

# 非定常環境下での教師付き学習： データの入力分布が変化する場合

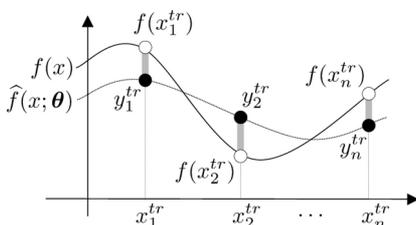
東京工業大学  
杉山 将

文字認識や顔認識などの認識問題は、データの背後に潜む生成規則を推定する教師付き学習の問題として定式化できる。教師付き学習では、識別機の学習に用いる訓練データと識別機の性能を評価するときに用いるテストデータが同じ規則に従うと仮定することが多い。しかし、実問題ではこの仮定が成り立たないことがある。例えば、顔認識システムでは背景や光の当たり方が訓練データとテストデータで異なることが多い。迷惑メール分類システムでは、分類システムだけでなく迷惑メール製作者も学習するため、迷惑メールデータの分布が変化する<sup>6)</sup>。脳波でコンピュータを操作するブレイン・コンピュータインターフェースでは、脳の非定常性のため脳波データの分布が変化する<sup>15)</sup>。その他、ロボット制御<sup>14)</sup>、バイオインフォマティクス<sup>3)</sup>、アンケートデータ処理<sup>8)</sup>など様々な分野の学習問題で分布の変化が見受けられる。

訓練データとテストデータが全く異なる規則に従って生成されれば、訓練データからテストデータに関する情報が全く抽出できない。従って、意味のある学習を行なうためには、訓練データとテストデータが何らかの共通点を持つ必要がある。本稿では、データの入力分布が変化する共変量シフト<sup>10)</sup>と呼ばれる状況を考え、その影響を緩和するための方法について解説する。

## 共変量シフト下での教師付き学習問題

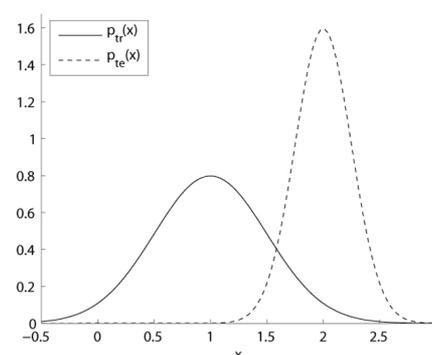
ある入出力関数  $f(x)$  を訓練データ  $\{(x_i^{tr}, y_i^{tr})\}_{i=1}^{n_r}$  から学習する問題を考えよう。但し、 $\{x_i^{tr}\}_{i=1}^{n_r}$  は入力点、 $\{y_i^{tr}\}_{i=1}^{n_r}$  は (一般に雑音を含む) 出力値、 $n_r$  は訓練データ数である。学習にはパラメータ  $\theta$  を持つモデル  $\hat{f}(x; \theta)$  を用い、 $\theta$  を訓練データから決定する (第1図



第1図 教師付き学習問題。訓練データ  $\{(x_i^{tr}, y_i^{tr})\}_{i=1}^{n_r}$  から関数  $f(x)$  を学習する。学習にはパラメータ  $\theta$  を持つモデル  $\hat{f}(x; \theta)$  を用いる。

参照) 学習したパラメータ  $\hat{\theta}$  を用いれば、訓練データに含まれないテスト入力  $x^{te}$  に対する出力  $y^{te}$  を  $\hat{f}(x^{te}; \hat{\theta})$  によって予測できる。

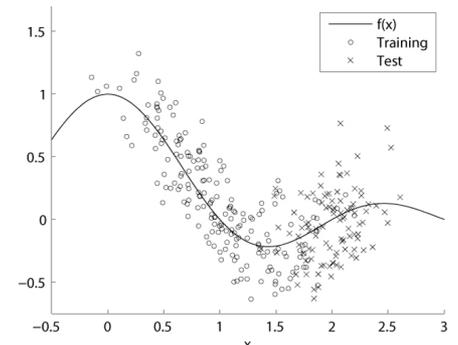
この問題をもう少し厳密に議論するために、次のような統計的な設定を考えることにしよう。訓練入力  $x_i^{tr}$  は密度



