

深層学習を用いた脳波信号からの運動意図推定

佐野 光
Sano Hikaru

2017 年 12 月 27 日

目次

1	研究背景	2
1.1	研究背景	2
1.2	BCI の概要	2
1.2.1	外部刺激型 BCI	2
1.2.2	運動想起型 BCI	2
2	BCI のための要素技術	3
2.1	特徴量工学	3
2.1.1	他分野での特徴量工学	3
2.1.2	脳波解析における特徴量工学	3
2.2	フーリエ解析と時間周波数解析	3
2.2.1	周波数スペクトル	3
2.2.2	スペクトログラム	4
2.3	多変量解析	4
2.3.1	Principal Component Analysis(PCA)	4
2.3.2	Indipendent Component Analysis(ICA)	6
2.3.3	Blind Source Separation(BSS)	6
2.4	分類手法	9
2.4.1	Linear Discriminant Analysis(LDA)	9
2.4.2	Support Vector Machine(SVM)	10
2.4.3	Logistic Regression(LR)	13
3	従来の運動想起型 BCI	14
3.1	Common Spatial Pattern とその派生手法	14
3.1.1	脳波信号の定式化	14
3.1.2	Common Spatial Pattern(CSP)	14
3.1.3	Common Spatiospectral Pattern(CSSSP)	15
3.1.4	Filter Bank Common Spatial Pattern(FBCSP)	16
3.2	従来手法のまとめ	16
3.2.1	従来の運動想起 BCI の構成と問題点	17
4	提案手法	19
4.1	ニューラルネットワーク	19
4.1.1	フィードフォワードニューラルネットワーク	19
4.1.2	リカレントニューラルネットワーク	19
4.1.3	畳み込みニューラルネットワーク	19
4.2	FIR フィルタと畳み込み層	19
5	実験と結果	20

6	結論と考察	21
	Appendix	22
A.1	セクション	22
A.1.1	サブセクション	22
	参考文献	24

数学的表記について

本文中で用いられる数学的な表記の中で、慣習的に用いているものを記載する。ここに記載した数式は再度本文中で明示する場合もあるが、特に誤解を招かない場合には予告なしに用いる。

x	ベクトル（スカラーを表す場合には本文中で明記して用いる）	(1)
x^T	転置記号	(2)
$ x $	ユークリッドノルム	(3)
$ x _p$	l_p ノルム	(4)
X	行列（他の用途の場合には明記する）	(5)
X^{-1}	逆行列記号	(6)
\mathbb{R}	実数集合	(7)
\mathbb{R}^m	m 次元実数集合	(8)
$\mathbb{R}^{m \times n}$	m 次元実数集合と n 次元実数集合の直積集合	(9)
$\mathbb{I}(\cdot)$	指示関数	(10)
$a \in A$	a は集合 A の元	(11)
$B \subset A$	B は集合 A の真部分集合	(12)
$B \subseteq A$	B は集合 A の部分集合	(13)
π	円周率	(14)
e	ネイピア数	(15)
$\log(\cdot)$	ネイピア数を底とした対数関数	(16)
$p(\cdot)$	確率密度関数あるいは確率質量関数	(17)
$q(\cdot)$	確率密度関数あるいは確率質量関数	(18)
$p(x_1, x_2)$	x_1, x_2 の同時確率	(19)
$p(x_1 x_2)$	x_2 に条件付けられた x_1 の条件付き確率	(20)
$\mathcal{N}(\mu, \sigma^2)$	スカラーの平均 μ 、分散 σ^2 の 1 次元ガウス分布	(21)
$\mathcal{N}(\mu, \Sigma)$	平均ベクトル μ 、分散共分散行列 Σ の多次元ガウス分布	(22)
$\mathbb{E}[\cdot]$	期待値演算	(23)
$\mathbb{E}[x_1 x_2]$	x_2 に条件付けられた x_1 の条件付き期待値	(24)
$\mathcal{KL}(p q)$	p から q へのカルバック・ライブラーダイバージェンス	(25)
\mathcal{D}	データ集合	(26)

研究背景

1.1 研究背景

Brain Computer Interface(BCI) とは、脳とコンピュータを何らかの方法で接続し脳活動とコンピュータの動作を関連付ける技術の総称である。BCI は脳とコンピュータの接続方法に関して「侵襲式」と「非侵襲式」の2つに分類され、実応用を考える上では非常に重要な項目となる。侵襲式と非侵襲式とは、BCI を構築する際に外科手術を要するか否かを指す言葉であり、非侵襲式では外科手術を伴わない接続方法を用いて BCI を構築する。

非侵襲式は侵襲式に比べ人間への負担が明らかに少なく、BCI の実用化において大きなアドバンテージを有する。一方で、脳活動の計測精度という観点では劣り、BCI の構築のためには多くの工学的問題を解決する必要がある。具体的には、計測時に混入するノイズの除去の問題や、脳活動の信号源の推定の問題を解決しなければならないと考えられている。本稿では特に脳活動を頭皮上の電位変化である Electroencephalogram (EEG) として計測し、計測データに応じてコンピュータが特定の処理を行うタイプの BCI に着目する。

1.2 BCI の概要

1.2.1 外部刺激型 BCI

1.2.2 運動想起型 BCI

本研究で着目する運動想起型 BCI とは、人間が肢体運動を能動的に想像したことを検知する BCI である。主な応用としては麻痺患者のリハビリテーションが想定されているが、他にも肢体運動を外部機器の操作に対応させることで依り汎用的な応用が期待できる。

2

BCI のための要素技術

2.1 特徴量工学

運動想起 BCI の標準的な役割は、脳波信号 $x(t) \in \mathbb{R}^D$ に対応した運動想起部位を識別することである。ここに D は電極の個数である。通常は脳波計によって計測された $x(t)$ から直接運動想起部位の識別を行うのは困難である。そこで、何らかの変換 $f(\cdot)$ を脳波信号 $x(t)$ に対して施すことで、識別に有利な特徴量 $z = f(x(t))$ の獲得を目指すのが特徴量工学である。

2.1.1 他分野での特徴量工学

特徴量工学は BCI に限らず、データ解析を必要とする多くの分野で活用されている。例えば画像処理の分野では、画像に写っている物体の輪郭を抽出するエッジ処理や、画像に対してぼかしを入れる平滑化処理などが特徴量抽出手法として用いられる。また音声認識の分野では、音声の周波数に重要な情報が含まれていると考え、周波数スペクトルやケプストラムなどを特徴量として用いる試みがなされてきた。

2.1.2 脳波解析における特徴量工学

特に脳波

2.2 フーリエ解析と時間周波数解析

2.2.1 周波数スペクトル

周波数スペクトルは、脳波研究の知見を使った有効な特徴量であると考えられる。1 章で述べたように、運動想起時には想起部位に応じた頭皮上において、脳波の特定の周波数のパワーが減少する事象関連脱同期が生じることが知られている。従って、脳波の周波数スペクトルは運動想起時と休息時で異なると期待され、良い特徴量になることが想定される。[?]。

この手法の主な流れを記す。まず脳波信号を $x(t) \in \mathbb{R}^D$ と表記する。ここに、 D は電極の個数である。まず電極に対しての重み付けベクトル (空間フィルタ) である $w \in \mathbb{R}^D$ を何らかの方法で決定することで、

$$z(t) = w^T x(t) \in \mathbb{R} \quad (2.1)$$

とスカラー値の時間波形に変換する。空間フィルタとしては着目したい頭皮領域に対して“スモークラプリアンフィルタ”のような決定論的な手法を用いる場合もあれば、PCA や ICA、また次節で述べる Common Spatial Pattern(CSP) のような統計的な指標に基づいた手法を用いる場合もある。次に $z(t)$ に対してフーリエ変換 $\mathcal{F}(\cdot)$ を行い、スペクトル

$$A(f) = \mathcal{F}(z(t)) \quad (2.2)$$

を獲得する。運動想起が行われた場合には特定の周波数において、パワースペクトル $A(f)^2$ の減少が生じると期待できる。

2.2.2 スペクトログラム

フーリエ変換の代わりに $z(t)$ に対して短時間フーリエ変換 $STFT(\cdot)$ を行い、

$$A(t, f) = STFT(z(t)) \quad (2.3)$$

を獲得し、スペクトログラムを特徴量として用いることもできる。運動想起開始から事象関連脱同期がどのタイミングでどのくらいの長さ継続するのも反映することができる。あるいはウェーブレット変換などの他の時間周波数解析を用いることも可能である。

2.3 多変量解析

多変量解析とは多変量データを統計的に扱う手法である。扱うデータの科学的知識に基づいた特徴量抽出に加え、実データの統計的な性質を考慮した特徴量抽出も様々な分野で行われている。この章では脳波解析で用いられている多変量解析について紹介し、その有用性と限界について考察する。

2.3.1 Principal Component Analysis(PCA)

Principal Component Analysis(PCA) は特徴量抽出手法として幅広い分野で活用されている。簡単のため時間平均が 0 の多次元信号 $x(t) \in \mathbb{R}^D$ に対して PCA による特徴抽出を考える。PCA では変換行列 $W^T \in \mathbb{R}^{D \times D}$ を左から作用させ、特徴量

$$z(t) = W^T x(t) \in \mathbb{R}^D \quad (2.4)$$

を獲得するが、この際、 $z(t)$ の各成分が互いに無相関になるように W を決定する。 $z(t)$ が無相関となるためには、その分散共分散行列 $\Sigma_z = \mathbb{E}[z(t)z^T(t)]$ が対角行列になることが要請される。ここで $x(t)$ の分散共分散行列を $\Sigma_x = \mathbb{E}[x(t)x^T(t)]$ とすると、 $z(t)$ の分散共分散行列 Σ_z について

$$\begin{aligned} \Sigma_z &= \mathbb{E}[z(t)z^T(t)] \\ &= \mathbb{E}[W^T x(t)x^T(t)W] \\ &= W^T \mathbb{E}[x(t)x^T(t)]W \\ &= W^T \Sigma_x W \end{aligned} \quad (2.5)$$

と表すことができる。(2.5) が対角行列になるような W は、 Σ_x の固有値分解によって求まる。実数信号の分散共分散行列は一般に正定値実対称行列となっており、直交行列によって固有値分解が可能である。また、固有値は必ず正の値となる。これらの数学的な扱いやすさから PCA は非常に広く普及している。

図 2.1 は人工的に作成した 2 次元波形 $x(t)$ と各成分の散布図である。この $x(t)$ に対して PCA を用いると、図 2.2 に示す $z(t)$ が得られ、散布図から $z(t)$ の各成分は無相関となっていることが確認できる。図 2.3 は $x(t)$ に対して PCA を施した場合に得られる新たな直交基底を表しており、 $z(t)$ は新たな直交基底に $x(t)$ を射影したものに他ならない。PCA では基底を取り直すことで各成分が無相関な信号を獲得でき、その結果、新たに得られた信号の各成分がどのような意味を持つのかを考察しやすくなる。

しかし、応用上は単に基底を取り直すことを目的とするケースは少ない。通常は固有値分解によって求めた固有ベクトル（新たな基底）を全て利用するのではなく、値の大きな固有値に属する固有ベクトル w を $d(< D)$ 個選び、 $W_d = (w_1, \dots, w_d)$ によって変換行列を構成することで、

$$z(t) = W_d^T x(t) \in \mathbb{R}^d \quad (2.6)$$

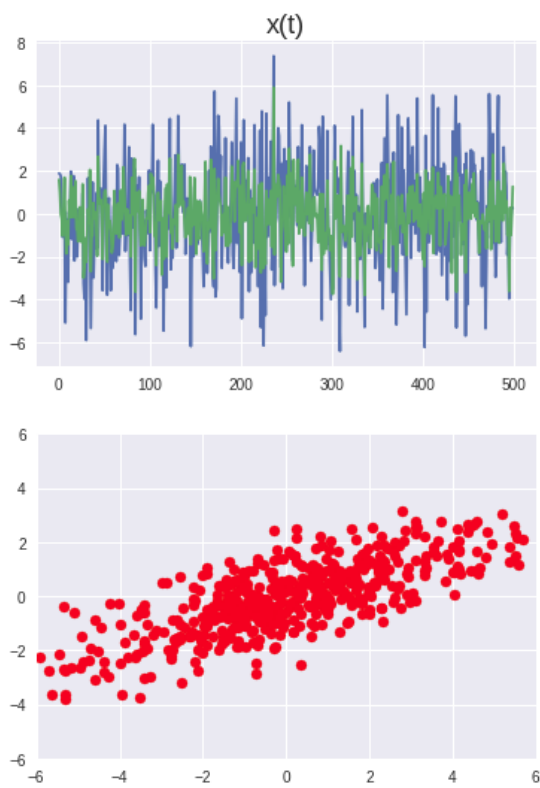


図 2.1: $x(t)$ の波形と散布図

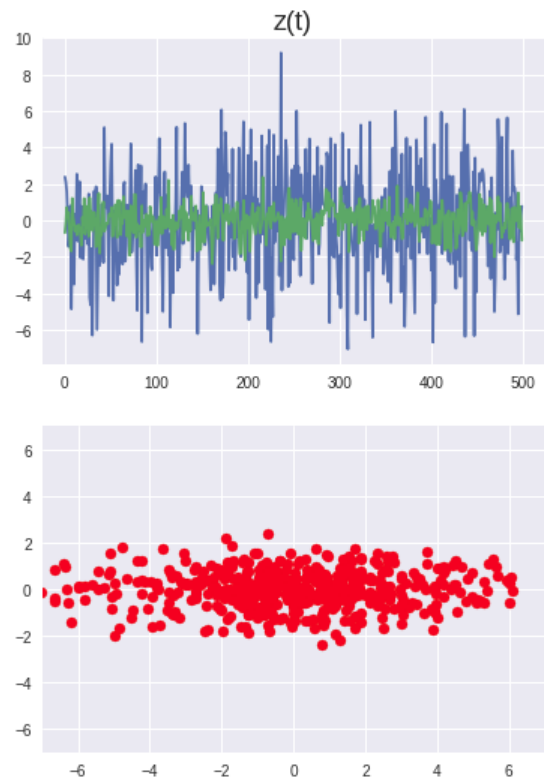


図 2.2: $z(t)$ の波形と散布図

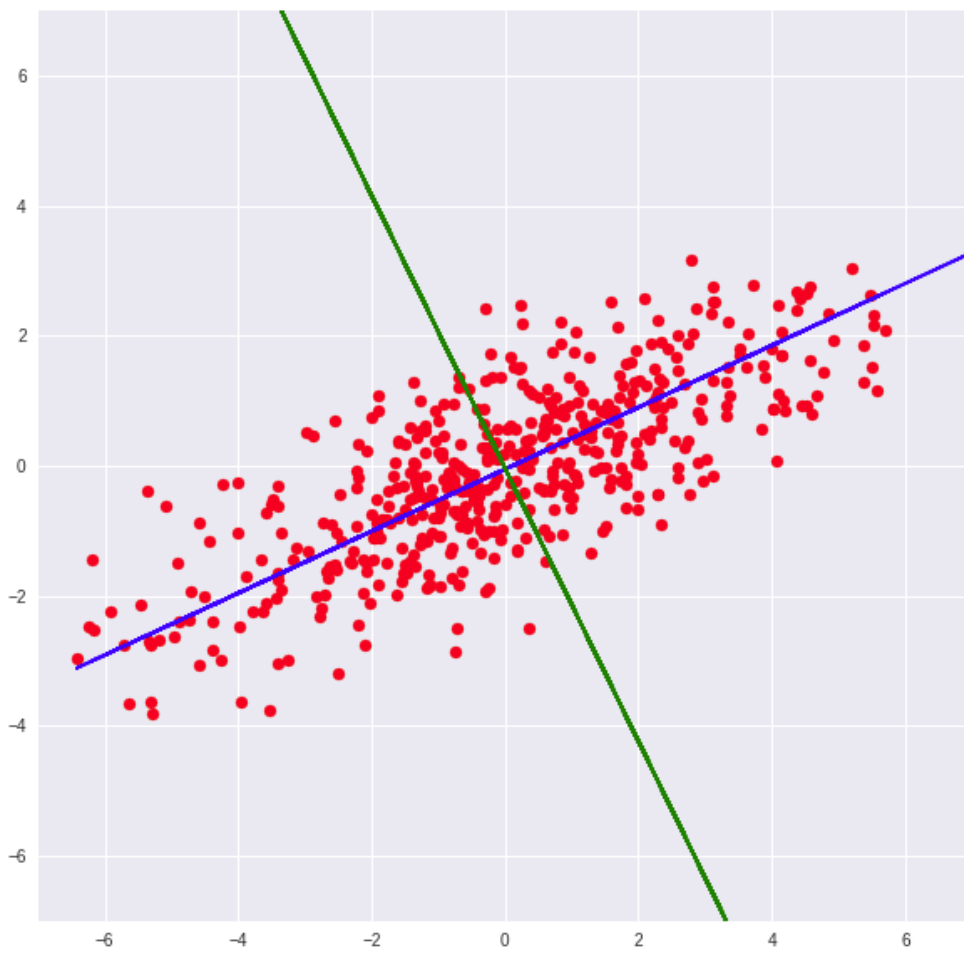


図 2.3: PCA によって得られる基底

と次元削減を行う。値の大きな固有値に属する固有ベクトルによって基底を構成することは、 W_d による基底の元で、信号の分散（あるいは振幅）が最大化されることを要請することと等価である。また、射影先での分散最大化に伴い、(2.6) での変換行列 W_d は、元々の信号 $x(t)$ と、 $x(t)$ を \mathbb{R}^D の部分空間 \mathbb{R}^d へ射影した信号 $z(t)$ との二乗誤差を最小化する変換行列となっている。

以上から PCA は、元々の信号 $x(t)$ の情報損失を二乗誤差の意味で最小限に抑えながら、射影先で大きな変動を有し、かつ各成分が無相関となる特徴量 $z(t)$ を抽出する。しかし、脳波への応用を考える上では PCA の性質は必ずしも有効には働かない。運動想起 BCI を考える上では、脳波に含まれる全ての情報の中から識別したい身体部位に関する情報のみを抽出する必要がある。この場合、脳波信号のごく一部のみが重要である可能性があり、射影先で大きな分散を持つような信号となっているかは定かではないためである。

2.3.2 Independent Component Analysis(ICA)

Independent Component Analysis(ICA) は PCA を発展させた比較的新しい信号解析手法である。ICA は PCA と同様に変換行列 $W^T \in \mathbb{R}^{D \times D}$ を左から作用させ、特徴量

$$z(t) = W^T x(t) \in \mathbb{R}^D \quad (2.7)$$

を獲得するが、この際、 $z(t)$ の各成分が互いに独立になるように W を決定する。独立性は無相関性の十分条件であり、PCA に比べて $z(t)$ により強い条件を要請する。PCA では無相関性が固有値分解という数学的によく知られた問題と関連していたが、独立性は簡単な問題への定式化は困難であるため、通常は独立性を測る目的関数を設定し、勾配法などの逐次最適化法を用いる。最も普及している求解アルゴリズムは FastICA として知られており、独立性の必要条件である無相関性を要請した後、独立性を最大化することで解を得る。図 2.4 に PCA と ICA の振る舞いをトイデータを用いて比較を掲載する。図 2.4 の base vectors of ICA and PCA から分かるように ICA で得られる基底は直交するとは限らない。また、PCA の基底の大きさは等しく 1 であり正規直交基底を構築するが、ICA では正規性も持つとは限らない。

脳波への応用では、脳波計測時に混入した脳波以外の信号（筋電位、眼電位など）が脳波とは統計的に独立であると考え、脳波以外の信号成分を除去する目的で利用される。一方で運動想起 BCI において身体部位に関する脳波成分を直接抽出することは PCA 同様に難しい。筋電などの場合、脳波と独立であるという仮定は妥当であり、かつ振幅が目視可能なほど脳波に比べて大きくなる。従って、ICA によって分解された独立な成分から筋電などを見分けるのは比較的容易である。しかし、一方で身体部位に関する脳波成分がその他のあらゆる脳波と独立であるかは定かではなく、仮に独立であった場合にも分解された信号から目視によって特定することは困難であると推察される。

2.3.3 Blind Source Separation(BSS)

上記では PCA と ICA が次元削減として用いられることを見た。一方でこれらの手法は Blind Source Separation(BSS) 問題の解法として解釈されることも多いため、ここで簡単に述べておく。まず信号源 $s(t) \in \mathbb{R}^d$ を直接観測できない場合に、 D 個のセンサで $x(t) \in \mathbb{R}^D$ という信号を観測したとする。このとき、観測信号 $x(t)$ のみから信号源 $s(t)$ を推定する問題が BSS 問題である。計測機器や環境に応じて、信号源 $s(t)$ は何らかの変換 $f(\cdot)$ を受けて観測されると考えられる。従って $x(t)$ は

$$x(t) = f(s(t)) \quad (2.8)$$

と表記できる。このときに観測に伴う変換 $f(\cdot)$ が線形変換 A であると仮定した場合、

$$x(t) = As(t) \quad (2.9)$$

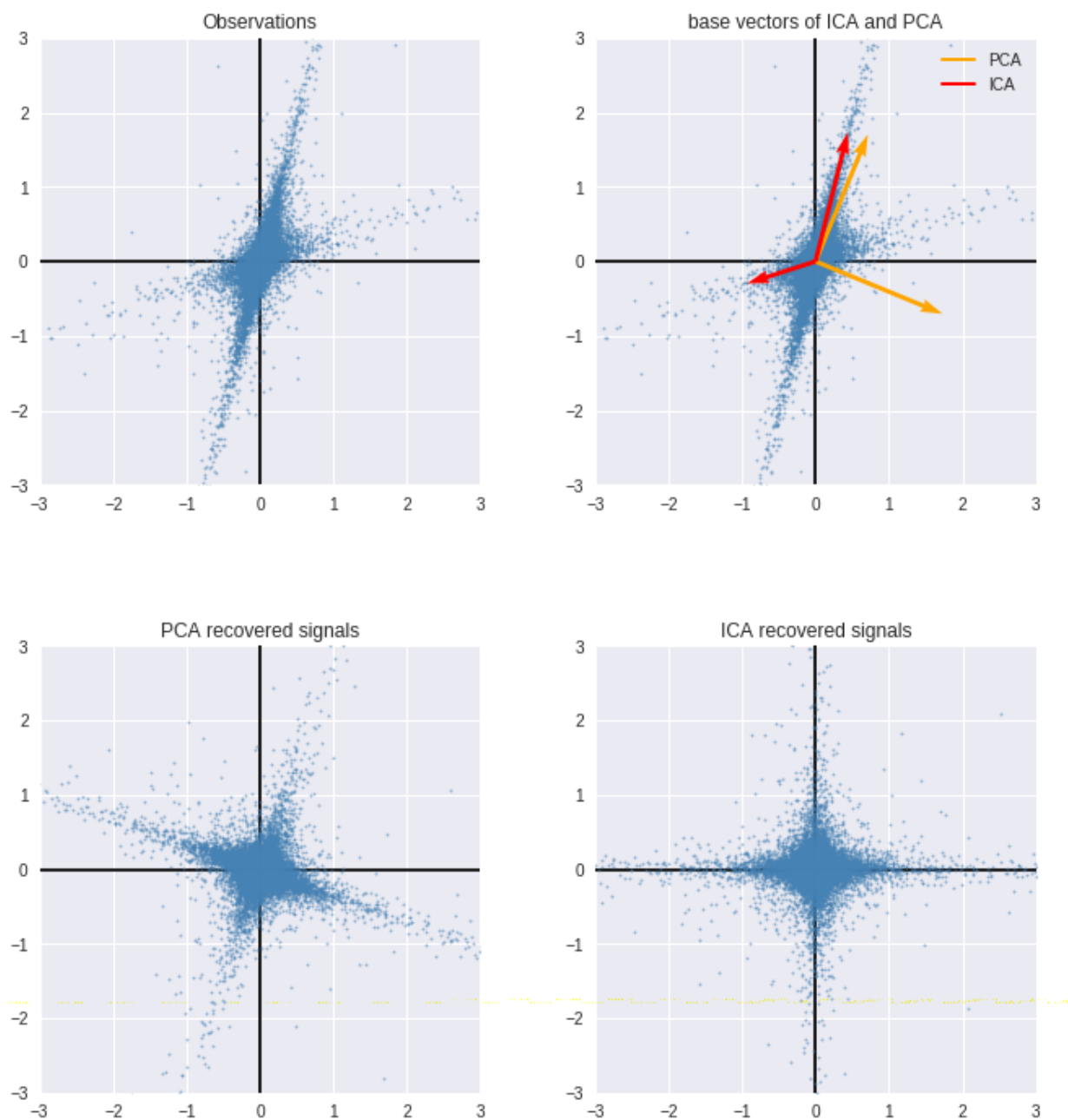


図 2.4: 無相関な基底を得る PCA と独立な基底を得る ICA の比較

と表記することができ、BSS 問題は $x(t)$ から A と $s(t)$ を同時に推定する問題であると見なせる。ここで仮に適当な線形変換によって、観測信号 $x(t)$ を

$$z(t) = Wx(t) \quad (2.10)$$

と変換することを考える。 W を上手く選ぶことに成功すれば、 $z(t) = Wx(t) \simeq s(t)$ となることが期待できる。ここで $D = d$ 、すなわち信号源の次元と観測信号の次元が一致している場合を考える。このとき (2.9) において、 A が正則であるとし、

$$s(t) = A^{-1}x(t) \quad (2.11)$$

と表すことが可能になる。従って、(2.10) の W を $W = A^{-1}$ とすることができれば、

$$z(t) = Wx(t) = A^{-1}x(t) = s(t) \quad (2.12)$$

と信号源を求めることが可能である。ただし、 $z(t) = s(t)$ となる W が存在するとしても、既知の $x(t)$ に対して未知の $W, s(t)$ を求めようとしている状況に変わりはなく、BSS 問題は基本的に不良設定問題である。また、実データでは信号源と観測信号の次元が一致しない場合が多く A は逆行

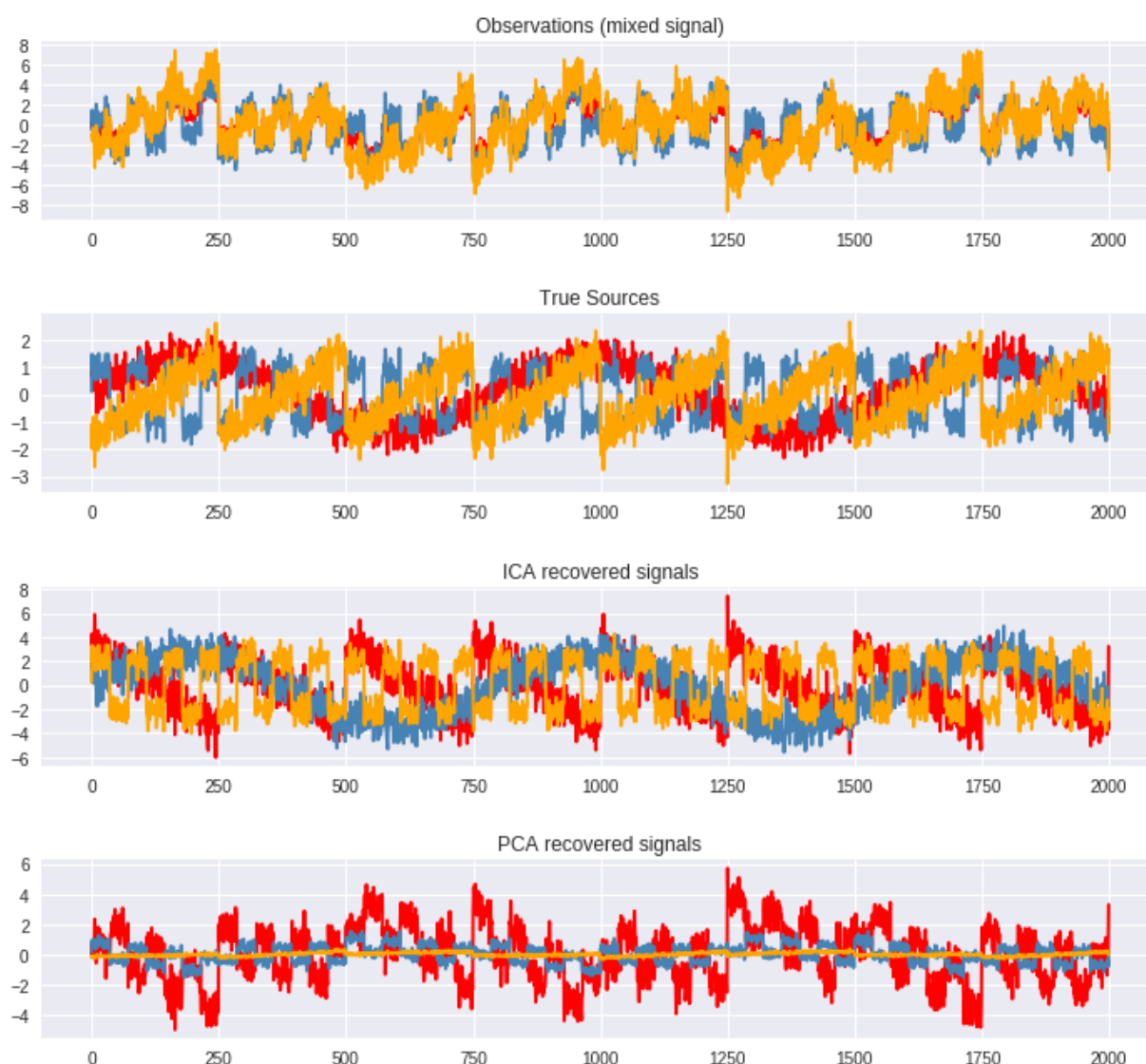


図 2.5: BSS 問題に対する PCA と ICA

列を持たないため、状況はより複雑である場合が多い。通常は BSS 問題を解くためには何らかの条件を追加するか、正則化の手法を導入する必要がある。PCA や ICA は信号源 $s(t)$ が各成分について無相関あるいは独立であると仮定することで条件式を追加し、観測信号 $x(t)$ から条件式を満たすような W と $z(t)$ を求め、 $z(t)$ が信号源 $s(t)$ の良い近似になっていると考える BSS 問題の解法の一つである。

PCA と ICA の BSS 問題への振る舞いを確認するために、トイデータによる実験結果を図 2.5 に示す。周波数の異なる 2 つの正弦波と 1 つのノコギリ波にそれぞれガウスノイズを加算した信号源 (図 2.5 の True Sources) を準備し、適当な線形変換を施して観測信号 (図 2.5 の Observations) とする。図 2.5 の ICA recovered signals が FastICA によって推定された信号源に適当なゲインを加えたものであり、図 2.5 の PCA recovered signals が PCA によって推定された信号源に適当なゲインを加えたものである。ICA では周波数の異なる正弦波とノコギリ波を明確に分解できており、信号源に近い波形が得られていることが確認できるが、PCA では信号源と異なる信号が得られている。PCA の振る舞いは各成分を無相関にしつつ、射影先で分散を最大化するような基底を求めるためノコギリ波と位相が一致している正弦波を 1 つの成分に集約してしまっている。

トイデータによる実験では ICA が BSS 問題に対して有効に働くことが確認できるが、本来信号源がどのようなものであるかは未知であるため、信号源推定が正しく行われたかを確認するのは実データでは困難である。運動想起 BCI を想定して脳波に BSS 問題を適用する場合には、観測された信号 $x(t)$ から脳波信号の根源である $s(t)$ を復元することを目的とするが、ICA によって推定さ

れた信号源のいずれの成分が運動想起と関連しているのかを判別するのは極めて難しい。ただし、脳波と筋電では波形が明らかに異なるため、筋電と脳波が独立であるという仮定を用いて ICA によって筋電成分を取り出すことは可能である。一般に、不良設定問題である BSS 問題に対しては何らかの仮定を置かざるを得なく、複雑な脳波に対して適切な仮定を設けることが重要な課題となる。

2.4 分類手法

分類問題の説明を行う。

2.4.1 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis(LDA) は統計分析において伝統的に用いられてきた歴史ある手法である。特に BCI では CSP による特徴量抽出の後に用いられることが多い。LDA では多次元データを部分区間で切り取り、部分空間で分類超平面を構築することでクラス分類を行う。分類超平面を構築する手段を与えなければ、LDA は特徴量抽出手法としても機能する。まず、多次元データ $x \in \mathbb{R}^D$ を基底 $w \in \mathbb{R}^D$ へ射影すると、以下の式で表されるスカラー値を獲得できる。

$$z = w^T x \in \mathbb{R} \quad (2.13)$$

z に対してある閾値を設定し、 $z \geq -w_0$ の場合はクラス C_1 とし、そうでない場合はクラス C_2 であるとする事で分類器を獲得できる。多次元データを 1 次元空間へ射影した場合には多くの情報損失が生ずるが、 w の取り方を上手く調整することによって、クラス分離を行いやすい射影を選択できる。クラス分類性能を向上させる w を得るために、まず以下のようにクラス毎の平均ベクトル m_1, m_2 を定義する。

$$m_1 = \frac{1}{|C_1|} \sum_{x \in C_1} x \quad (2.14)$$

$$m_2 = \frac{1}{|C_2|} \sum_{x \in C_2} x \quad (2.15)$$

ここに、 $|C_i|$ はクラス C_i に属するデータの数である。クラス C_1 とクラス C_2 の平均が射影先で大きな値となれば分離が行われやすいと想定できる。従って、以下の距離の最大化によって分類性能が向上すると考えられる。

$$d = |w^T(m_1 - m_2)| \quad (2.16)$$

しかし実際にはクラス毎の平均を考慮しただけでは分類が上手く行くとは限らない(図 2.6)。射影先での各クラスのデータの分散が大きい場合には、異なるクラスのデータが重なってしまう場合が生じるからである。この問題を解決するためにはデータの分散を考慮する必要がある。まず射影先での各クラスの分散は以下で表記できる。

$$\sigma_1^2 = \sum_{x \in C_1} \{w^T(x - m_1)\}^2 \quad (2.17)$$

$$\sigma_2^2 = \sum_{x \in C_2} \{w^T(x - m_2)\}^2 \quad (2.18)$$

ここで、全データのクラス毎の分散の和を総クラス内分散として以下で定義する。

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \quad (2.19)$$

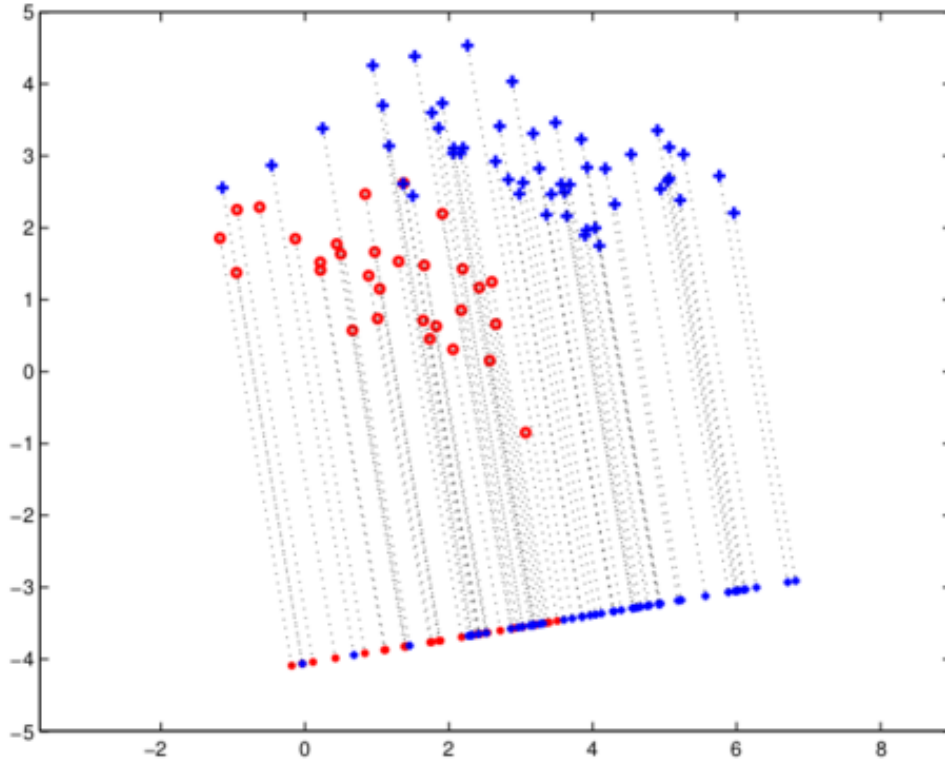


図 2.6: クラス毎の平均間の距離を射影先で最大化した判別分析

総クラス内分散 (2.19) を小さくしながらクラス間の平均の距離 (2.16) を大きくすることを考慮し LDA では以下の評価関数を用いる。

$$J(w) = \frac{d^2}{\sigma^2} = \frac{\{w^T(m_1 - m_2)\}^2}{\sum_{x \in C_1} \{w^T(x - m_1)\}^2 + \sum_{x \in C_2} \{w^T(x - m_2)\}^2} \quad (2.20)$$

このクラス間の平均とクラス内の分散を考慮した評価関数を用いることで、射影先でデータがクラス毎に小さくまとまり、かつ異なるクラスのデータがなるべく離れるようになる (図 2.7)。

1 次元の空間 (数直線上) でクラス毎にデータが上手く分離できた後、数直線上に閾値 $w_0 = \frac{m_1 + m_2}{2}$ を設けることで分離平面が得られる。これはクラス毎の平均値の平均値によって分離平面を設定したことに相当する。しかし、射影のされ方によってはこの閾値は適当ではない。より精密な分類を行うためには、 $z = w^T x$ がクラス毎に異なるガウス分布から生じる確率変数だと考え、クラス C_i の条件付き確率 $p(z|C_i)$ を算出した判別基準を設け、条件付き確率が大きいクラスへ分類するなどの方法を取る。

2.4.2 Support Vector Machine(SVM)

脳波の分類では Support Vector Machine(SVM) の応用例もある。基本的に SVM はマージン最大化の考えによって汎化性能の向上に成功した 2 クラス分類のための線形分類器である。まずマージン最大化という概念について説明する。マージンとは端的に述べるとデータ点と分類超平面との距離のことを表す。学習データに対してマージンを最大化することで、学習データが空間上で僅かに移動した際にも誤分類を起こしづらくなると期待できる。SVM ではこのマージン最大化によって以下の分類超平面を定める。

$$y(x) = w^T x + w_0 \quad (2.21)$$

ここに x は D 次元のデータベクトルであり、 w は D 次元のパラメータベクトルである。 w_0 もスカラーパラメータであり閾値の役割を担う。分類面の役割により $y(x)$ は x がクラス C_1 に属する

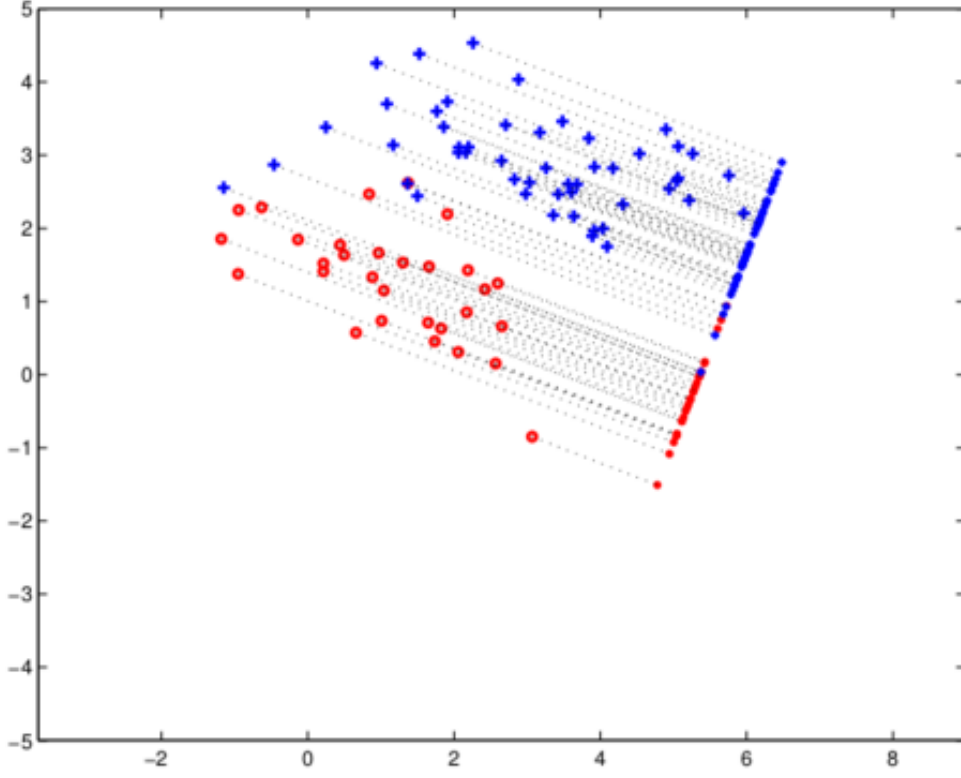


図 2.7: クラス毎の平均間の距離を射影先で最大化した判別分析

場合には正の値を、 C_2 に属する場合には負の値を取るよう学習される。ここで、(2.26) の超平面と、あるデータ点 x_n との距離は以下で表される。

$$|r| = \frac{|y(x_n)|}{|w|} \quad (2.22)$$

ここで、 x_n がクラス C_1 に属する場合は $t_n = 1$ とし、クラス C_2 に属する場合には $t_n = -1$ と定めた t_n を導入する。さらに x_n には分類面から最も近いデータ点のみを考慮することとし、そのときの $|r|$ をマージンと呼び以下で表す。

$$|r|_{margin} = \min_{x_n} \frac{t_n y(x_n)}{|w|} \quad (2.23)$$

この (2.23) を最大化するようにパラメータを決定することでマージン最大化を実現することができる。従って、SVM のパラメータ決定は以下の最適化問題によって定式化される。

$$\arg \max_{w_0, w} \left(\min_{x_n} \frac{t_n y(x_n)}{|w|} \right) \quad (2.24)$$

しかしこの最適化問題において、 w, w_0 の大きさは本質的ではない。なぜなら w, w_0 を同時に kw, kw_0 と定数倍した場合にも (2.23) の値は変化しないためである。従って、 w, w_0 の大きさに関して制約を設ける必要がある。そこで分類面から最も近い x_n に関して $t_n(w^T x + w_0) = 1$ となるような制約を w, w_0 に対して要請する。この条件式に伴って、任意のデータ点において $t_n(w^T x + w_0) \geq 1$ という制約が与えられる。最終的にマージン最大化問題 (2.24) は以下で定式化される。

$$\begin{aligned} \arg \min_{w_0, w} \quad & \frac{|w|^2}{2} \\ \text{s.t.} \quad & t_n(w^T x + w_0) \geq 1 \end{aligned} \quad (2.25)$$

目的関数の分母は、単に勾配を計算する際に約分できるというテクニックによるものである。分子の二乗は、最適化問題の解を変更せずに勾配計算などを容易に行うための変形である。図 2.8 に分類面の定め方によりマージンが異なっている様子を見ることができる。左右いずれの図も学習デー

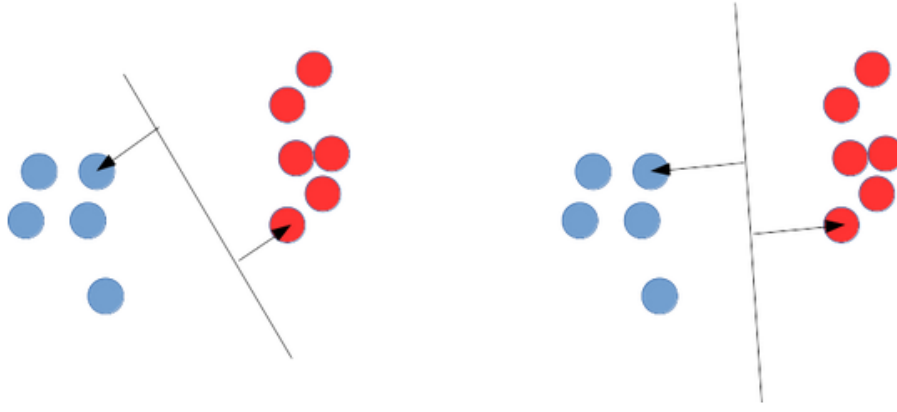


図 2.8: 分類面によってマージンの大きさが異なる様子

タに対して正しく分類が行える分類面になっているが、新規のデータに対しての分類結果が異なってくる。SVM では右図の分類面の方が優れていると考える場合に用いる手法である。ここまで線形分類を行う場合の SVM を見てきたが、一般的に線形分類器はデータ点を非線形関数 $\phi(\cdot)$ によって別の特徴空間へ写像し、特徴空間上で線形分類を行う問題へ拡張することができる。SVM では分類超平面を以下の式によって構築することに相当する。

$$y(x) = w^T \phi(x) + w_0 \quad (2.26)$$

この場合においてもこれまでと同様の議論で最適化問題を以下のように定式化できる。

$$\begin{aligned} \arg \min_{w_0, w} \quad & \frac{|w|^2}{2} \\ \text{s.t.} \quad & t_n(w^T \phi(x) + w_0) \geq 1 \end{aligned} \quad (2.27)$$

実データは線形分離不可能な場合が多いため、通常 SVM を用いる場合は上記のような非線形に拡張されたものを用いる。また実データは異なるクラスのデータが重なって分布することも多々あるため、厳密な分類を行うことは不可能な場合が多い。そういった場合に対応したソフトマージンと呼ばれる考えがあり、学習データの誤分類に対して寛容になる指標を導入する。このソフトマージンの考え方は機械学習で過学習抑制に用いられる正則化の考えと本質的には変わらない。また、制約付き最適化問題をラグランジュ法によって変形することで双対問題を獲得することができる。双対問題においては非線形変換後の内積のみが必要となり、具体的な非線形変換の計算をデータ x に対して実施する必要はない。すなわち、非線形変換 $\phi(x)$ を具体的に考える代わりに、その内積 $\phi(x_i)^T \phi(x_j)$ を定義することで SVM の非線形への拡張が可能である。このような方法はカーネル法として知られており、このときに用いられる内積計算をカーネル $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ と呼ぶ。カーネルは内積の公理を満たせば直ちに適用できるわけではないため、適用可能条件について注意する必要がある。ただし、BCI も含めた多くの応用では既に知られた優れた性質を持つカーネルを活用することがほとんどである。以下に代表的なカーネルについて記載する。

- ・線形カーネル：非線形変換を行わないことに対応。

$$k(x_i, x_j) = x_i^T x_j$$

- ・多項式カーネル：多項式関数による非線形変換に対応。

$$k(x_i, x_j) = (x_i^T x_j + c)^M$$

- ・ガウス基底カーネル：特徴空間が無限次元となる。

$$k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right)$$

通常、カーネルを用いる場合にはハイパーパラメータが付随する。多項式カーネルの場合は c, M 、ガウス基底カーネルの場合は σ がハイパーパラメータとなり、これらの調整次第で得られる分類面は異なる。応用上はソフトマージンカーネル SVM を用いればよく、ソフトマージンのハイパーパラメータとカーネルの設計を変えることで通常の線形 SVM の動きをさせることも可能である。ソフトマージンカーネル SVM の識別関数は以下で表される。

$$y(x) = \sum_{n=1}^N a_n t_n k(x, x_n) + b \quad (2.28)$$

ここで学習データの数 N としている。また最適化問題は以下で定式化される。

$$\begin{aligned} \arg \max_{a_1, \dots, a_N} \quad & \sum_{n=1}^N a_n - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j t_i t_j k(x_i, x_j) \\ \text{s.t.} \quad & 0 \leq a_n \leq C \\ & \sum_{n=1}^N a_n t_n = 0 \end{aligned} \quad (2.29)$$

ここに C はソフトマージンのハイパーパラメータであり、誤分類に対しての厳しさを表す。最適化問題の解である a_1, \dots, a_N と任意の i を選んで以下の条件式に代入することで閾値 b も求まる。

$$t_i \left(\sum_{n=1}^N a_n t_n k(x, x_n) + b \right) = 1 \quad (2.30)$$

ただし、(2.30) は無駄な計算も含まれている。 $a_n = 0$ となるような x_n に対して $\sum_{n=1}^N a_n t_n k(x, x_n)$ の計算を実行する必要はない。 $a_n \neq 0$ となっている x_n のことをサポートベクトルと呼び、実際にはサポートベクトルのみ計算に考慮すれば良い。このことは新規のデータに対して (2.28) の計算を行うときも同様である。従って学習後に保持しておかねければならないデータはサポートベクトルのみに限定でき、実用上省メモリに貢献できる。

2.4.3 Logistic Regression(LR)

3

従来の運動想起型 BCI

3.1 Common Spatial Pattern とその派生手法

運動想起型 BCI に対しては“Common Spatial Pattern(CSP)” [?] が 1999 年に提案されて以降、非常に活発に応用研究がされている。その後、ベンチマークとして引用されてきただけでなく、CSP の拡張手法も数多く提案されてきた。CSP は脳波解析の分野で発展した手法であるため、その定式化も脳波計測時のデータ構造に合わせた形式となっている。そのため、今一度脳波信号の定式化を述べた後に CSP について記す。

3.1.1 脳波信号の定式化

これまで多次元の信号を $x(t) \in \mathbb{R}^D$ と表記し、連続時間信号として扱ってきた。しかし、通常は計測された脳波はコンピュータで処理するために離散時間信号 $x_n \in \mathbb{R}^D$ に変換される。従って、以降、脳波信号を離散時間信号として取り扱う。また、運動想起時の脳波信号を計測する際には、被験者に対して定められたタイムスケジュールで運動想起を行うように指示がなされる。例として 64 個の電極を用い、サンプリング周波数 100Hz で 10 秒間の計測を行った場合、計測された運動想起 1 回分の脳波信号は $X \in \mathbb{R}^{64 \times 1000}$ と表すことができる。従って、以降統一のため、 M を電極の個数、 N を計測時間点数とした場合の脳波信号を以下で定義する

$$X = (x_1, \dots, x_N) \in \mathbb{R}^{M \times N} \quad (3.1)$$

。

運動想起を K 回行った場合には、 K 個の X が得られる。通常は運動想起時の脳波を数個から数十個集め、統計的な指標を元に有用な特徴量を抽出する。CSP は運動想起 BCI に対して極めて有効に働くとされている特徴量抽出手法である。

3.1.2 Common Spatial Pattern(CSP)

CSP を脳波に用いる際は、脳波信号 X を直接扱うのではなく、何らかの前処理を施した信号 $\hat{X} = \mathcal{H}(X)$ を用いる。通常、 \mathcal{H} には、運動想起に関連のある周波数帯域のみを通過させるバンドパスフィルタを用いる。バンドパスフィルタ通過後の脳波信号を以下のように表記する。

$$\hat{X} = (\hat{x}_1, \dots, \hat{x}_N) \in \mathbb{R}^{M \times N} \quad (3.2)$$

