

Lecture 4: Adaptivity and Machine Learning

Introduction to Modern Brain-Computer Interface Design

> Christian A. Kothe SCCN, UCSD



Outline

- 1. Adaptivity in BCIs
- 2. Machine Learning
- 3. Concrete Case Study
- 4. Performance Evaluation





4.1 Adaptivity in BCIs



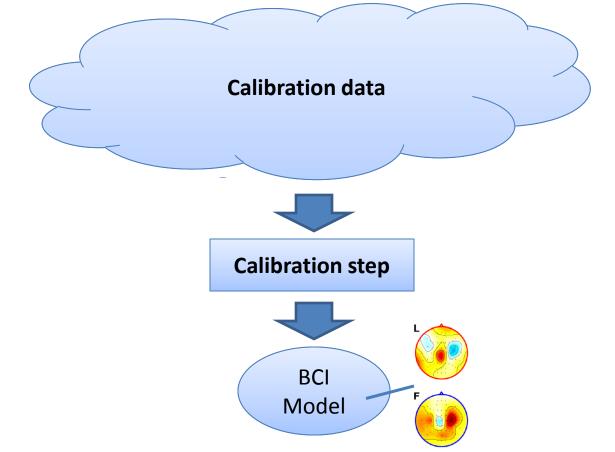
Unknown Parameters

- for most BCI questions and implementations, the parameters leading to best accuracy (W,b, ...) are a priori unknown...
 - Depend on highly variable factors (e.g., sensor placement, subject state)
 - Different for every person, task, montage, etc.
 - Depend on hard-to-measure factors (e.g., brain functional map)
 - Depend on expensive-to-measure factors (e.g., brain folding)
- How to solve this problem?



Model Calibration

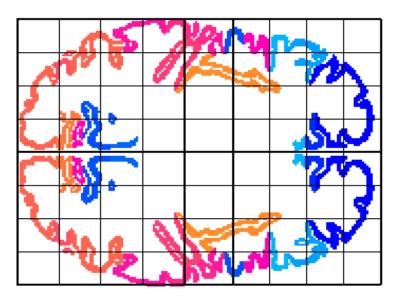
• Can use *calibration / training data* to estimate parameters from, and a separate *calibration step*





Prior Knowledge

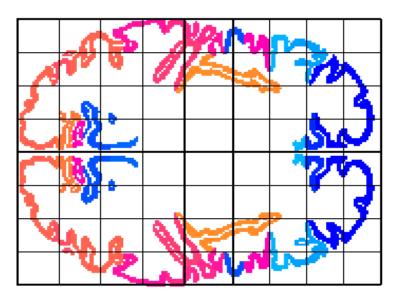
- Prior knowledge is neuroscientific, such as:
 - Anatomical atlases
 (e.g. Talairach, LONI)
 - Functional atlases (if available)





Prior Knowledge

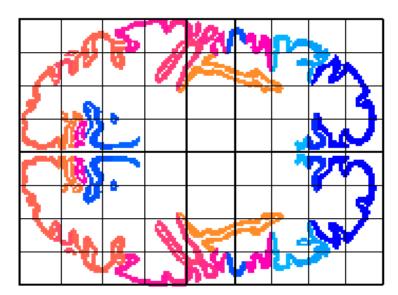
- Prior knowledge is neuroscientific, such as:
 - Anatomical atlases
 (e.g. Talairach, LONI)
 - Functional atlases (if available)
 - Timing information (e.g. neural latencies, reaction times)

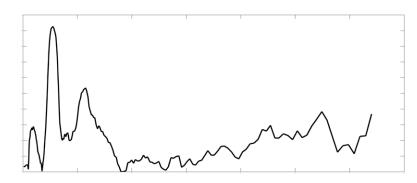




Prior Knowledge

- Prior knowledge is neuroscientific, such as:
 - Anatomical atlases
 (e.g. Talairach, LONI)
 - Functional atlases (if available)
 - Timing information (e.g. neural latencies, reaction times)
 - Frequency bands of oscillatory processes (alpha, beta, theta, ...)







Calibration Data

• Example/calibration data is used to calculate optimal parameters of a BCI, and is *extremely important*