(a) 訓練入力密度、テスト入力密度

$p_{tr}(x)$  を持つ確率分布に従うと仮定し、訓練出力  $y_i^{tr}$  は条件付き密度  $\mu(y | x = x_i^{tr})$  を持つ条件付き確率分布に従うと仮定する。そして、 $\mu(y | x)$  の条件付き期待値を学習したい真の関数  $f(x)$  とみなすことにする。

一般的な教師付き学習の枠組みでは、



(b) 真の関数、訓練データ、テストデータ

第2図 共変量シフトの例。訓練データは主に図の左側の領域から、テストデータは主に図の右側の領域から採取される。これは、訓練入力データがほとんどない場所での出力を予測する外挿問題である。

テストデータ ( $x^{te}$ ,  $y^{te}$ ) は訓練データと同じ確率分布に従うと仮定する。しかし本稿では、テスト出力  $y^{te}$  は訓練出力と同じ条件付き確率分布に従うが、テスト入力  $x^{te}$  は密度  $p_{te}(x)$  を持つ別の確率分布に従う場合を考える。このような状況を共変量シフトと呼ぶ<sup>10)</sup>。共変量シフトの例を第2図に示す。これは、関数  $f(x) = \sin(x)$  を学習する問題であるが、訓練入力データがほとんどない場所での出力を予測しなければならない外挿問題であり、教師付き学習の難問の一つである。

教師付き学習問題の目標は、全てのテストデータに対する誤差の期待値  $G$  を最小にするようにパラメータ  $\theta$  を学習することである。即ち、

$$\min_{\theta} [G], \quad G = \mathbb{E}_{te} [(\hat{f}(x^{te}; \theta) - y^{te})^2]$$

但し、 $\mathbb{E}_{te}$  はテストデータ ( $x^{te}, y^{te}$ ) に関する期待値を表す。

### 重要度重み付き最小二乗法によるパラメータ学習

直線モデル  $\hat{f}(x) = \theta_0 + \theta_1 x$  を例に、共変量シフト下でのパラメータ学習法について説明することにする。

標準的なパラメータの学習法は、訓練誤差を最小にする最小二乗法 (least squares; LS) であろう。

$$\min_{\theta} \left[ \sum_{i=1}^{n_{tr}} (\hat{f}(x_i^{tr}; \theta) - y_i^{tr})^2 \right]$$

最小二乗法による学習結果の例を第3図(a)に示す。訓練データ('o')はうまく追従できているが、この学習結果を用いてもテストデータ('x')は正しく予測できない。

共変量シフト下では、重要度重み付き最小二乗法 (importance weighted LS; IWLS) を用いるのがよい<sup>10)</sup>。

$$\min_{\theta} \left[ \sum_{i=1}^{n_{tr}} w(x_i^{tr}) (\hat{f}(x_i^{tr}; \theta) - y_i^{tr})^2 \right]$$

$w(x)$  は重要度と呼ばれ、

$$w(x) = \frac{p_{te}(x)}{p_{tr}(x)}$$

で定義される。重要度重み付き最小二乗法では、テスト入力密度が高く訓練入力密度が低い訓練データを強く重み付けすることによって、共変量シフトの影響を吸収する。第2図の外挿問題の重要度  $w(x)$  を第8図に示す。このグラフより、第2図の左側にある訓練データの重要度は非常に小さく、右側に

