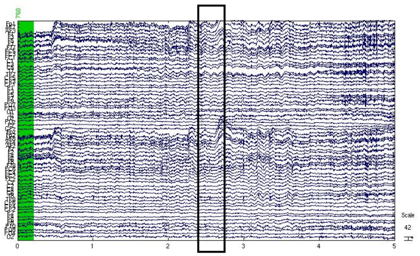
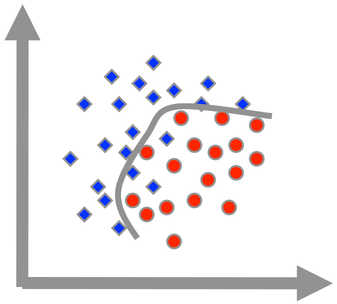
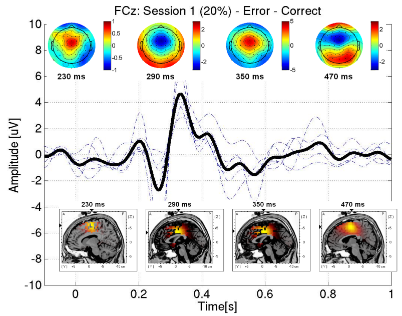




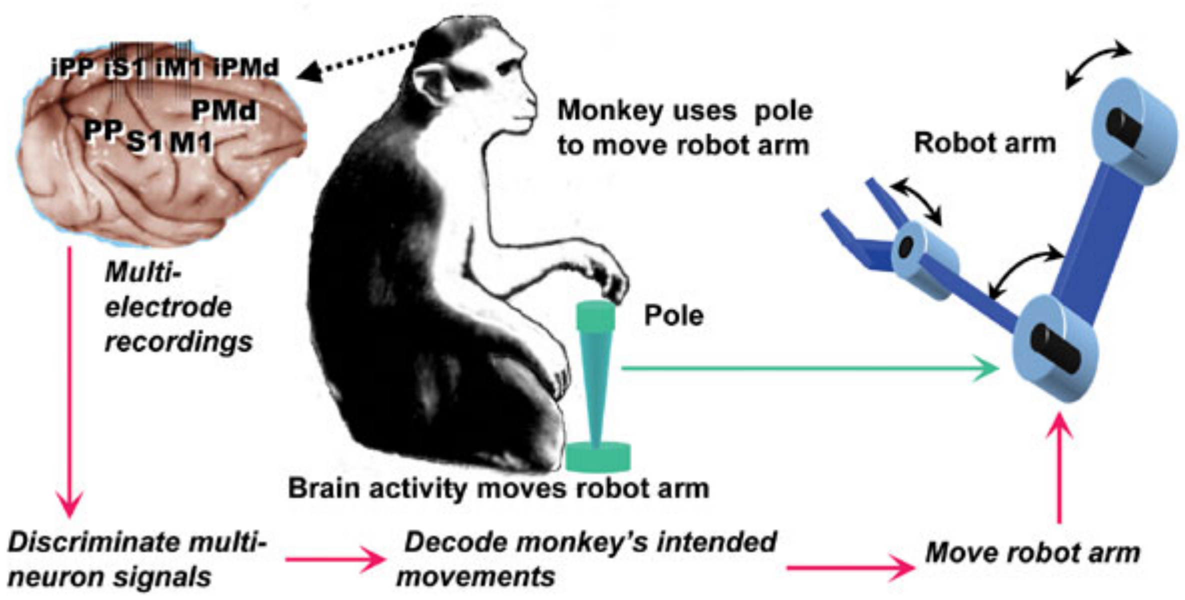


Data analysis and model classification

Regression

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

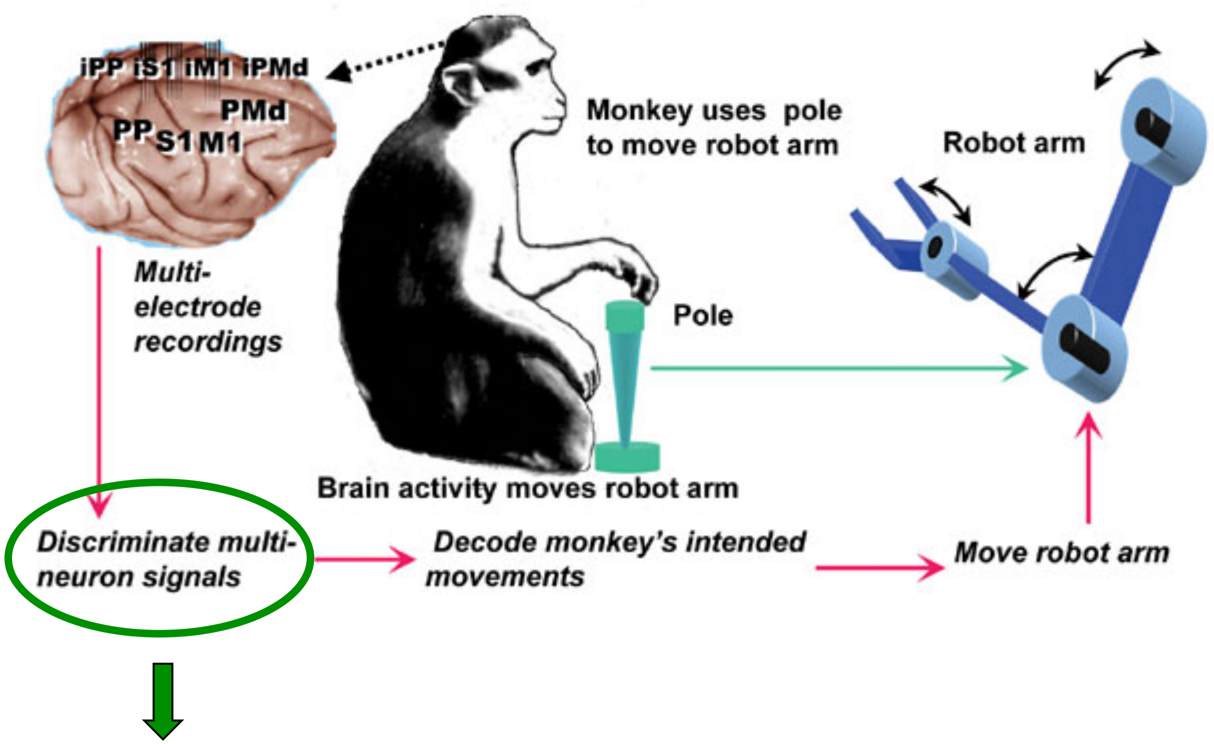
Example: Neuroprosthetics



Goal:

Decode arm movement direction from neural activity

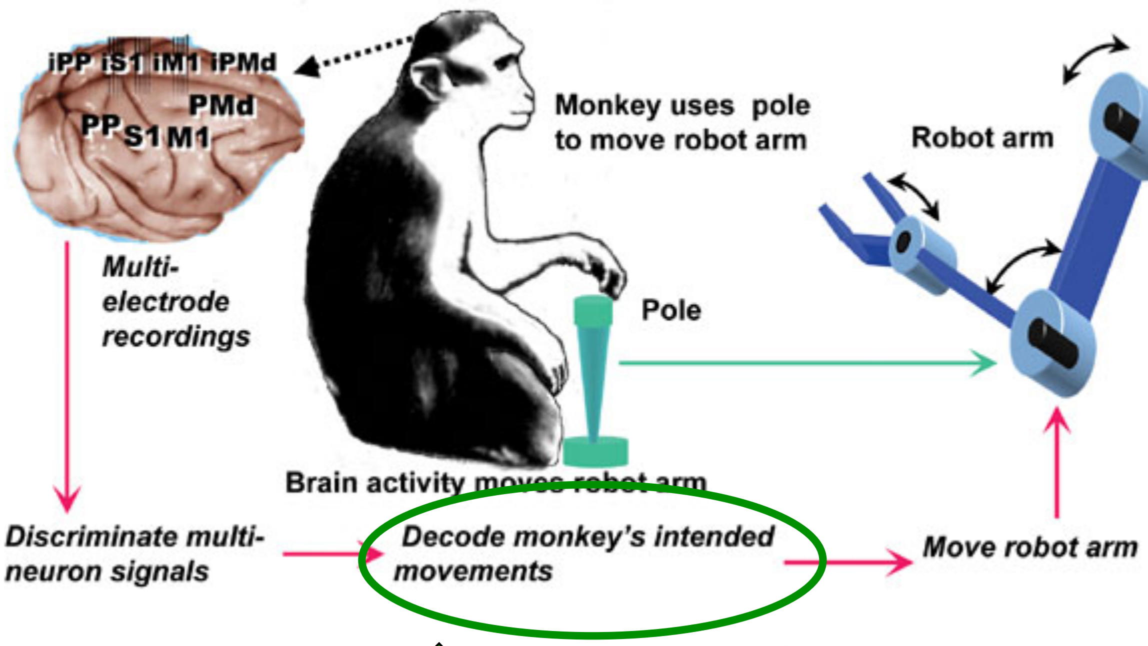
Example: Neuroprosthetics



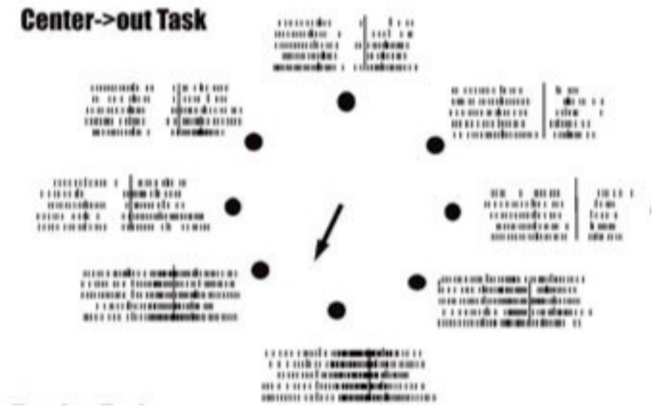
Each implanted electrode records signals from several <u>unknown</u> neurons

Individual neurons are identified using Unsupervised learning

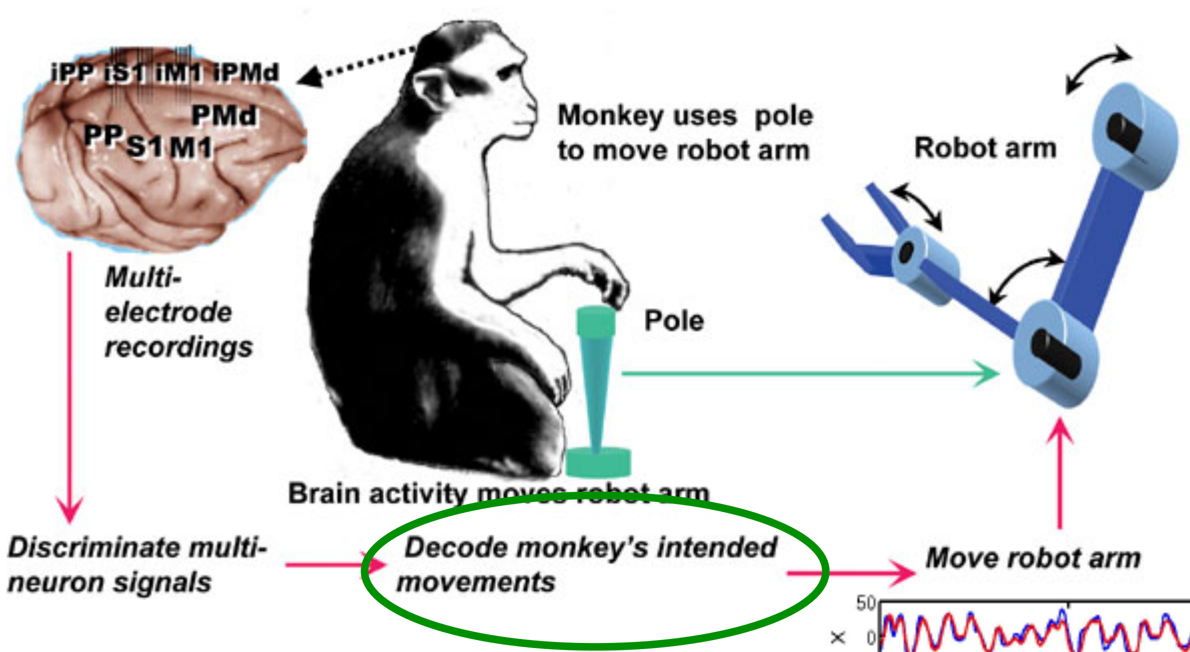
Example: Neuroprosthetics

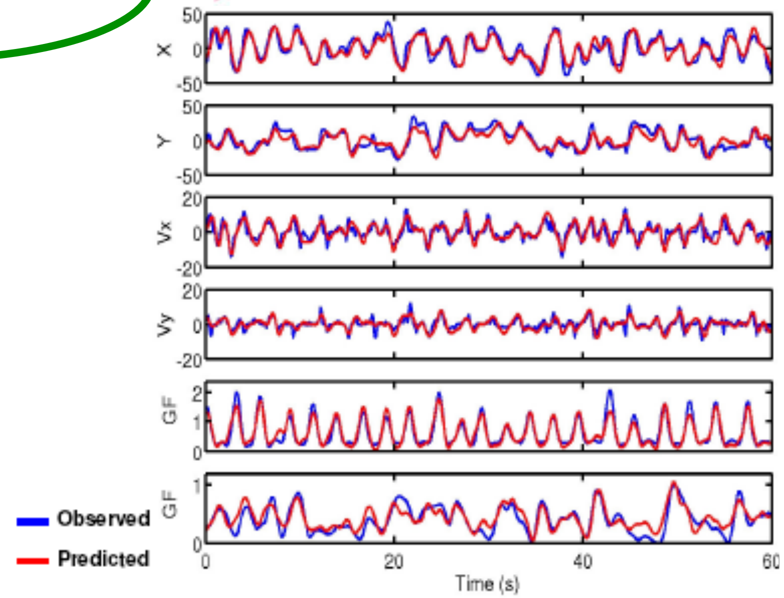


Relationship between neural activity (spikes) and arm intended movements are learned through Supervised learning