Fp1	French and a start of the start	. Areast de	half noorhoutherman	Marin man Maran	land a state of the second of the second	www.waterapater	ᠵ <i>᠁᠕ᢦᠵᡑᠽᡗᠯᢂᠰᢦᡐ</i> ᡁᠵᡧᠮᢍᢦ	walka walawa ye va hay	and Malanananan L.	a sealer and a sealer	Manuf Manufacture
	composition and	Marine Million Start	Mark Bark Marker	Martine	outhouse the second	and the second second	warman water and the	a contraction of the set	man and the second of the	a manufacture and	endered and read
F3 Fz	in the second	An and a start and a start of the start of t	ale and many market	in a company war	and the second s	montenation	and the second second	and the second and the	month a strange when a strange when	munumun	mer manung mer
F4		marine and the second									
C5		the stand when the st					COLORAN LASS	un and a state of the state of			
-C3 -C1	Same and	Man Allan and and and and and and and and and a	and the second sec	with the state of	the second s	and the second		and the second	munder marine	the second will when a second	mound
=Čz			the most have a carden	Maria Maria	and an an an and an and an a start	We charling and the	and the second second	a subserve and a second	martine White Analytics	and the second	and a second and the second
FC2	a manuna	A CONTRACTOR OF THE OWNER	the state of the s	a de antine and a state	and a provide the state	and we have been and the second	when a when when	water water and a second	and the second states	have the second man and	water the and prover the
=C4 =C6	District Annual State		NO DIAL CALLOUT	chineses at the	r and a construction of the	try 2000 and 2000	and the second second	Mar William Mar	and a start water and the start water and	mary warman	WINNERSONAVANTE
C5	american	- MANANA ANA ANA ANA ANA ANA ANA ANA ANA	Are What was a provide	an and a fair a fair an							VANAMANA
C3	munung	Winner with a lot of the second	Man Harren Jaren Wy workers	and a contract of the second	in the man age to	my strangent	and support the second second	and and the second of the seco	where the address of the second second	many han all an arth han have a gang	a new and the market of
C1 Cz	a water and	www.handhall.	The second strates and a second	and the second	and the second second with the second	and the state of the second second	when he have	www. here we wanted	www. Wine All man way when man	And the second when the second	manan
C2	and the stand	Walterhand - Mr. March 1999 . and	the alman	anti-advisional analysis	٢٠٠٠ ٣٠٠ ٢٠٠٠ ٢٠٠٠ ٢٠٠٠ ٢٠٠٠ ٢٠٠٠ ٢٠٠٠	al a patra a se la	Marks months there		hursday When we want the	the and the share with the	warman many met
C4 C6	External for the same	A REAL PROPERTY AND A REAL	Metable millians	المعالم الشغية المعلمات المعلم	Mary Sochand Million they	hout the make the	White Automatical and	we wanted the state	attended in a subar being	will be a call of an an	UNERFANCE AND ALLER A
CB CP5	2 minutes and	PARTICIPATION OF THE PARTY	MAN AND AND AND AND AND AND AND AND AND A	and it is the shirt of the	AN INVESTIGATION OF A	of the state of the state of	COLUMN STATISTICS	ADDIL MARK SACHAN MARCH	Adding to the second second	and the second of the	ADMAR THE POINT OF PROT
CP3	Same an and	Man man Jan Martin	man a ser way way way	hand a hour harder	مريك مركز المعالم معادية المركز ا مريك مركز المركز الم	a Manada and		White the Area and a second	the she down to she when the	and the stand have	www.manner
CP1 CPz	an and a start with	And the second with the second with the	prover and a survey of some	WWW. we want and a start of the	man the server where the street was	and a spectral and a second	and the second and the second	man and a second and a second se	m the Mune was shown	and and and and a series when	Northe Hard Carles Secondar - 10-
	E	and a second and a second with the	there and the second	WWWWWWWWWWWWWWWWWW	فالمان والمحمد المراجع والمصحر والمحمرين	14 D. T. A.	A Received and a second s				
CP4	and the second	enter danie	TANAPART		P A Photo A Marcha	How WHITH MARKER AND	with the first states	paralipe de l'ale de parte de	the month of the state of the	AT THE PART STORE	where many shares
P6 P3	and a stand and a stand	THE PERSON NET 1 1919 1 1 1	M919. La si i i i i	hand in A source of the new	undersidence Auch	a shara and a m	where a same note	leng du an ar men aread	an all then the same	and the standard stand	where the state of
Pz	a multiple and	KAN AREN BARRAN	and Manual and	hall when a will	Contraction of the second s	h white the second	Marken Williams Mary	water an entropy of the	CALL AND	and the same share to an include	
P4	Procession and a start	A Company of the second s	The state of the second	Autor and served	Marine Auro	the stand and a second	and some frances	and the second sec	and the state where	Ner all and the property of the	and the second second
0z	-	internet internetity	Mar Markel barrow	Americaninerant	Annon With a Maria Mar	How we want the second	Adams, and the finder of the	Viewallowshipstoneyrandy	mal Protitional where	Mush which when the starter	unonal with some de
01 02	-	http://www.http://www.hellen.en.		The second states and	here secondary mente		and and a set	the second being	a da a la constantia da a d	a anna a'	-
OGh	F										-
OGv MGI	F										1
MGr	F										-
4	5		46		47		48		49		50



- Collected with the same/similar measurement apparatus as used for online runs
 - otherwise extra transformations and uncertainty incurred
- Comprises multiple independent realizations / repetitions / trials (to quantify variability)
 - one-shot learning (one exemplar) is *much* harder



- Collected under conditions that are as close to those of the online runs as possible (i.e., drawn from the same statistical distribution)
 - Same person is preferable
 - Same sensor arrangement is preferable
 - Same session is preferable
 - Task parameters (stress level, ...) should be similar
- Obviously a cost/benefit tradeoff:
 - Would trade off some performance for being able to reuse one recording for multiple sessions and persons



- If there is a bias (e.g., different session), data should cover multiple realizations (e.g., multiple sessions) to capture variability
- A plain EEG recording is "unlabeled" (no knowledge about the association between raw observed signal and the cognitive state variable of interest)
- Labeled data (person is "surprised" / "not surprised") is *far* more useful than unlabeled