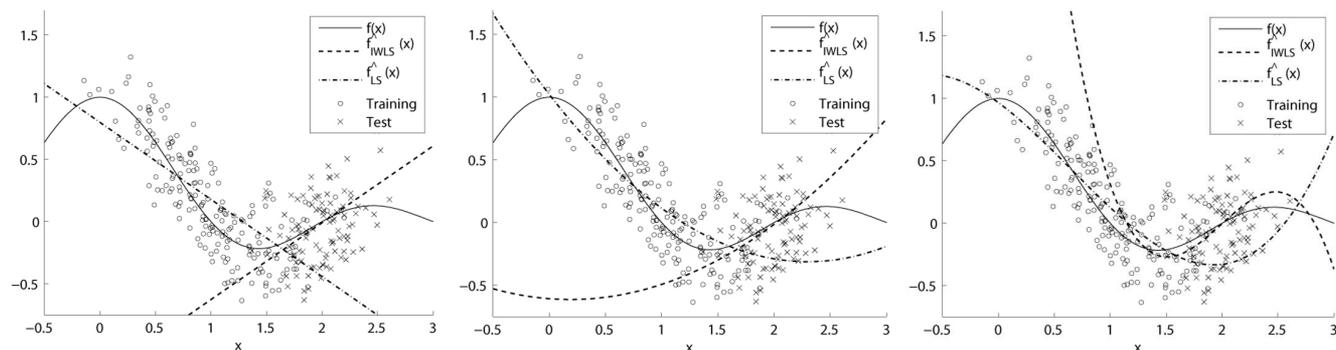
ある幾つかの訓練データだけが大きな重要度を持つことが分かる。つまり、訓練データの中でテストデータに似ていないものは排除され、テストデータに似ているものが自動的に選ばれる。

このときの重要度重み付き最小二乗法による学習結果の例を第3図(a)に示す。但し、重要度は第5節で紹介する方法で推定している。この結果より、共変量シフトが起こっても、重要度重み付き最小二乗法を用いることによってテストデータ('x')をかなり正確に予測できることが分かる。念のため、学習にはテストデータは用いておらず、訓練データ('o')しか用いていないことを明記しておく(但し、重要度の推定にはテスト入力データを用いている)。

線形判別分析やサポートベクトルマシンなど訓練誤差に基づくほとんど全ての学習法は、重要度で重み付けすることにより共変量シフト下での予測性能が向上する<sup>11) 9)</sup>。

### 重要度重み付き交差確認法によるモデル選択

教師付き学習ではモデルの選択が重要である<sup>2)</sup>。前節の外挿問題の例では



(a) 最大次数が1の多項式モデル (直線モデル)

(b) 最大次数が2の多項式モデル

(c) 最大次数が3の多項式モデル

第3図 第2図の外挿問題に対する通常の最小二乗法の学習結果  $\hat{f}_{LS}(x)$ 、及び、重要度重み付き最小二乗法の学習結果  $\hat{f}_{IWLS}(x)$ 。通常の最小二乗法の学習結果は訓練データを良く近似するが、テストデータの良い近似にはなっていない。一方、重要度重み付き最小二乗法の学習結果はテストデータを良く近似している。いずれの結果を求める際にもテストデータ('x')は用いず、訓練データ('o')しか用いていない(但し、重要度の推定にはテスト入力データを用いている)。

直線モデルを用いたが、直線に限る必然性はなく、例えば一般の多項式を用いてもよい。ここでは次式で表される多項式モデルを例にとり、その最大次数  $b$  を決定する方法を説明することにしよう。

$$\hat{f}_b(x; \theta) = \theta_0 + \sum_{\ell=1}^b \theta_\ell x^\ell$$

第2図の外挿問題の例に対して、最大次数  $b=1,2,3$  のモデルを用いて学習

した結果を第3図に示す。一般的に、通常の最小二乗法の学習結果は訓練データを良く近似するが、テストデータの良い近似にはなっていない。一方、重要度重み付き最小二乗法の学習結果はテストデータを良く近似している。

モデル選択の標準的な方法は、交差確認法 (cross validation; CV) であろう。しかし、交差確認法も共変量シフトの影響を受けるため、重要度を用いて補正した重要度重み付き交差確認法 (importance weighted CV; IWCV) を使う方がよい<sup>11)</sup>。これは、予測誤差  $G$  の推定値を与える方法であり、この値が小さいほど  $\hat{f}(x; \hat{\theta})$  は真の関数  $f(x)$  の良い推定である。重要度重み付き交差確認法を用いたモデル選択アルゴリズムを第4図に示す。