Example: Neuroprosthetics

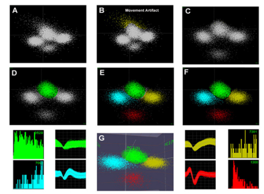




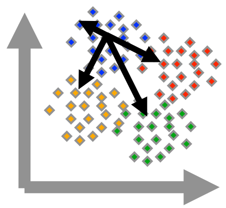
Relationship between neural activity (spikes) and arm intended movements are learned through Supervised learning

Summary: Unsupervised Learning

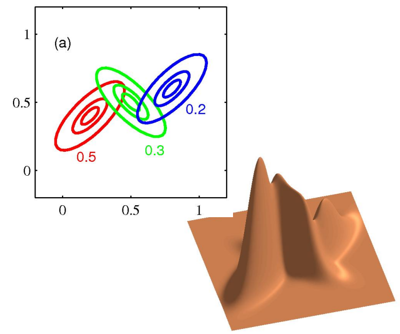
Unsupervised learning is used to process unlabelled data



Data can be characterized by a set of different clusters of data points: K-means algorithm

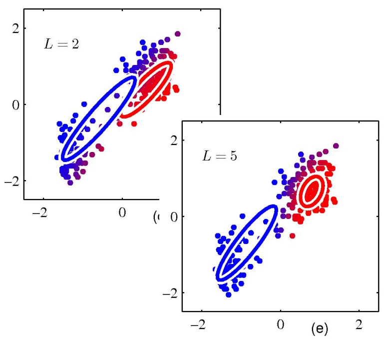


Data can alternatively described as a mixture of density functions (Gaussians)



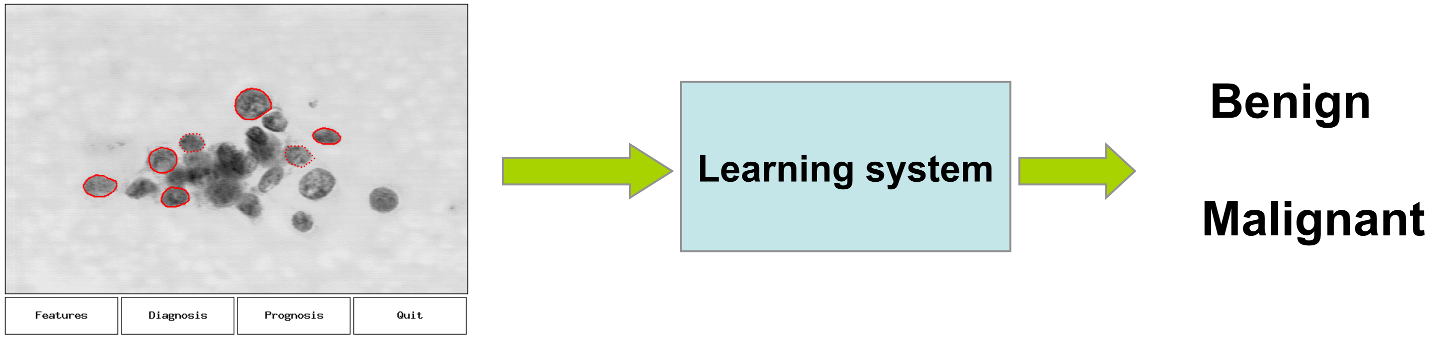
Parameters of the mixture are obtained by maximizing the likelihood of the observed samples

They can be obtained iteratively using the EM algorithm



Summary: Supervised learning

A classifier system defines a function maps the inputs into one of several *discrete* classes.

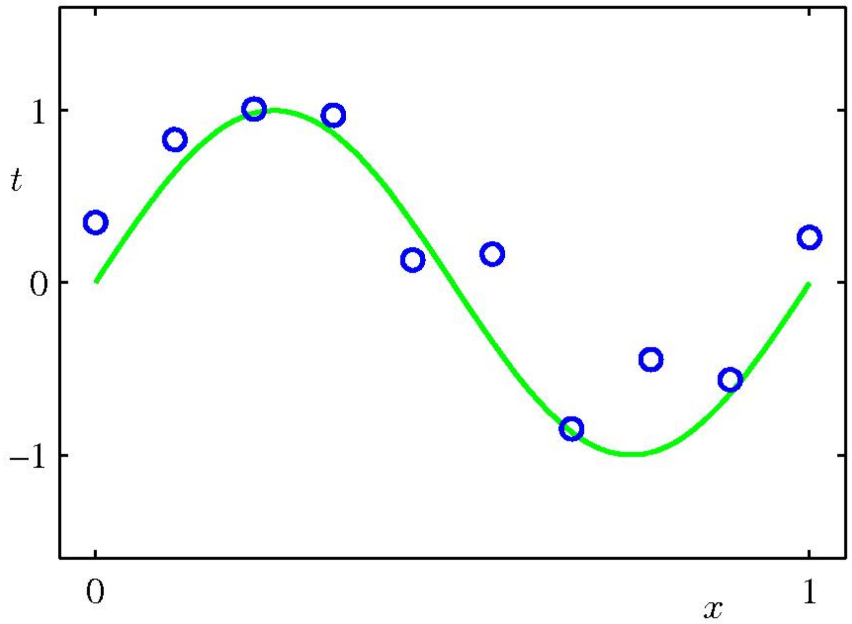


This function is learned from training examples so as to minimize some error function (e.g. the number of miss-classifications)

Supervised learning can also be used to approximate a <u>continuous</u> function...

Other examples?