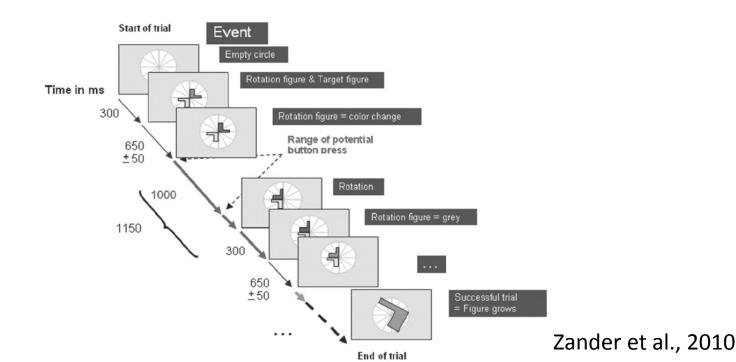
• Labels are assigned per realization (e.g., per trial) and *index the output that the BCI shall produce for this class of data*

En1	www.www.	manual and manual	es priving to far with an of prover station of the second states and	www.water	and an address of the second and the	and and the second second second second second
Fp1 Fp2 F3 Fz	connegatived	New York Control of the second s		· · · · · ·		
E3	- more mark	the second second second second	an a	a share a far a share a	and the second	a the second sec
Fz F4	Ann and a set			and a state of the second	and the second	
FC5	Participation of the second se	bland and a state of the state			and the second second second	
FC3	Same and		and the second			Man was a second s
FC1 FCz	Samana		and a second second constrained and a second s	and be a second and	the second and the second s	and the second s
FC2	and many	manufacture and a manufacture and	alpertananter and a set of a s	- and a second rest of the second of the sec	and the second se	When the second and the second s
FC4 FC6	Total providence of	and the second	the Advistment of Advisor With the average and the set	www.allerander.		A man an a
C5	a monormust		and the second se	and the second	and when we have a second of the second of the second	man with the second and the second and the second of
Č3	minimum	and the start of t	angenet server a lander of the server a strong and the server and	and the second states a	WANNE ZHU UNINGANI WANGAN KANGA UNING WANGUNANGAN UNING ANG UNING UNING UNING UNING UNING UNING UNING UNING UN UNING NA UNING U UNING UNING UNI	¹ Maria (1977-1964) (kathar 1967) (kathar 1977) Maria (kathar 1977) (kathar 1977) (kathar 1977) (kathar 1977) (kathar 1977)
C1 Cz	and a start way	ALL A LOCAL AND	and a hard a product to the state of the second state of the second states and the second states and the second	the start of the s	water water water and a strategy of the second strategy and the second strategy and the second strategy and the	the second and a second second second
C2	- martine	and the second	and a second second and a second s	with a relation in a la shakka handhar Mar	and a second and a s I a second a I a second a second I a second a second I a second a second I a second a I a second a s	allow and and and an arrange and and
C1	-	warmen to bland the short the ball of a set	eder, fand kalt i skrift af ta de derekt instrument after og	the particulation in the data of the particulation of the	in the second	beland by a share of a second share of the second states of the second s
C6 CP5	1 1 1 1 2	A HIT WAS A THINK I WAS A THINK OF THE REAL OF	and the second	Karlina a karolina a karalar a karalar kar	address and the second and the second	
CP3	Lange Marine	the survey of the second s	Manhad have been and the transformed and the	have been and a superior of the second second	with the reason and the when the water the	month have here have me and an and
CP1	manning	Martin all all and an and martin	All and a second the second se	have been a second and a second a		and a second and a
CPz CP2	Frank March	and a second with the second have been a free to be a second with the second second second second second second	man some and some man and and and and and and and and and a	way a gran and a south and the second	months non mu and have been been at the set of	
CP4	Sincer Hard		Part Shines	athe of the Manual and a well the Manual The State of the out		
CP6	Same Stan and	the state of the s	All a finally showing been been by the set from a		an a sha a da ar a a sha ar a sha ar a sha a	The set of
P3 Pz	1 mul north and	the second show the second show here have	an a	1 st. Wheeling and the second statements of the	and a sub-the second and the second	and a second where the second second
P4	Reserved	and the second second second second second second	Alter and the second	alle a stand and a stand a		A AMAL AN AND AN ANALY ALLAND AND AN AND AND AND AND AND AND AND A
-Oz	-	when the structure with an author because a	and an an an and a second a standard the states, the	a set working the second we define a set we behave	Mandeter Mander of payres Andrew Apart Harry Argenting	a late a selection of the selection of the second s
01 02	- International terration	Printing a participation of the second second	a a standard model the last to be a start of the second to a	Maranda i puli Ali danan Matada adalar a stala i an isa	a same stress the salid finds of a loss brade while and should be the sales	and a real of the second of the second of the second s
OGh	-					
OGv	E					
MGr	F					
	15	46		48	19	50
_		40	41	40	40	50



Summary

 The required data to calibrate a BCI resembles data produced by *controlled psychological experiments*