第3図に示した重要度重み付き最小二乗法の学習結果に対して、重要度重み付き交差確認法、及び、通常の交差確認法を用いて予測誤差を推定した結果を第5図に示す。このグラフより、真の予測誤差と重要度重み付き交差確認法による推定値にはほぼ同じ増減傾向があることが分かる。一方、通常の交差確認法による推定値の増減は真の予測誤差とは異なっている。

モデル選択を行った後に得られる最終的な学習結果の予測誤差の箱髴図を第6図に示す。この結果より、重要度重み付き最小二乗法と重要度重み付き交差確認法を組み合わせることによって、高い予測性能が獲得できることが分かる。

赤池の情報量規準などのモデル選択規準も、重要度で重み付けすることにより共変量シフト下での予測誤差推定精度が向上する<sup>10), 11)</sup>。

**手順1:**  
訓練データ  $\mathcal{T} = \{(x_i^{tr}, y_i^{tr})\}_{i=1}^{n_{tr}}$  を  $R$  個の部分集合  $\{\mathcal{T}^{(r)}\}_{r=1}^R$  に分割する。

**手順2:**  
 $\mathcal{T}^{(r)}$  以外の訓練データ  $\mathcal{T} \setminus \mathcal{T}^{(r)}$  を用いてモデル  $\hat{f}_b(x; \theta)$  を学習し、 $\hat{f}_b(x; \hat{\theta})$  を求める。

**手順3:**  
学習に用いていない訓練データ  $\mathcal{T}^{(r)}$  を用いて、 $\hat{f}_b(x; \hat{\theta}^{(r)})$  の予測誤差の  $\hat{G}_b^{(r)}$  を次式で求める。

$$\hat{G}_b^{(r)} = \frac{1}{|\mathcal{T}^{(r)}|} \sum_{(x,y) \in \mathcal{T}^{(r)}} w(x) (\hat{f}_b(x; \hat{\theta}^{(r)}) - y)^2$$

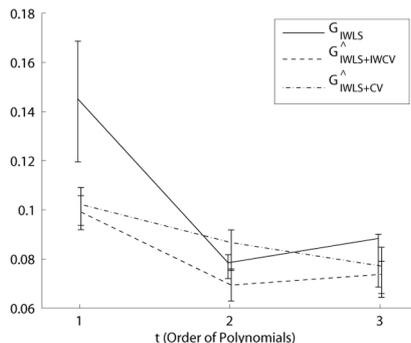
**手順4:**  
手順2から手順3を  $r=1,2,\dots,R$  に対して繰り返し、モデル  $\hat{f}_b(x; \theta)$  の予測誤差の推定値  $\hat{G}_b$  を次式で求める。

$$\hat{G}_b = \frac{1}{R} \sum_{r=1}^R \hat{G}_b^{(r)}$$

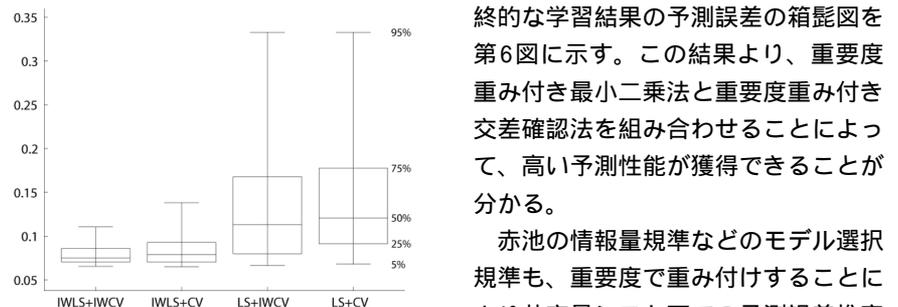
**手順5:**  
手順2から手順4を全てのモデル  $b$  に対して繰り返し、 $\hat{G}_b$  が一番小さいモデル  $\hat{b}$  を選ぶ。