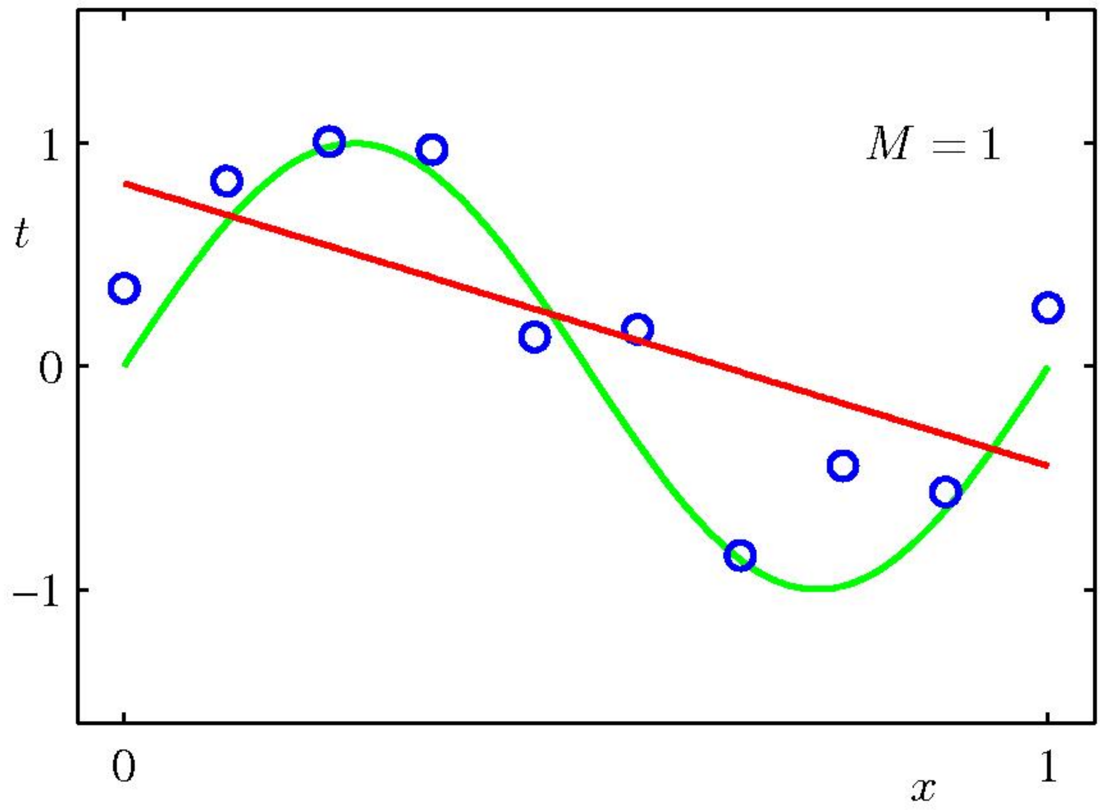
Linear Regression



- Given training data: pairs *(x^n, y^n=t^n)*, find a set of parameters (w_i) that minimizes some error measure

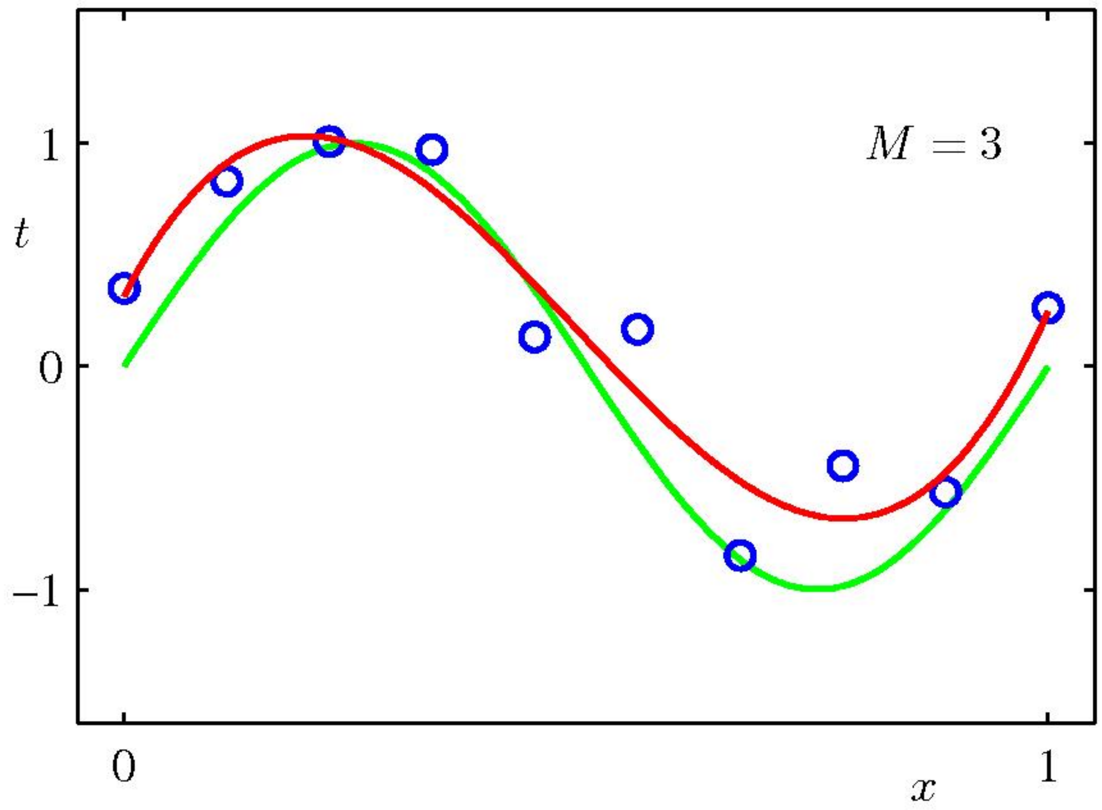
$$E = \frac{1}{2} \sum_{n=1}^{N} \{y(x^n, \mathbf{w}) - t^n\}^2$$

Linear Regression



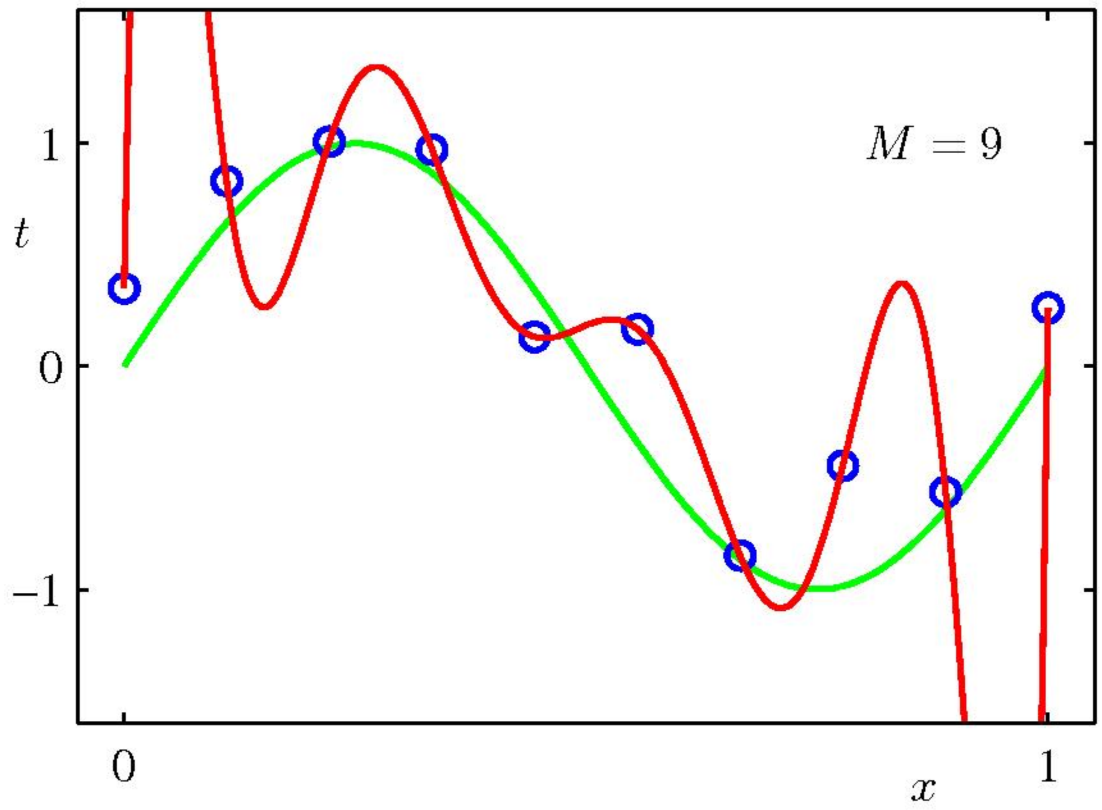
Linear discriminant function: $y(x) = w_0 + w_1 x$

Polynomial Regression



$$y(x) = w_0 + w_1 x + \ldots + w_M x^M = \Sigma_{j=0}^{M} w_j x^j$$

Polynomial Regression



$$y(x) = w_0 + w_1 x + \ldots + w_M x^M = \Sigma_{j=0}^{M} w_j x^j$$

Polynomial Regression

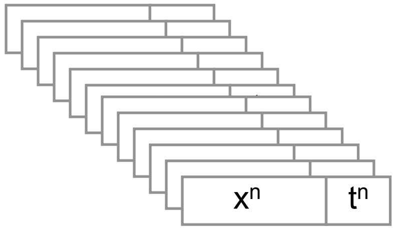
When to stop?