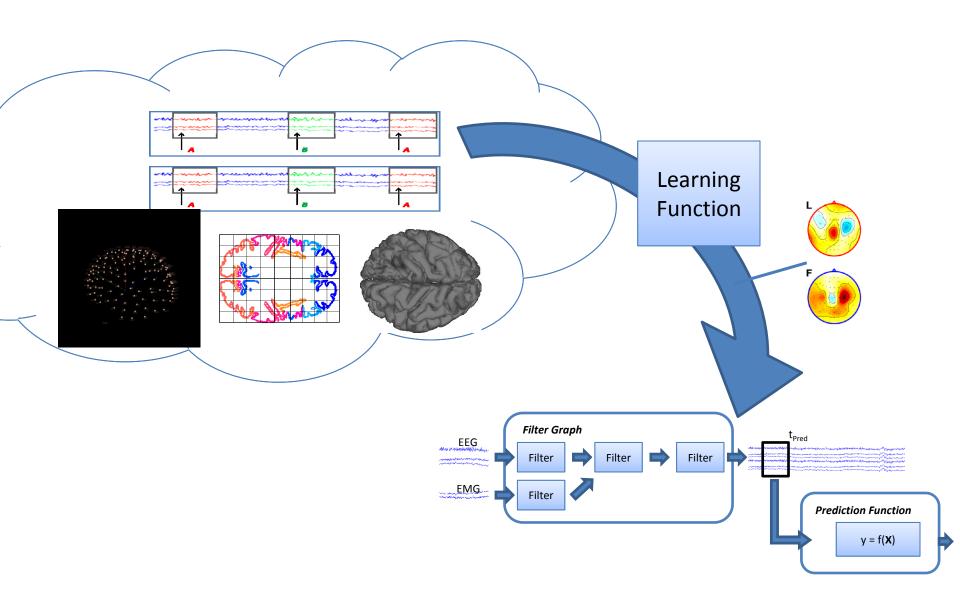
Summary

- Features
 - continuous EEG (or other)
 - multiple trials/blocks (capturing variation)
 - randomized (eliminating confounds)
 - event markers to encode cognitive state conditions of interest, e.g., stimuli/responses (called *"target markers"* in BCILAB)
- Can also be used for offline performance tests

and man man man have been and the second sec						
			\uparrow			
₅₂	_{S1}	R1	S1			



Big Picture







4.2 Machine Learning



Machine Learning

- Large field with 100s of algorithms
- Most methods conform to a common framework of a *training function* and a *prediction function*



- Intermediate model parameters $\boldsymbol{\theta}$ capture the learned relationship
- Data $X \in \mathbb{R}^{N \times F}$ and Labels / target values $y \in \mathbb{R}^{N \times D}$ N = #trials, F = #features, D = #output dims.



Machine Learning

- The Machine Learning Framework is largely trial-based (learning from exemplars)
- Most methods come in form of two functions: *learning function* and *prediction function*
- Learning function is often *far* more complex than the prediction function



Sub-Types In ML

- Supervised Learning: given a set of (input,output) pairs as training data, learn a parametric (or "non-parametric") model M that encodes the mapping from input to output
- Unsupervised Learning: given a set of training examples, learn the structure in the input space (e.g. clusters, manifolds, probability density)
- Semi-Supervised Learning: Some training examples have labels, others do not
- **Others:** e.g., Active Learning, Online Learning, ...



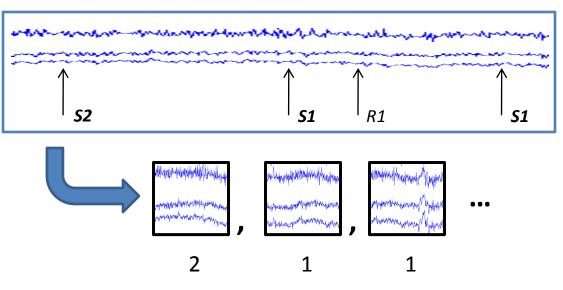
Related Areas

- Probability Theory
- Statistics
- Optimization
- Neural Networks
- Artificial Intelligence, ...



Using Machine Learning

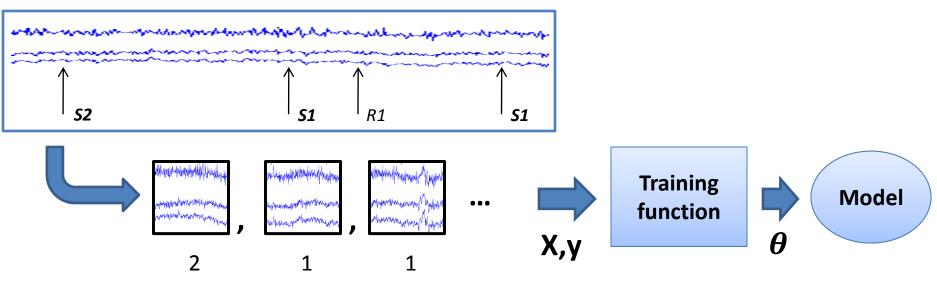
- Often, one trial segment (sample) is extracted for every target marker in the calibration recording and is used as *training exemplar* X_k
- Its associated label y_k can be deduced from the target marker





Using Machine Learning

- The training function computes a parameter (here θ) of the prediction function such that the performance on the given example data is optimal
- What is considered optimal depends on *extra assumptions* (a.k.a. *priors*)

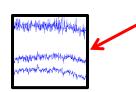




Detour: Feature Extraction

- **Caveat:** Off-the-shelf machine learning methods often do not work very well when applied to raw signal segments of the calibration recording
 - too high-dimensional (too many parameters to fit)
 - too complex structure to be captured (too much modeling freedom)

1000s of degrees of freedom!