**手順6:**  
選んだモデルを全ての訓練データ  $\mathcal{T}$  を用いて学習し、学習結果  $\hat{f}_{\hat{b}}(x; \hat{\theta})$  を求める。

第4図 重要度重み付き交差確認法を用いたモデル選択アルゴリズム。重要度重み付き交差確認法は、予測誤差  $G_b = \mathbb{E}_{(x,y) \in \mathcal{T}} [(f_b(x; \theta) - y)^2]$  の推定値を与える方法である。この値が小さいほど  $\hat{f}_b(x; \hat{\theta})$  は真の関数  $f(x)$  の良い推定である。通常の交差確認法では、手順3で重要度  $w(x)$  による重み付けを行わない。



第5図 第3図の重要度重み付き最小二乗法による学習結果の予測誤差  $G_{IWLS}$ 、それを重要度重み付き交差確認法で推定した結果  $\hat{G}_{IWLS+IWCV}$ 、及び、 $G_{IWLS}$  を通常の交差確認法で推定した結果  $\hat{G}_{IWLS+CV}$  の100回平均、上四分位点、下四分位点。  $\hat{G}_{IWLS+IWCV}$  の増減は  $G_{IWLS}$  とほぼ同傾向であるが、 $\hat{G}_{IWLS+CV}$  は異なる傾向を示している。



第6図 モデル選択とパラメータ学習を含めて、最終的に得られた学習結果の予測誤差 (100回の実験結果の箱髴図)。重要度重み付き最小二乗法と重要度重み付き交差確認法の組み合わせによって、高い予測性能が実現されている。

手順1:  $\alpha$ を適当に初期化する。  
 手順2:  $\alpha \leftarrow \alpha + \varepsilon A^\top (1/A\alpha)$   
 手順3:  $\alpha \leftarrow \alpha + (1-b^\top \alpha)b / (b^\top b)$   
 手順4:  $\alpha \leftarrow \max(0, \alpha)$   
 手順5:  $\alpha \leftarrow \alpha / (b^\top \alpha)$   
 手順6: 収束するまで手順2から手順5を繰り返し、 $\hat{\alpha} = \alpha$ を出力する。

第7図 KLIEPによる重要度推定のアルゴリズム。  $\varepsilon$ は小さい正の定数、 $A$ は $(j,i)$ 要素が $\exp(-(x_j^{te} - x_i^{tr})^2 / (2\sigma^2))$ の行列、 $\top$ は転置、 $\mathbf{1}$ は全ての要素が1のベクトル、 $\cdot / \cdot$ は要素ごとの除算、 $\mathbf{0}$ は全ての要素が0のベクトル、 $b$ は $i$ 番目の要素が $\sum_{r=1}^{n_{tr}} \exp(-(x_j^{tr} - x_i^{tr})^2 / (2\sigma^2))$ のベクトルである。また、ベクトルに対する $\max$ 演算は要素ごとの $\max$ 演算を表す。手順2は勾配上昇、手順3から手順5は拘束条件を満たすための射影にそれぞれ対応している。 $\alpha$ を十分に小さい値に設定すれば、上記のアルゴリズムは初期値の選び方によらず常に同じ解に収束する。

## KLIEPによる重要度の推定

これまでの議論から、共変量シフト下では重要度 $w(x) = p_{te}(x) / p_{tr}(x)$ をうまく用いることにより予測性能が向上することが分かった。しかし一般に重要度は未知であるため、何らかの方法で推定する必要がある。本節では、訓練データに加えて、密度 $p_{te}(x)$ を持つ確率分布に従うテスト入力データ $\{x_j^{te}\}_{j=1}^{n_{te}}$ が与えられる場合を考え、これらのデータから重要度を推定する方法について説明する。

訓練入力データ $\{x_i^{tr}\}_{i=1}^{n_{tr}}$ とテスト入力データ $\{x_j^{te}\}_{j=1}^{n_{te}}$ を用いて訓練入力密

度 $p_{tr}(x)$ とテスト入力密度 $p_{te}(x)$ を推定し、その比を取れば重要度を推定することができる。しかし、確率密度関数の推定は非常に難しい問題であり、この方法は精度が良くない。ここでは、確率密度関数を推定することなく直接重要度を推定できるKLIEP (Kullback-Leibler importance estimation procedure) と呼ばれる方法<sup>13)</sup>を紹介することにしよう。

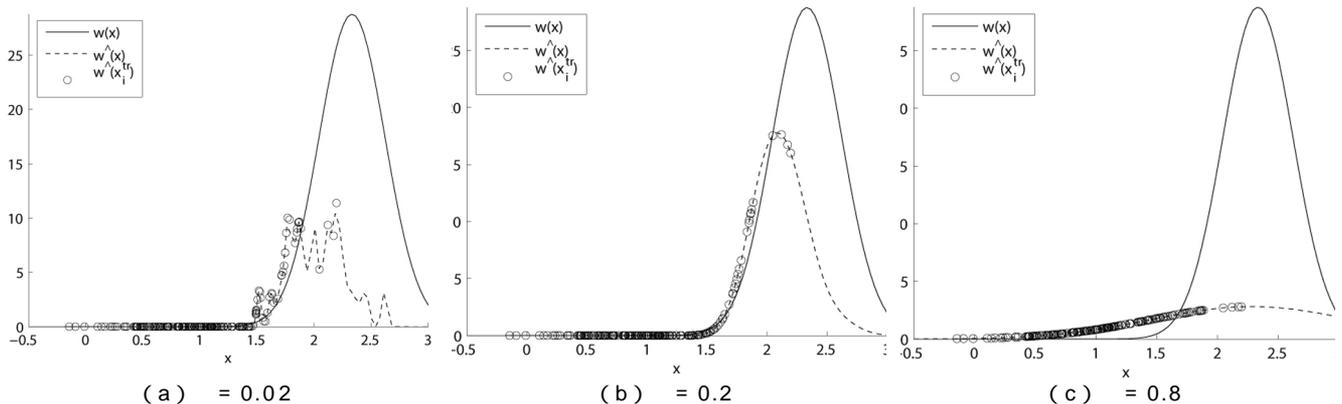
KLIEPでは、重要度 $w(x)$ を $x$ の関数とみなし、適当なモデルで近似する。ここでは次式のモデルを用いることにしよう。

$$w_\sigma(x; \alpha) = \sum_{\ell=1}^{n_{te}} \alpha_\ell \exp\left(-\frac{(x - x_\ell^{te})^2}{2\sigma^2}\right)$$

テスト入力データ数 $n_{te}$ が非常に大きい場合は、その一部だけの線形結合を考えてもよい。KLIEPによるパラメータの推定アルゴリズムを第7図に示す。これは、真のテスト密度 $p_{te}(x)$ からその推定 $\hat{w}(x) p_{tr}(x)$ へのKullback-Leiblerダイバージェンスが小さくなるようにパラメータを決定するアルゴリズムである。

第2図の外挿問題の例に対してKLIEPを適用した例を第8図に示す。この結果より、 $\sigma = 0.2$ のとき重要度をうまく推定できていることが分かる。しかし、KLIEPの性能は $\sigma$ の値に依存するため、 $\sigma$ の値を適切に決める必要があることもこの例から分かる。

$\sigma$ の値は尤度交差確認法 (likelihood CV; LCV) を使えばうまく決定することができる。尤度交差確認法を用いたKLIEPのモデル選択アルゴリズムを第9図に示す。尤度交差確認法は、対数尤度規準 $J = \mathbb{E}_{p_{te}} [\log \hat{w}(x; \hat{\alpha})]$ の推定法であり、値が大きいほど重要度の良い推定である。第8図の重要度推定結果の対数尤度規準の値を尤度交差確認法によって推定した結果を第10図に示す。このグラフより、尤度交差確認法によって対数尤度規準の非常に良い近似値が得られることが分かる。



第8図 KLIEPによる重要度の推定結果 $\hat{w}(x)$ 。KLIEPの推定性能はパラメータ $\sigma$ の値に依存するが、 $\sigma$ を適切に決定すれば重要度の良い推定値が得られる。

手順1:

テスト入力データ  $X = \{x_j^{(te)}\}_{j=1}^{m_{te}}$  を  $R$  個の部分集合  $\{X^{(r)}\}_{r=1}^R$  に分割する。

手順2:

$X^{(r)}$  以外のテスト入力データ  $X \setminus X^{(r)}$  と全ての訓練入力データ  $\{x_i^{(tr)}\}_{i=1}^{m_{tr}}$  を用いて、KLIEPにより  $\hat{w}_\sigma(x; \hat{\alpha}^{(r)})$  を求める。

手順3:

学習に用いていないテスト入力データ  $X^{(r)}$  を用いて、 $\hat{w}_\sigma(x; \hat{\alpha}^{(r)})$  の対数尤度規準の推定値  $\hat{J}_\sigma^{(r)}$  を次式で求める。

$$\hat{J}_\sigma^{(r)} = \frac{1}{|X^{(r)}|} \sum_{x \in X^{(r)}} \log \hat{w}_\sigma(x; \hat{\alpha}^{(r)})$$

手順4:

手順2から手順3を  $r=1, 2, \dots, R$  に対して繰り返し、モデル  $\hat{w}_\sigma(x; \hat{\alpha}^{(r)})$  の対数尤度規準の推定値  $\hat{J}_\sigma$  を次式で求める。

$$\hat{J}_\sigma = \frac{1}{R} \sum_{r=1}^R \hat{J}_\sigma^{(r)}$$

手順5:

手順2から手順4を幾つかの異なる  $\sigma$  の値に対して繰り返し、 $\hat{J}_\sigma$  が一番大きいモデル  $\hat{\sigma}$  を選ぶ。

手順6:

選んだモデルを全てのテスト入力データ  $X$  と全ての訓練入力データ  $\{x_i^{(tr)}\}_{i=1}^{m_{tr}}$  を用いて学習し、重要度の推定  $\hat{w}_\sigma(x; \hat{\alpha})$  を求める。

第3、4節で示した数値例では、KLIEPと尤度交差確認法によって重要度を推定していた。このことから、共変量シフト下では、KLIEP、尤度交差確認法、重要度重み付き最小二乗法、重要度重み付き交差確認法の四つの手法を組み合わせることにより、単純に最小二乗法と交差確認法を用いるよりもはるかに高い予測性能が獲得できることが分かる(第6図参照)。

重要度の推定法には、他にもカーネル法を用いた手法<sup>9)</sup> やロジスティック回帰を用いた手法<sup>5)</sup> などがある。

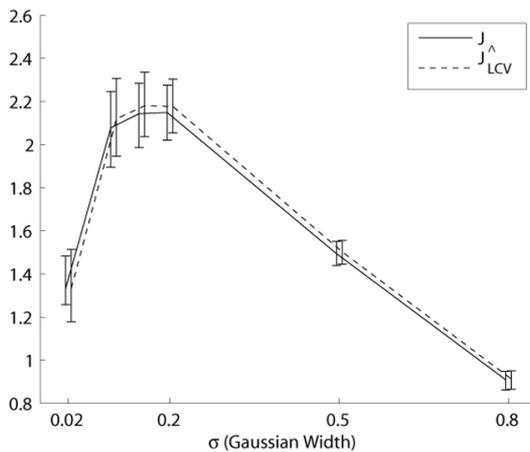
## まとめと今後の展望

本稿では、データの入力分布が変化する共変量シフト<sup>10)</sup> と呼ばれる状況における学習法について解説した。これらの方法の理論的背景や数値的特性についてはそれぞれの論文、及び、解説論文1)を参照されたい。

分布が変化するような状況は様々な応用場面で起こる。そのため、このような状況下での機械学習研究が近年注目されている。例えば、この話題に関連するワークショップが開催されている<sup>4), 7)</sup>。現実的な応用場面では、共変量シフトだけでなく他の様々な種類の分布の変化が起こる可能性がある。現在、様々な形態の分布の変化に対処できる学習法が開発されつつある。今後は、分布の変化を検出する方法、及び、分布の変化の種類を同定する方法を開発していくことが重要な研究課題である。

本研究の一部は、科学研究費補助金(若手研究(B)17700142、及び、基盤研究(B)18300057)、財団法人大川情報通信基金、マイクロソフト産学連携研究機構CORE3プロジェクトの支援を受けて行われた。

第9図 尤度交差確認法を用いたKLIEPのモデル選択アルゴリズム。尤度交差確認法は、対数尤度規準  $J = \mathbb{E}_{x \in \mathcal{X}} [\log \hat{w}(x; \hat{\alpha})]$  の推定値を与える推定法である。この値が大きいほど、 $\hat{w}(x; \hat{\alpha})$  は重要度  $w(x)$  の良い推定である。



第10図 対数尤度規準  $J$  の真の値とその尤度交差確認法による推定値  $\hat{J}_{LCV}$  の100回平均、上四分位点、下四分位点。 $\hat{J}_{LCV}$  は  $J$  の非常に良い近似である。

## 参考文献

- 1) 杉山 将. 共変量シフト下での教師付き学習. 日本神経回路学会誌, 13(3):111-118, 2006.
- 2) H.Akaike. A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19(6):716-723, 1974.
- 3) P.Baldi and S.Brunak. Bioinformatics: The Machine Learning Approach. MIT Press, Cambridge, 1998.
- 4) S.Bickel. ECML2006 workshop on discovery challenge, 2006.
- 5) S.Bickel, M.Brückner, and T.Scheffer. Discriminative learning for differing training and test distributions. In Proceedings of the 24th International Conference on Machine Learning, 2007.
- 6) S.Bickel and T.Scheffer. Dirichlet-enhanced spam filtering based on biased samples. In B.Schölkopf, J.Platt, and T.Hoffman, editors, Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA, 2007.
- 7) J.Q. Candela, N.Lawrence, A.Schwaighofer, and M.Sugiyama. NIPS2006 workshop on learning when test and training inputs have different distributions, 2006.
- 8) J.J. Heckman. Sample selection bias as a specification error. Econometrica, 47(1):153-162, 1979.
- 9) J.Huang, A.Smola, A.Gretton, K.M. Borgwardt, and B.Schölkopf. Correcting sample selection bias by unlabeled data. In B.Schölkopf, J.Platt, and T.Hoffman, editors, Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA, 2007.
- 10) H.Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of Statistical Planning and Inference, 90(2):227-244, 2000.
- 11) M.Sugiyama, M.Krauledat, and K.-R. Müller. Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research, 8:985-1005, May 2007.
- 12) M.Sugiyama and K.-R. Müller. Input-dependent estimation of generalization error under covariate shift. Statistics & Decisions, 23(4):249-279, 2005.
- 13) M.Sugiyama, S.Nakajima, H.Kashima, P.von Bünau, and M.Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. Advances in Neural Information Processing Systems 20(to appear).
- 14) R.S. Sutton and G.A. Barto. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, 1998.
- 15) J.R. Wolpaw, N.Birbaumer, D.J. McFarland, G.Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. Clinical Neurophysiology, 113(6):767-791, 2002.

## 【筆者紹介】

杉山 将

東京工業大学

大学院情報理工学研究科

計算工学専攻

准教授

〒152-8552

東京都目黒区大岡山2-12-1-W 8-74

TEL : 03-5734-2699

FAX : 03-5734-2699

E-mail : sugi@cs.titech.ac.jp

主なる業務歴及び資格

研究内容及び論文は

<http://sugiyama-www.cs.titech.ac.jp/> sugi  
をご覧ください。

<b>Keyword</b>	教師付き学習	質問とその答えの組から背後に潜む規則を推定する学習の問題。
	汎化能力	学習機械のテストデータに対する予測性能。
	共変量シフト	教師付き学習において訓練データとテストデータの入力分布が異なる状況。
	重要度	各訓練データのテストデータに対する相対的な重要度。
	交差確認法	学習機械の汎化能力を推定する方法の一つ。