How to do it?

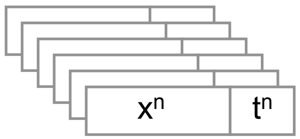
Generalization

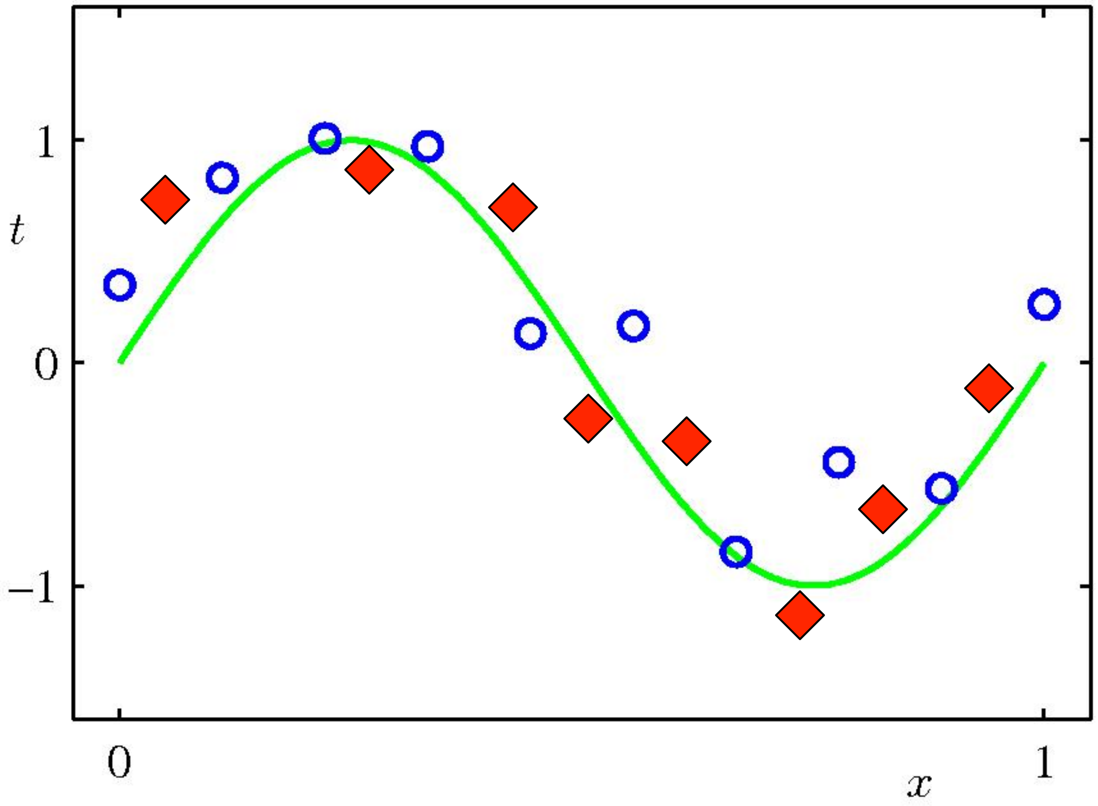
Examples
(input, target)

Training set



Validation set

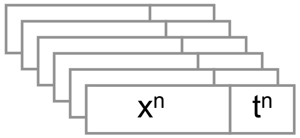




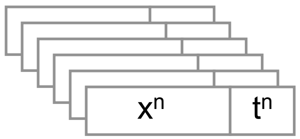
Generalization

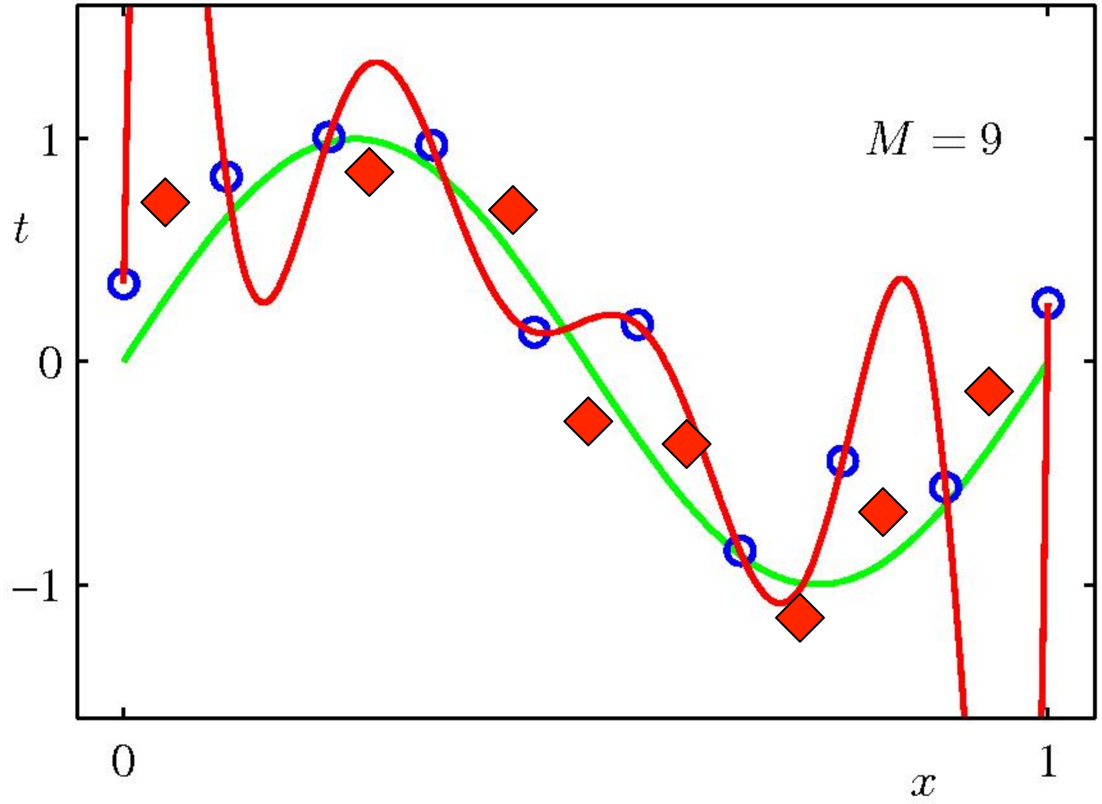
Examples
(input, target)