Detour: Feature Extraction

- Solution: Introduce additional mapping (called *"feature extraction")* from raw signal segments onto feature vectors
 - output is often of lower dimensionality
 - hopefully statistically "better" distributed (easier to handle for machine learning)



Examples of Feature Extraction

- Depends on the process of interest (e.g. oscillation, ERP peak, complex phenomenon)
- For oscillations, e.g.:
 - log-Variance (logarithm yields more convenient data distributions)
 - Part of the Fourier spectrum
- For ERPs, e.g.:
 - Peak latency, height, width (artificial example)
 - Mean in one or more time ranges relative to an event
 - Subset of Wavelet coefficients





4.3 Concrete Case Study



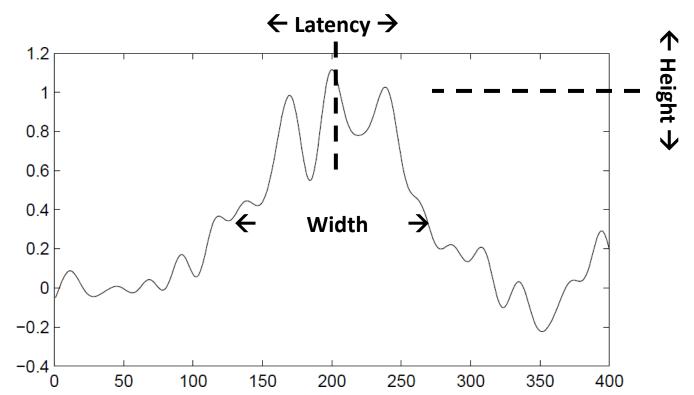
Example Calibration Problem

- **Task:** A person is presented with a sequence of 300 images (one ever 2 seconds). Half of the images are exciting, the other half are not. One channel of EEG (at Cz location) is recorded.
- Question: How to design a BCI that can determine whether a person is shown an exciting or a non-exciting image?
- Approach: For each trial k, cut out an epoch X_k of 1s length, extract a short vector of features f_k, and assign a label y_k in {E,NE}. Use machine learning to find an optimal statistical mapping from f_k onto y_k.



Extracting Features of a Peak

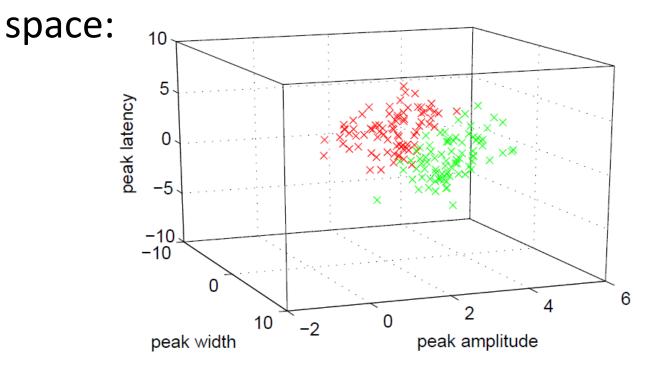
 A supposed characteristic peak in a time window (relative to an event) could be characterized by three parameters:





Resulting Feature Space

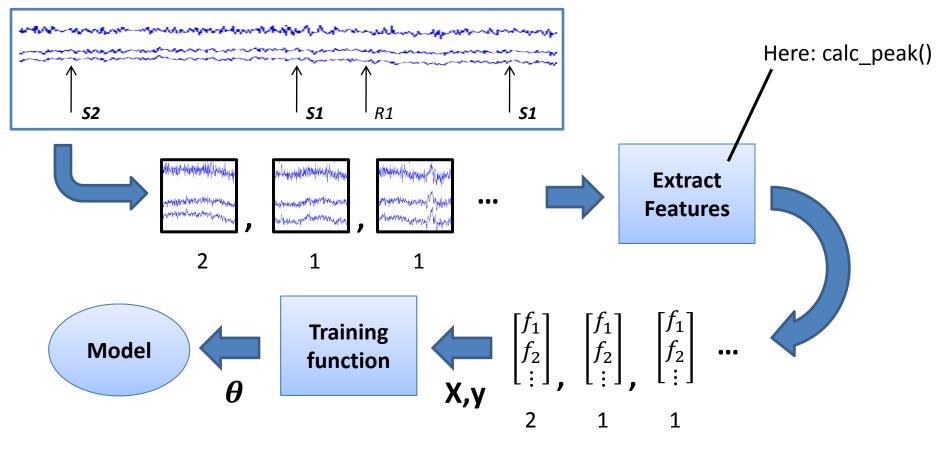
 Plotting the 3-element feature vectors for all exciting trials in red, and non-exciting trials in green, we obtain two distributions in a 3d





ML with Feature Extraction

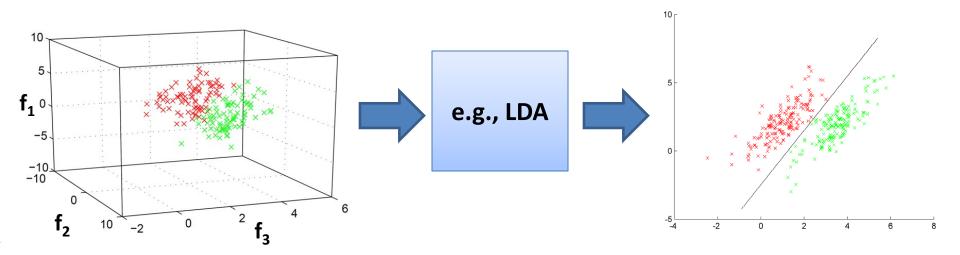
• Including the feature extraction, the analysis process is as follows:





Using Machine Learning

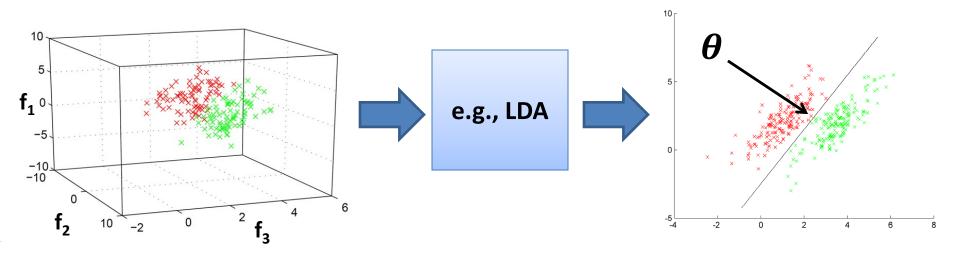
• The feature vectors are passed on to a machine learning function (e.g., Linear Discriminant Analysis)





Using Machine Learning

- The feature vectors are passed on to a machine learning function (e.g., Linear Discriminant Analysis)
- ... which determines a parametric predictive mapping



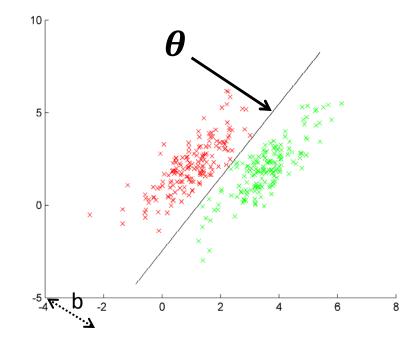


LDA In a Nutshell

• Given trial segments x_k (in vector form) in \mathcal{C}_1 and \mathcal{C}_2 ,

$$\boldsymbol{\mu}_i = \frac{1}{|\mathcal{C}_i|} \sum_{k \in \mathcal{C}_i} \boldsymbol{x}_k, \qquad \Sigma_i = \sum_{k \in \mathcal{C}_i} (\boldsymbol{x}_k - \boldsymbol{\mu}_i) (\boldsymbol{x}_k - \boldsymbol{\mu}_i)^{\mathsf{T}}$$

$$\boldsymbol{\theta} = (\Sigma_1 + \Sigma_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \qquad \mathbf{b} = -\boldsymbol{\theta}^{\mathsf{T}} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$$





Resulting Predictive Map

• LDA generates parameters of a linear mapping

$$y = \theta x + b$$

For classification, the mapping is actually *non-linear*:

$$y = sign(\theta x + b)$$



More on LDA

• Assumptions:

- Data in each class is distributed according to a Gaussian distribution
- Shape of the distribution is identical for all classes

• Benefits:

Simple, fast and optimal in the large-sample limit if assumptions are true

• Problems:

- Very sensitive to outliers (non-Gaussian)
- Covariance matrix estimates become unreliable / unusable for too few trials and too many dimensions



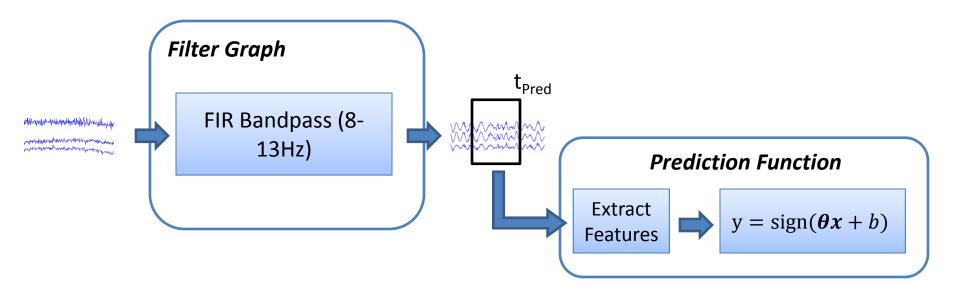
Putting it Together

- 1. Apply band-pass filter to calib. recording
- 2. Extract epochs relative to target markers
- 3. Extract features for each epoch (here: peaks)
- Submit all feature vectors & target labels to LDA to calculate *θ* and *b* of the predictive mapping



Putting it Together

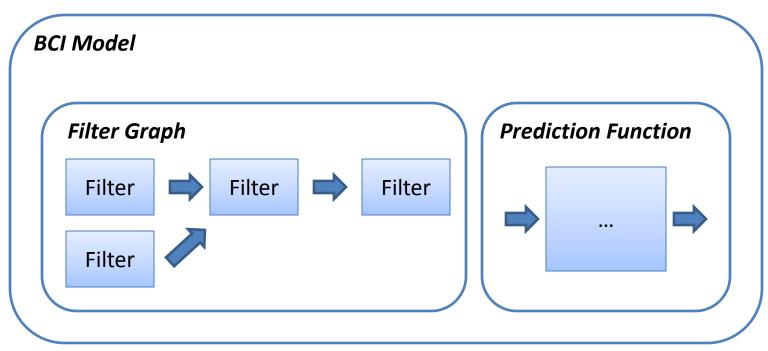
- For online operation, the overall prediction function consists of the feature extraction followed by the predictive map
- This yield a primitive "excitement detector"





BCI Models

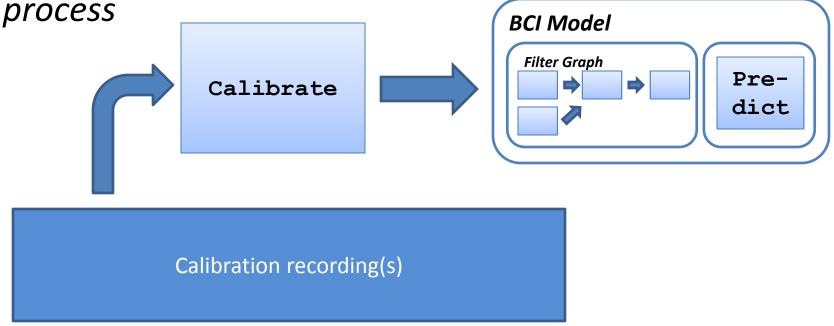
- BCIs are described by "BCI models" that specify both the *filter graph* and the *prediction function* (incl. its parameters)
- These models are the result of the calibration step





"BCI Paradigms"

- BCI paradigms are a notion that was first developed in the BCILAB framework
- A BCI paradigm is the *full description and codification* of a particular type of calibration and prediction



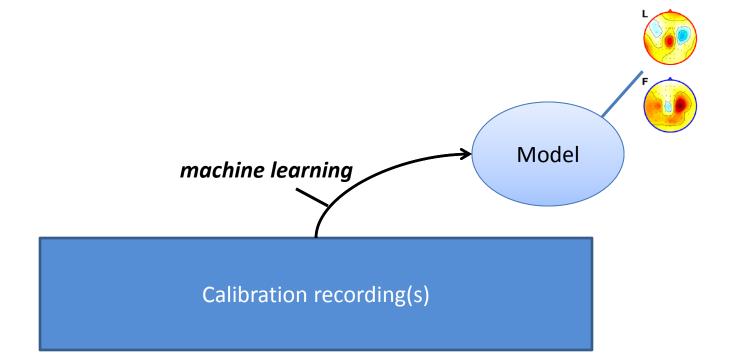




4.4 Performance Evaluation

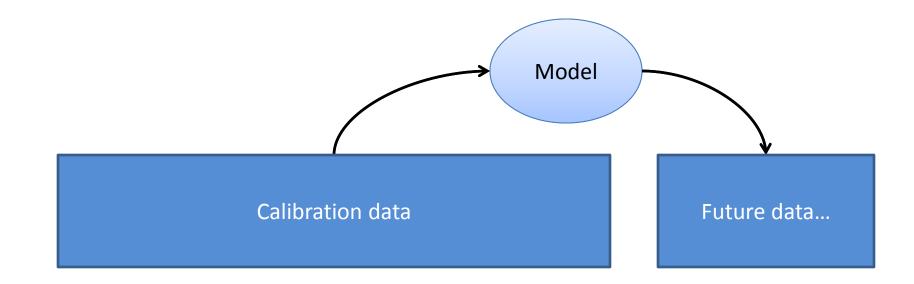


How to Evaluate a Model?





- When given calibration data and test data...
- Estimate model parameters (spatial filters, statistics)
- Apply the model to new data (online / single-trial)
- Measure prediction performance or loss





 Note: Overall *loss estimate* between a vector of predictions *p* and a vector of targets *t* is a summary statistic

• Mean-Square Error:

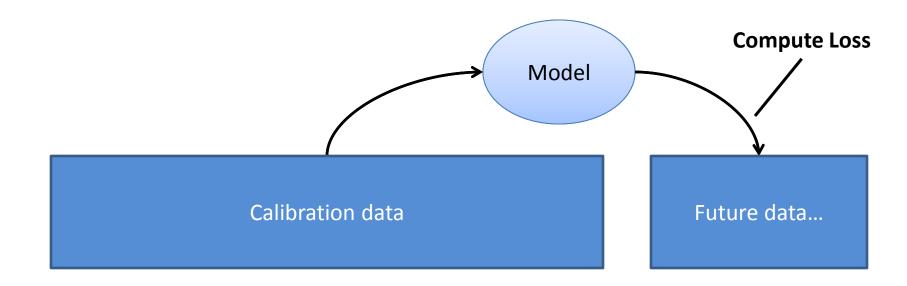
$$-L_{MSE}(\boldsymbol{p},\boldsymbol{t}) = \frac{1}{N}\sum_{k}(\boldsymbol{p}_{k}-\boldsymbol{t}_{k})^{2}$$

• Mis-Classification Rate:

$$-L_{MCR}(\boldsymbol{p},\boldsymbol{t}) = \frac{1}{N} \sum_{k} \begin{cases} 1, \boldsymbol{p}_{k} \neq \boldsymbol{t}_{k} \\ 0, \boldsymbol{p}_{k} = \boldsymbol{t}_{k} \end{cases}$$

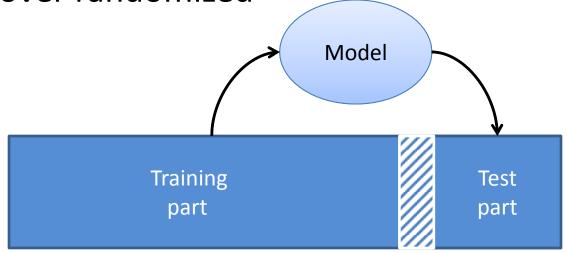


• What if there is no second data set?



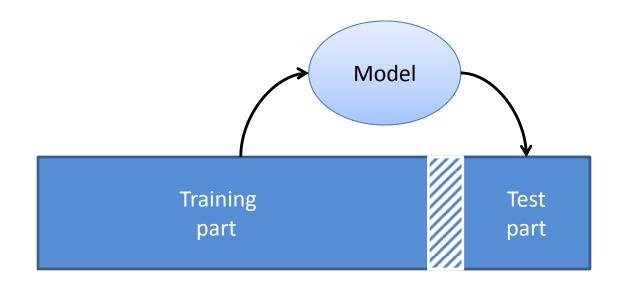


- Alternative: split one data set repeatedly into training/test blocks systematically, a.k.a. cross-validation
- Each trial is used for testing once
- Time series data: Prefer block-wise cross-validation over randomized



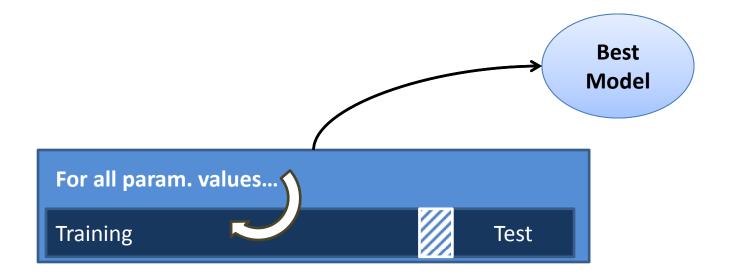


- Consideration: Since neighboring trials are more closely related than training and future online data, *leave a margin of several trials/seconds between training and test*
- Standard splitting schemes: 5x, 10x



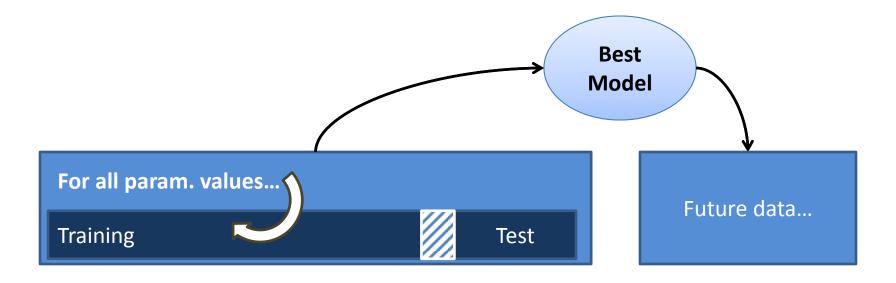


- **Parameter search** can be done using cross-validation in a grid search (try all values of free parameters)
- Quite general (e.g. can search for best method)



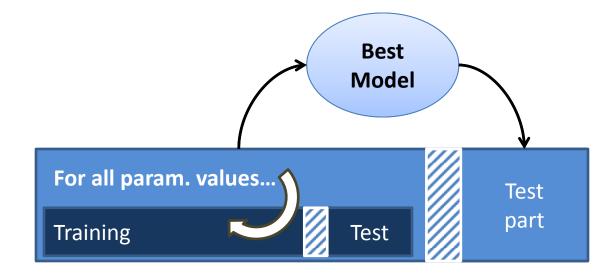


- **Parameter search** can be done using cross-validation in a grid search (try all values of free parameters)
- Quite general (e.g. can search for best method)
- However: Cannot directly report "best performance" estimates (=cherry-picked), except on future data





- **Parameter search** can be done using cross-validation in a grid search (try all values of free parameters)
- Alternatively: Parameter search can be nested within an outer cross-validation ("nested cross-validation")







L4 Questions?