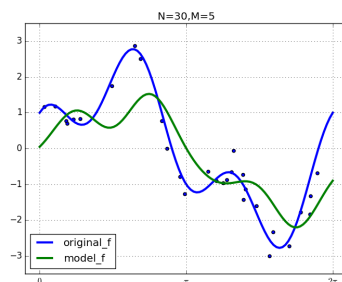
    def f(z,k):
        tmp=0
        for m in range(M):
            tmp+=m_N[k,m]*gauss_basis(z,m,mu[m],s)
        return tmp

def model_f(z):
    tmp=0
    for l in range(L):
        tmp+=f(z,l)
    return tmp/L

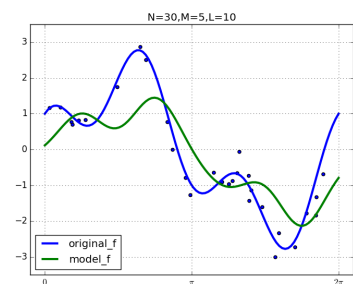
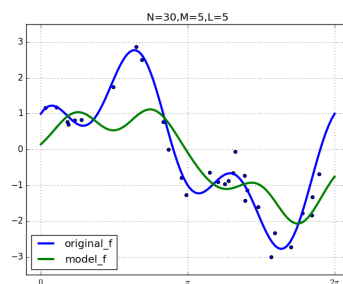
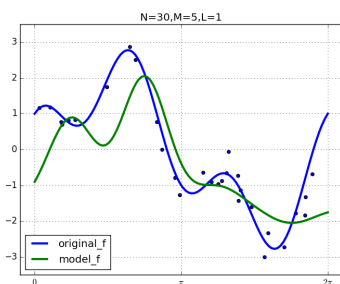
```

4 結果

ここでは例として、ガウス基底を用いた線形基底関数モデルで実験をした。
コミッティの影響が見えた例として、通常であれば



となったものが、コミッティを行うと $L = 1, 5, 10$



となった。

誤差は

M \ N	30	100	500
5	0.82	0.89	0.88
10	0.35	0.29	0.28
15	0.35	0.28	0.29

表 1: 通常の線形基底関数モデル

が

M \ N	30	100	500
5	0.84	0.89	0.87
10	0.36	0.30	0.27
15	0.34	0.31	0.27

表 2: コミッティ $L = 1$

M \ N	30	100	500
5	0.87	0.90	0.89
10	0.37	0.28	0.28
15	0.37	0.29	0.27

表 3: コミッティ $L = 5$

M \ N	30	100	500
5	0.82	0.89	0.88
10	0.35	0.29	0.28
15	0.35	0.28	0.27

表 4: コミッティ $L = 10$

となった.

5 まとめ

コミッティ(ブートストラップ集約)で変わるのは、データへの依存性の部分である. 同じモデルを用いて、違うデータに対して平均をとるというだけなので、モデルの改善にはつながらないと考ええる. ブートストラップデータ集合を用いるという点からしてうさんくさく、結果だけを取り繕うとした感じがする.

また、ブートストラップもあまり効果的であるとは考えられない. ブートストラップを使うなら、交差確認を行ったほうが良いように感じる.