Training set



Validation set

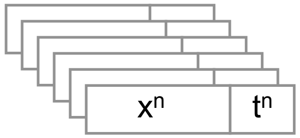




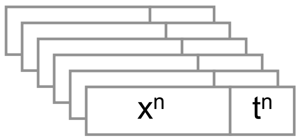
Generalization

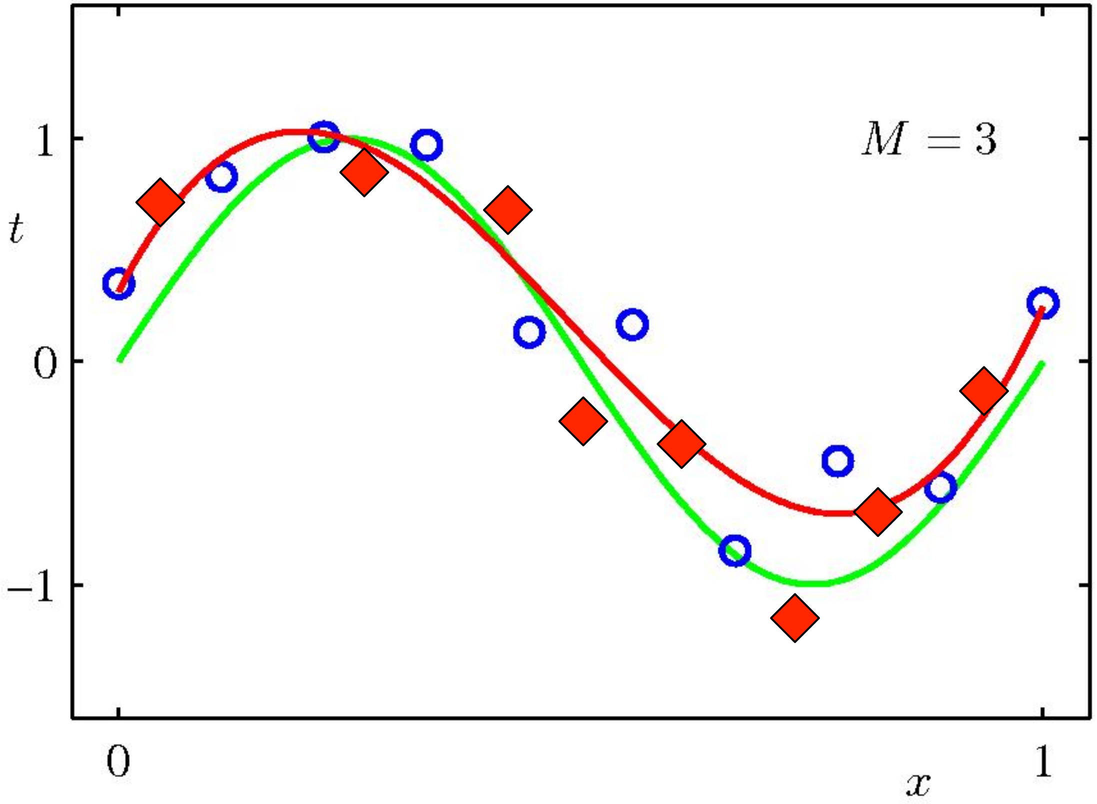
Examples
(input, target)

Training set

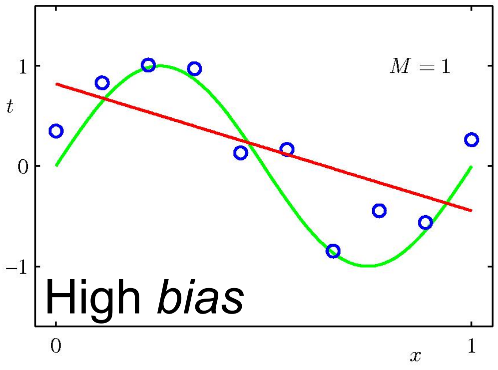


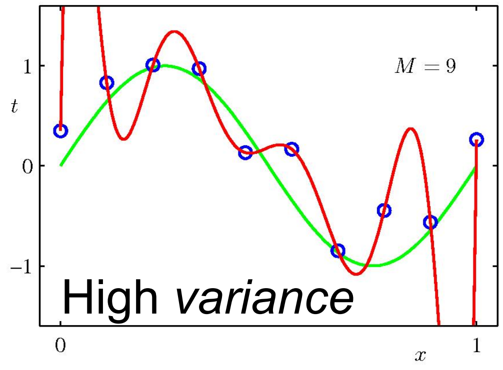
Validation set



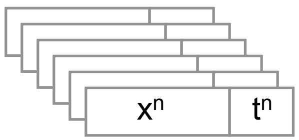


Generalization

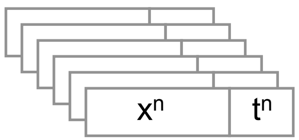


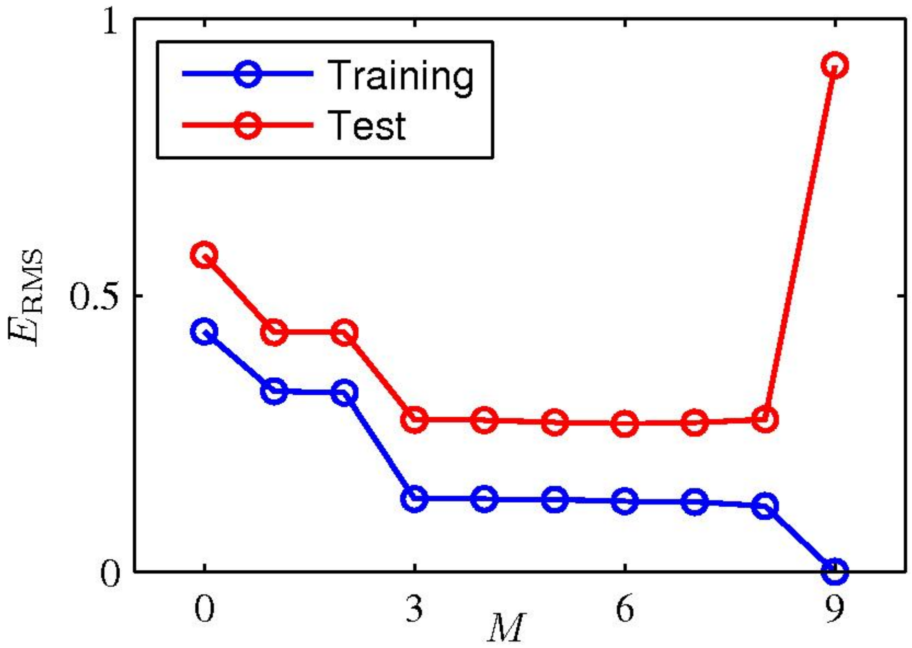


Training set



Validation set

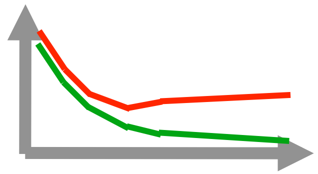




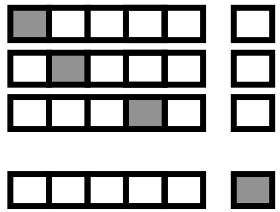
$$E^{RMS} = \sqrt{2\ E(\mathbf{w}^*)/N}$$

Training the model

- Accuracy of model should be assessed on a separate unseen test dataset



- Cross-validation can be used when not enough data is available



Linear Regression

Linear discriminant function: $y(x) = w_0 + w_1 x$

How to compute ***w***?