$\hat{x}_i \in \mathbb{R}^M$ における添字 i はサンプル時刻の添え字である。CSP では、新たな基底 $w \in \mathbb{R}^M$ にバンドパスフィルタ通過後の脳波信号を射影し、スカラー時間信号である $z = w^T \hat{X} \in \mathbb{R}^N$ を抽出する。この時の w の決め方を以下に記す。

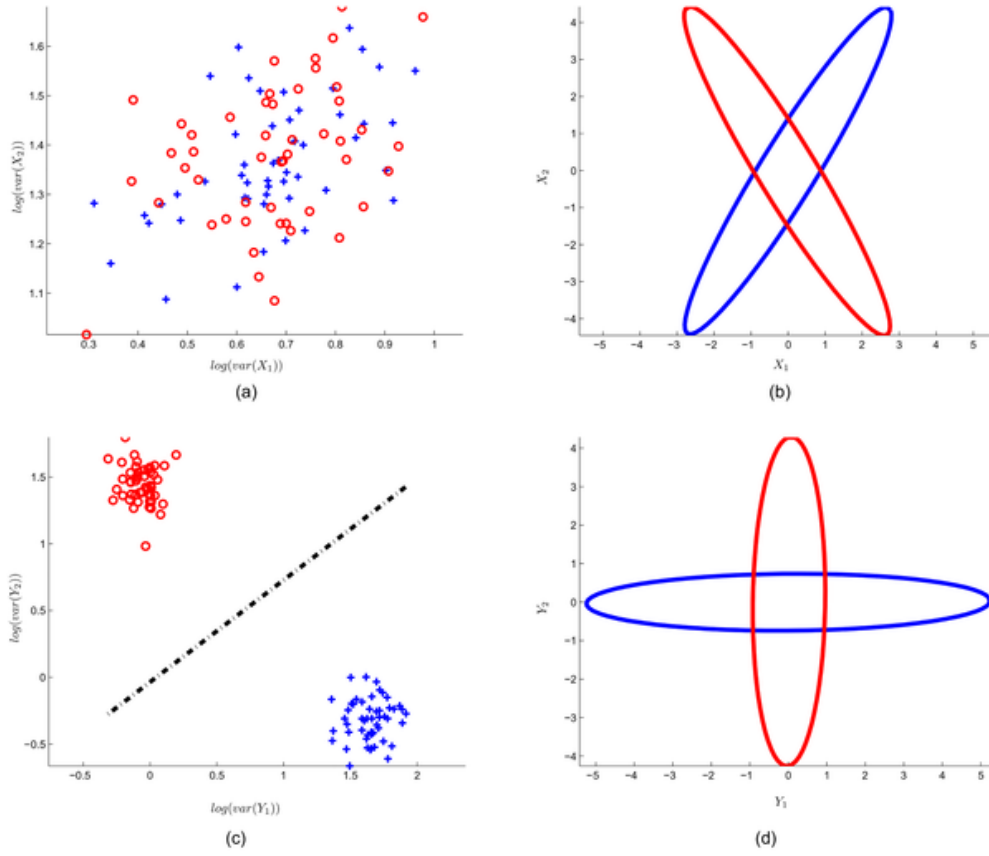


図 3.1: 林橋の図

まず \hat{X} を基底 w に射影した際の時間分散 $\sigma^2(\hat{X}, w)$ を以下で定義する。

$$\sigma^2(\hat{X}, w) = \frac{1}{N} \sum_{i=1}^N \left| w^T \left(\hat{x}_i - \frac{1}{N} \sum_{j=1}^N \hat{x}_j \right) \right|^2 \quad (3.3)$$

ここで計測された複数個の \hat{X} は必ず集合 C_1 か C_2 のいずれか一方に属するとし、 $C_1 \cap C_2 = \phi$ であるとする。CSP では、ベクトル $w \in \mathbb{R}^M$ を新たな基底とした電極空間において、一方のクラスに属する信号 $\hat{X} \in C_d$ についての時間分散 (3.3) が最大となるように、 w を決める。これは以下の最大化問題によって定式化される。

$$\begin{aligned} \max_w \quad & \mathbb{E}_{X \in C_1} [\sigma^2(\hat{X}, w)] \\ \text{s.t.} \quad & \sum_{d=1,2} \mathbb{E}_{X \in C_d} [\sigma^2(\hat{X}, w)] = 1 \end{aligned} \quad (3.4)$$

最大化問題 (3.4) を解いて得られる w は、クラス C_1 に属する脳波の振幅を最大化するような基底となっている。一方で、制約条件によってクラス C_2 に属する脳波の振幅を最小化する基底ともなっている。この最適化問題の解法はよく知られており、標準的な解法は一般化固有値問題を解くことと等価である。計算量の観点から効率の良い解法も提案されている。固有値が分散に、固有ベクトルが新しい基底に対応し、最大固有値に属する固有ベクトルと、最小固有値に属する固有ベクトルを新たな基底として用いる。通常は変換先での時間分散の対数値を特徴量として用いる。以下の図はトイデータを使った CSP による特徴量抽出の様子である。

3.1.3 Common Spatospectral Pattern(CSSSP)

CSP ではバンドパスフィルタ \mathcal{H} 通過後の信号を扱ったため、バンドパスフィルタの設計を終えてからでなければ CSP の問題を解くことができない。しかし、バンドパスフィルタの設計次第では CSP の解が変わることが想定される。そこで、バンドパスフィルタの設計を CSP の問題の中に

取り込んだのが CSSSP [1] である。

まず脳波信号 $X = [x_1, \dots, x_N] \in \mathbb{R}^{M \times N}$ に対して、観測時間点を $\tau > 0$ だけずらした $X_\tau = [x_{1+\tau}, \dots, x_{N+\tau}] \in \mathbb{R}^{M \times N}$ を考える。ここで $\tau = 1, \dots, T$ であり、 T が FIR フィルタの次数の役割を担う。更にクラス C_d に属する X に関して、

$$\Sigma_d^\tau = \mathbb{E}_{X \in C_d} [X X_\tau^T + X_\tau X^T] \quad (3.5)$$

を定義する。ただし、 $\Sigma_d^0 = \mathbb{E}_{X \in C_d} [X X^T]$ とする。FIR フィルタの実数係数ベクトルを $b = [1, b_2, \dots, b_T]^T$ として、CSSSP の最適化問題は以下で表現される。

$$\begin{aligned} \max_{b_2, \dots, b_T} \max_w \quad & w^T \left\{ \sum_{\tau=0}^{T-1} \left(\sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right) \Sigma_1^\tau \right\} w - \frac{\alpha}{T} |b|_1 \\ \text{subject to} \quad & w^T \left[\sum_{\tau=0}^{T-1} \left\{ \sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right\} (\Sigma_1^\tau + \Sigma_2^\tau) \right] w = 1 \end{aligned} \quad (3.6)$$

この最適化問題では空間重みベクトル $w \in \mathbb{R}^M$ と FIR フィルタの実数係数ベクトル b を同時に評価することが可能である。目的関数における第二項は正則化項であり、 α はハイパーパラメータである。この正則化によって FIR フィルタの系列ベクトルに対してスパース性が要請され、FIR フィルタが異常に複雑になることが避けられる。

3.1.4 Filter Bank Common Spatial Pattern(FBCSP)

CSSSP は FIR フィルタの設計を CSP の最適化問題に含むことで、バンドパスフィルタと CSP の同時最適化の定式化に成功した。しかし、CSSSP によって求まる特徴量は 1 つの FIR フィルタを通過し、1 つの CSP によって電極の重み付けが行われた脳波信号にすぎない。クラス分類を行うことが最終目標であることを考えれば、この特徴量は必ずしも最適なものであるとは言い難い。例えば左手の運動想起時と下肢の運動想起時には重要な周波数帯域が異なるためである。すなわち CSSSP によって決定された FIR フィルタは、左手の運動想起時に重要な周波数帯域を捉えている一方で、下肢に重要な帯域を失ってしまっているということが生じうる。FBCSP はこのような問題の解決を図る手法である。

まず、バンドパスフィルタバンク $\{\mathcal{H}_1, \dots, \mathcal{H}_F\}$ を定義する。各バンドパスフィルタを通過した $\hat{X}_f = \mathcal{H}_f(X)$ に対して、それぞれ CSP の問題を解くことで、複数の周波数帯域を反映した特徴量の抽出が可能となる。しかし、この段階においては CSSSP とは異なり、CSP とバンドパスフィルタの同時最適化を行ってはいないため、初めに定義したバンドパスフィルタバンクの CSP 特徴量を対等に扱うことが妥当とは言えない。FBCSP では各バンドパスフィルタに対しての CSP の解から、更に重要な特徴量の選定を行う方法までを含んだ提案がなされている [2]。

3.2 従来手法のまとめ

ここまで運動想起 BCI を構築するための特徴量抽出や分類器について述べてきた。具体的に運動想起 BCI を構築する際には、様々な手法を適宜組み合わせ、目的の性能を向上するために試行錯誤を重ねなければならない。その際に考慮しなければならない脳波に関する知見を以下に記す。

- 運動想起する身体部位に応じて、強く反応する頭皮領域は異なる。
- 事象関連脱同期は、特定の周波数領域に生ずる。
- 個人差や計測環境の影響を受けやすいため、ロバストな設計が必要となる。

上記の項目を考慮した上で運動想起 BCI を構築するために以下の施策を取る。

- 利用する電極の選定、あるいは空間フィルタなどによる電極の重み付けを行う。

- バンドパスフィルタによって特定の周波数波形のみを取り出す。
- 汎化性能の高い分類手法を用いる。

この章では、これらの施策を具体的に実行した BCI 構築の手順例について紹介・検討し、そこに生ずると考えられる問題点を把握し、研究のモチベーションを明らかにする。

3.2.1 従来の運動想起 BCI の構成と問題点

まず典型的な運動想起 BCI の構成について述べる。運動想起を 1 回行った際に計測された脳波を以下で表記する。

$$X = (x_1, \dots, x_N) \in \mathbb{R}^{M \times N} \quad (3.7)$$

ここに M は電極の個数、 N は計測時間点数である。典型的な運動想起 BCI は、運動想起に関連している周波数帯域を取り出すためにバンドパスフィルタ \mathcal{H} によって脳波の生データから $\hat{X}(\cdot)$ を獲得することが一般的である。

$$\hat{X} = \mathcal{H}(X) \quad (3.8)$$

次に、運動想起に関連している電極を選定するための空間フィルタ $f(\cdot)$ を適用する。

$$Z = f(\hat{X}) \quad (3.9)$$

続いて重要な周波数帯域と頭皮領域について考慮がなされている Z に対して、運動想起部位 Y を出力する分類器 $g(\cdot)$ を準備することで、運動想起 BCI が構成されている。

$$Y = g(Z) \quad (3.10)$$

このようにして構築された BCI はバンドパスフィルタ \mathcal{H} 、空間フィルタ f 、分類器 g を要素とした集合として記述することができる。

$$BCI = \{\mathcal{H}, f, g\} \quad (3.11)$$

また、時間周波数解析や行列分解・テンソル分解によって特徴量抽出を行う手法も提案されている。しかし、BCI を構築する上では（あるいは脳波解析をする上では）周波数帯域と頭皮領域の考慮は欠かせないことであり、バンドパスフィルタや空間フィルタは一般的に用いられる。従って独自の特徴量抽出手法を $h(\cdot)$ とすれば、以下のように典型的な BCI に要素を追加した形で表記することができる。

$$BCI = \{\mathcal{H}, f, g, h\} \quad (3.12)$$

以上の観点から、BCI を設計することは集合 (3.12) の要素をどのように決定するかを考えるということに相当する。最も神経科学的知見に忠実な運動想起 BCI は、バンドパスフィルタ \mathcal{H} を事象関連脱同期の周波数帯域によって決定し、空間フィルタ f を身体部位と対応する脳領域によって構成し、処理された信号に対して閾値を設けることで分類器 g を構築する。機械学習の手法を取り入れた代表的な BCI はバンドパスフィルタ \mathcal{H} を試行錯誤、あるいは何らかの評価指標に基づいて決定し、空間フィルタ f を CSP によって構成し、処理された信号に対して LDA などを用いて分類器 g を構築している。更に CSSSP はバンドパスフィルタ \mathcal{H} と空間フィルタ f を同時に考慮した同時最適化手法、FBCSP はバンドパスフィルタ \mathcal{H} を複数用意し、複数の周波数帯域から特徴量抽出を行う手法である。

従来手法に則って BCI を設計することの問題点を検討する。そのために典型的なワークフローの概略を示す。まず、予め (3.12) の要素を定める。具体的にはバンドパスフィルタ \mathcal{H} をバターワースフィルタ、空間フィルタ f を CSP、分類器 g を LDA などと定めることである。この時に検討される BCI を以下で表記する。

$$BCI = \{\mathcal{H}_{butterd}, f_{csp}, g_{lda}\} \quad (3.13)$$

ここで (3.13) の各要素も実数スカラー、実数ベクトル、実数行列などを要素に持つ集合であり、以下のように表される。

$$\mathcal{H}_{butterd} = \{w_p, w_s, r_p, r_s \mid w_p, w_s \in \mathbb{R}^2, r_p, r_s \in \mathbb{R}\} \quad (3.14)$$

$$f_{csp} = \{W_{csp} \mid W_{csp} \in \mathbb{R}^{m \times M}\} \quad (3.15)$$

$$g_{lda} = \{W_{lda} \mid W_{lda} \in \mathbb{R}^{d \times m}\} \quad (3.16)$$

w_p, w_s はそれぞれ通過帯域コーナー周波数、阻止帯域コーナー周波数である。 r_p, r_s はそれぞれ通過帯域リップル、阻止帯域の減衰量である。 W_{csp} は CSP における線形変換の表現行列であり、電極数 M の次元を持つ脳波を $m (< M)$ の任意の次元の電極空間に圧縮する。 W_{lda} は LDA における線形変換の表現行列であり、 m 次元の電極空間から分類クラスの数より小さな任意の d 次元への射影を担う。

改めて (3.13) についてその要素を書き下すと以下のように記述できる。

$$BCI = \{\{w_p, w_s, r_p, r_s\}, \{W_{csp}\}, \{W_{lda}\}\} \quad (3.17)$$

すなわちここで検討している BCI は (3.17) の要素を決定することによって構築され、第 1 の要素に含まれるパラメータがバンドパスフィルタを決定し、第 2 の要素に含まれるパラメータが空間フィルタを決定し、第 3 の要素に含まれるパラメータが分類器を決定する。脳波処理を適切に行うための手法を柔軟に検討できるかのように見えるが、ここで考慮した BCI には実際には明らかな制限が存在する。CSP のパラメータの決定方法については 3.1.2 で述べ、同様に LDA のパラメータの決定方法については 2.4.1 で述べた。いずれも解析的に決定されるパラメータであることを考えると、事実上、(3.17) において自由に決定できるパラメータは第 1 の要素に含まれているものだけであり、第 2 の要素と第 3 の要素は第 1 の要素を決めた時点で（計算の必要性は生じるが）決定されている。唯一、CSP と LDA における写像の次元についてだけは検討できるが、これらは最適化問題を解いた後、いくつかの次元を無視することで達成されるのであって、本質的に最適化問題の振る舞いが変わることはない。従って、ここまで検討してきた BCI に関してはバンドパスフィルタの設計に極めて敏感に性能が左右されると想定される。他方で、考慮しなければならないパラメータが極めて少ないために設計が容易であると捉えることも可能であるが、CSSSP や FBCSP などがここで検討した BCI よりも高い性能を示していることから、より大きな設計自由度を与える必要があると言える。

4

提案手法

4.1 ニューラルネットワーク

4.1.1 フィードフォワードニューラルネットワーク

特徴量工学は BCI に限らず、データ解析を必要とする多くの分野で活用されている。例えば画像処理の分野では、画像に写っている物体の輪郭を抽出するエッジ処理や、画像に対してぼかしを入れる平滑化処理などが特徴量抽出手法として用いられる。また音声認識の分野では、音声の周波数に重要な情報が含まれていると考え、周波数スペクトルやケプストラムなどを特徴量として用いる試みがなされてきた。

4.1.2 リカレントニューラルネットワーク

4.1.3 畳み込みニューラルネットワーク

4.2 FIR フィルタと畳み込み層

5

実験と結果

6

結論と考察

Appendix

A.1 セクション

A.1.1 サブセクション

謝辭

... ..

参考文献

- [1] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, Klaus-Robert Müller, “Combined optimization of spatial and temporal filters for improving Brain-Computer Interfacing ”, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, Vol. XX, No. Y, 2006.
- [2] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, Cuntai Guan “Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface”, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, in *Proc. 2008 Int. Joint Conf. Neural Netw.*, pp. 2390-2397, 2008.
- [3] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggenberger, Michael Tangermann, Frank Hutter, Wolfram Burgard, “Deep learning with convolutional neural networks for EEG decoding and visualization”, *Human Brain Mapping*, Vol. 38, No. 11, pp. 5391-5420, 2017.