From the data!!

Linear Regression

Assume that the data (x^n, y^n) is given by

$$y^i = wx^i + \varepsilon^i$$

where

- noise is independent
- noise has a normal distribution with mean 0 and unknown variance σ^2

so $p(y|w,x)$ has a normal distribution too

- mean wx
- variance σ^2

Bayesian Linear Regression

We want to infer w from the data (x^n, y^n)

$$p(w|x^1,x^2,\ldots,x^n,y^1,y^2,\ldots,y^n)$$

- Use Bayes rule to estimate a posterior distribution for w
- Use Maximum Likelihood Estimation

Maximum Likelihood estimation

For what w is

$p(y^1, y^2, \ldots, y^n | x^1, x^2, \ldots, x^n, w)$ maximized?

Equivalently: for what w is

independent samples

$\Pi_i p(y^i | x^i, w)$ maximized?

Equivalently: for what w is

normal distribution

$\Pi_i \exp(-\frac{1}{2}(y^i - wx^i)^2/\sigma^2)$ maximized?

Equivalently: for what w is

log transform

$\Sigma_i -\frac{1}{2}(y^i - wx^i)^2/\sigma^2$ maximized?

Equivalently: for what w is

$\Sigma_i (y^i - wx^i)^2$ minimized?

Maximum Likelihood estimation

The maximum likelihood of w minimizes the sum-of-squares error

$$E = 1/n\ \Sigma_i (y^i - wx^i)^2 =$$

$$= \Sigma_i y^{i2} - (2\Sigma_i y^i x^i) w + (\Sigma_i x^{i2}) w^2$$

w is then

$$w = \frac{\sum_i x^i y^i}{\sum_i x^{i2}}$$

Multivariate Linear Regression

What about w_0?

What about multiple variables?

$$y(\boldsymbol{x}) = \boldsymbol{wx} + \varepsilon$$

Multivariate Linear Regression

$$y^i = w_0 + w_1 x_1^i + w_2 x_2^i + \ldots w_m x_m^i + \varepsilon$$

In matrix form:

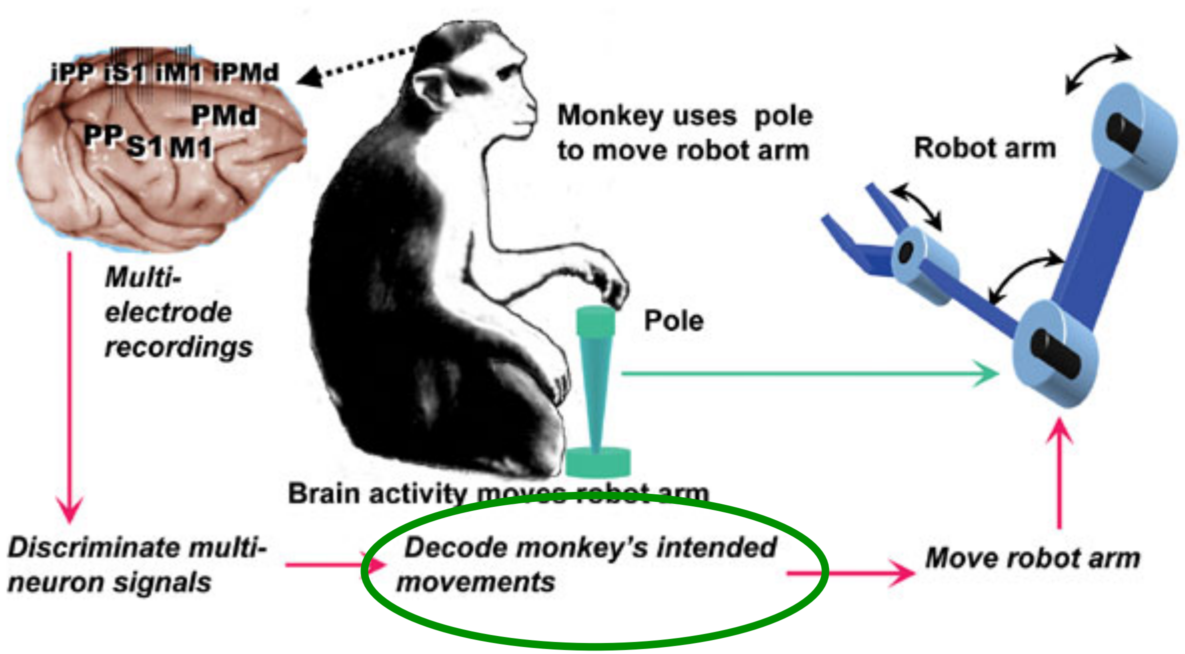
$$\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_m^1 \\ 1 & x_1^2 & x_2^2 & \cdots & x_m^2 \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_1^n & x_2^n & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{bmatrix}$$

$$Y = \mathbf{X}W$$

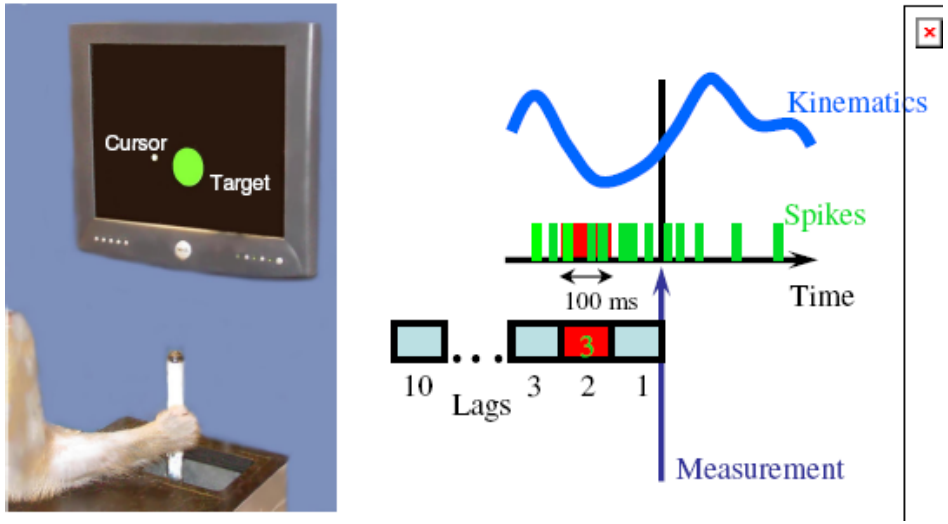
$$W = \boxed{(\mathbf{X}^T\mathbf{X})^{-1}}\mathbf{X}^T Y$$

pseudoinverse
computable when $\mathbf{X}^T\mathbf{X}$ is invertible
or via gradient descent otherwise

Example: Neuroprosthetics



Example: Neuroprosthetics

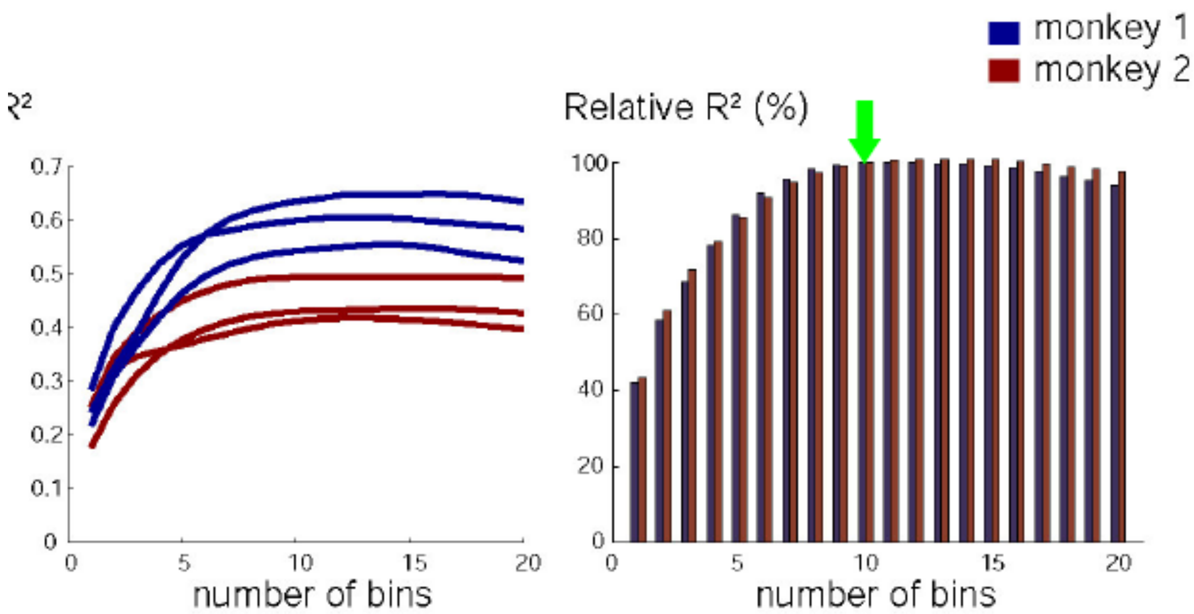


$$\mathbf{y}(t) = \mathbf{b} + \sum_{u=-m}^{n} \mathbf{a}(u)\mathbf{x}(t-u) + \varepsilon(t) \qquad \mathbf{Y}=\mathbf{X}\mathbf{A}, \qquad \mathbf{A}=inv(\mathbf{X}^T\mathbf{X})\mathbf{X}^T\mathbf{Y}$$

(Carmena et al., PLoS 2003)

Example: Neuroprosthetics

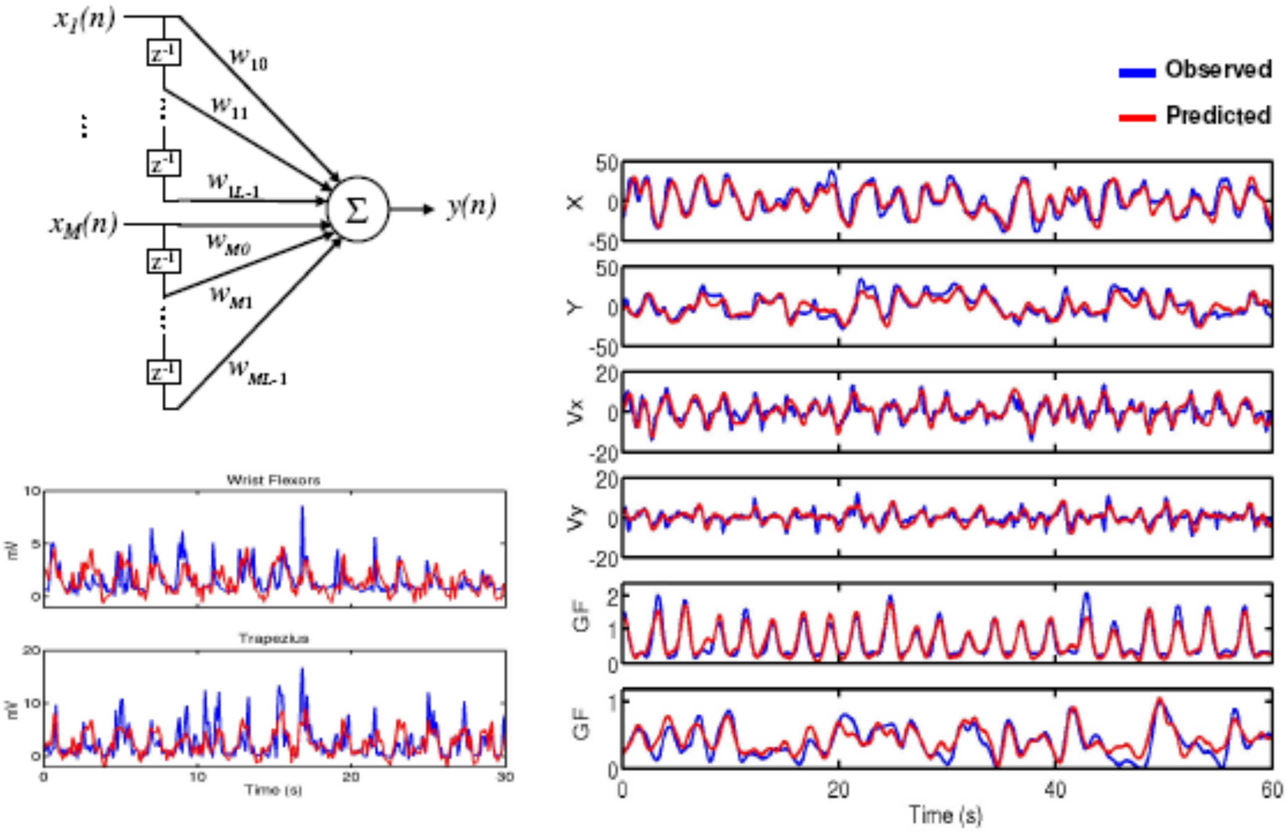
Parameter selection



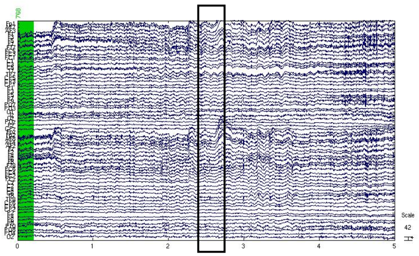
Optimal time lag
10 bins

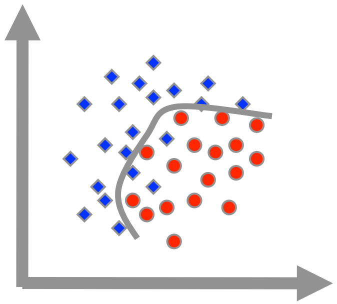
Example: Neuroprosthetics

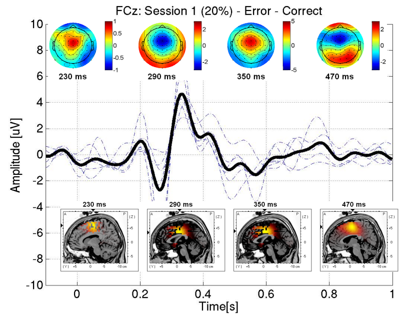
Prediction



(Carmena et al., PLoS 2003)







Data analysis and model classification

Regression

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE