machine learning for hand prosthetics
(and more)

barbara **CAPUTO**[1], claudio **CASTELLINI**[2], gerd **HIRZINGER**[3], luo **JIE**[1], giorgio **METTA**[2,4], francesco **ORABONA**[1], giulio **SANDINI**[2,4], patrick **VAN DER SMAGT**[3]

[1] *IDIAP*, martigny, switzerland
[2] *LIRA-Lab*, university of genova, italy
[3] *German Aerospace Research Center*, oberpfaffenhofen, germany
[4] *Italian Institute of Technology*, genova, italy

www.neurobotics.org

- robotic artifacts driven by humans (HBS), e.g.
  - intelligent prostheses
  - intelligent teleoperation platforms
- ...where by *intelligent* we mean *semi-autonomous* and/or *adaptive*
- solution: *learn* models of complex actions, e.g., reaching, grasping, etc. that can be used by HBSs

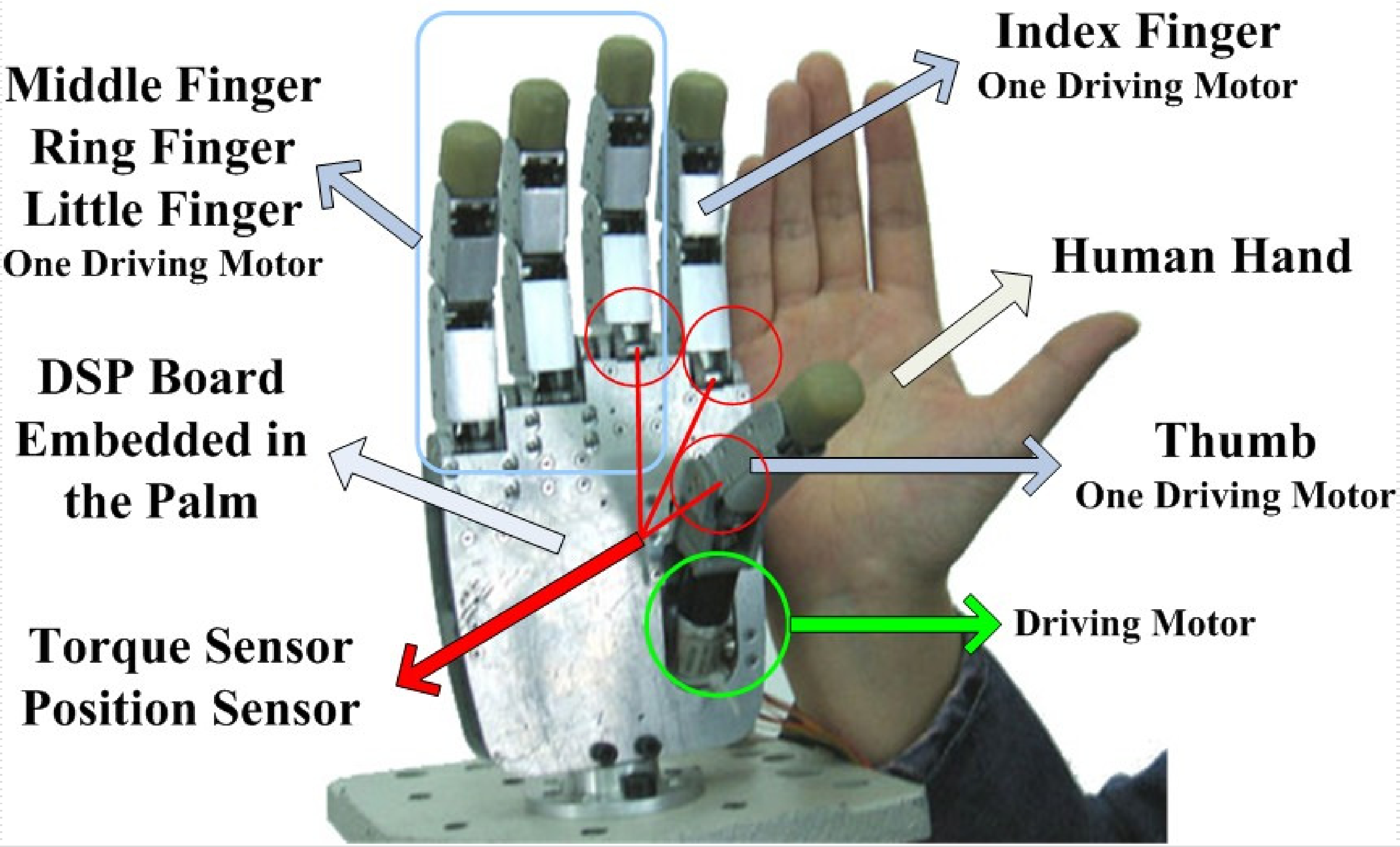
www.neurobotics.org

- in intelligent prosthetics:
  - guess *what* the patient wants his prosthesis to do
- in teleoperated reaching and grasping:
  - guess *when* the master wants to grasp
  - guess *how* the master wants to grasp
  - guess *what* the master wants to grasp

outline

- application of machine learning to
  - an emg-controlled hand prosthesis
  - indoor robotic navigation
  - human reaching and grasping
- some optimisations (on the road)
  - uniform sampling in the input space
  - linear independence in the feature space

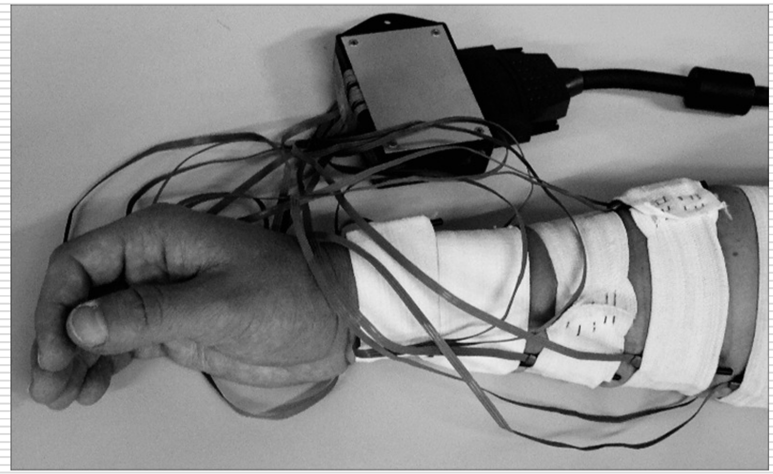
emg-driven hand prosthesis



problem

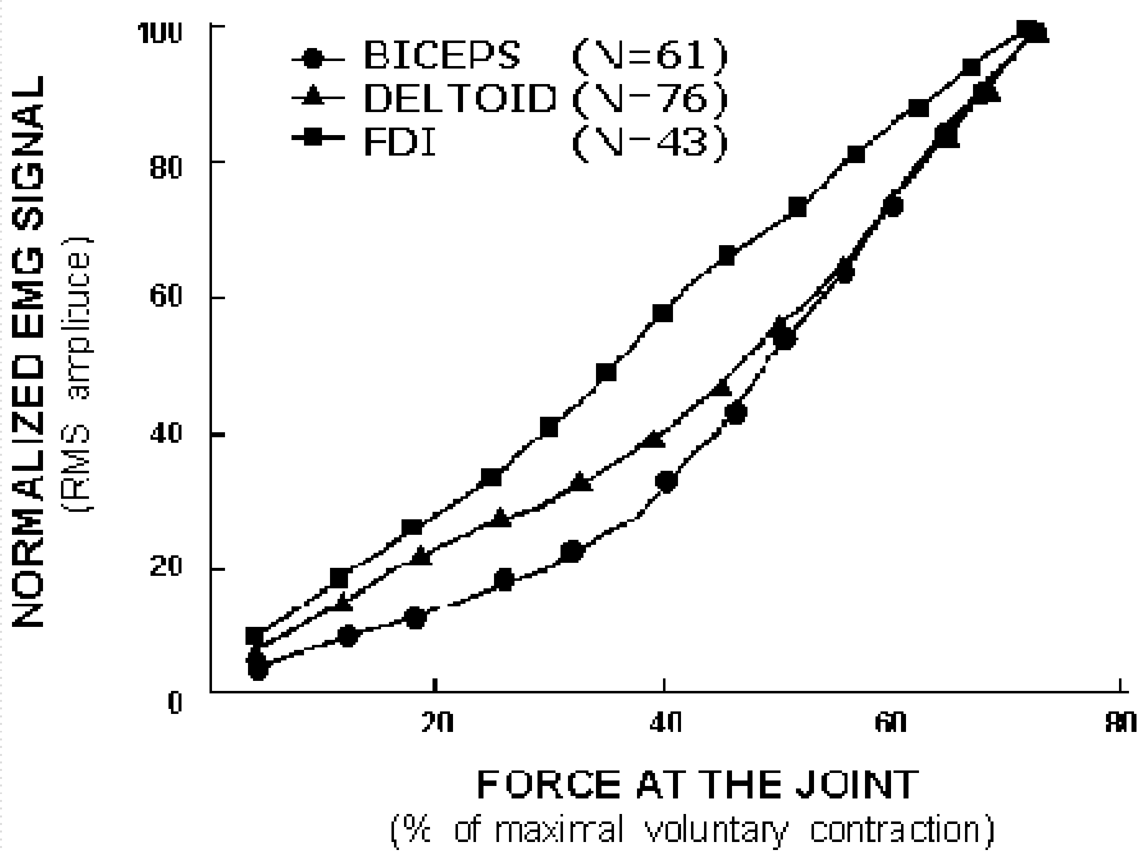
- high dexterity, but
- little possibility of *control* by the patient: what interface?
- we focus upon *non-invasive* interfaces, particularly upon
- **forearm surface electromyography**
- can a mechanical hand be swiftly driven using the emg?

setup



- 10 Ottobock emg electrodes
- 1 SpaceControl force/torque sensor
- 4 fingertip sensors
- detect ***type*** of grasps
- detect ***force*** involved in the grasp

emg (1)



- in principle, non-linearly related to the *force* applied by a muscle

emg (2)

- a very badly conditioned signal
- affected by a number of factors:
  1. *(long-term)* inter-arm differences
  2. *(long-term)* arm postures and movements
  3. *(short-term)* muscular fatigue, sweat
  4. *(medium-term)* electrode displacement, muscle cross-talk
- how to take into account all these problems?

emg (3)

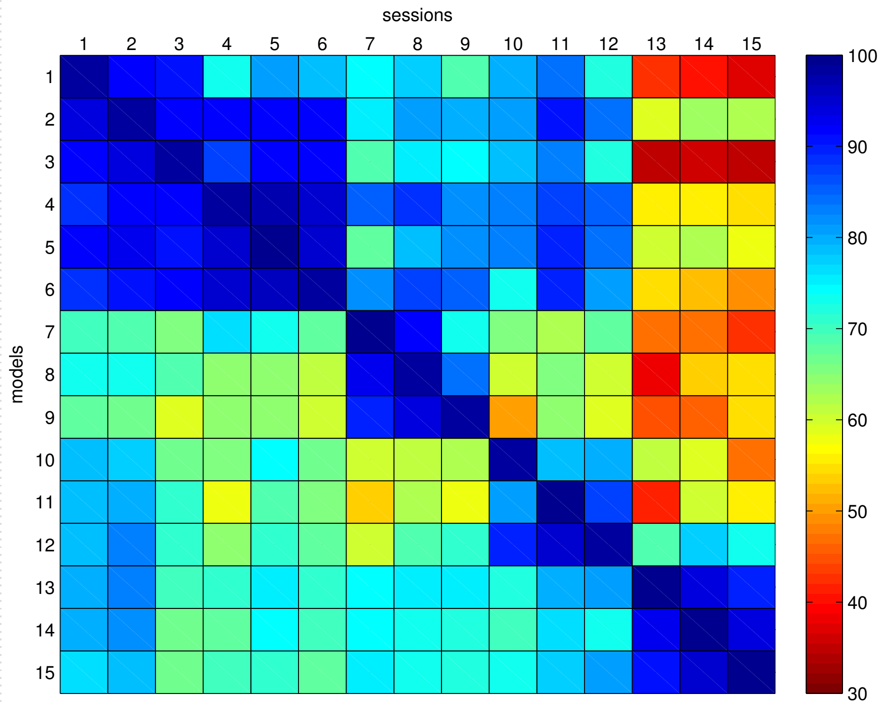
- how to take into account all these problems?
  1. (neglected) one subject only, able-bodied
  2. (neglected) relaxed on a table, fixed position
  3. three-and-a-half minutes of activity (one *session*)
  4. three sessions form a *group;* electrodes replaced between groups
- *all in all, 30 sessions in 10 groups across 2 days*

methods

- feed-forward neural network (classification, regression)
  - sigmoidal activation
  - one hidden layer w/10 neurons
  - backprop
- support vector machine (classification, regression)
  - gaussian kernel
- locally weighted projection regression (regression only)
  - online, incremental

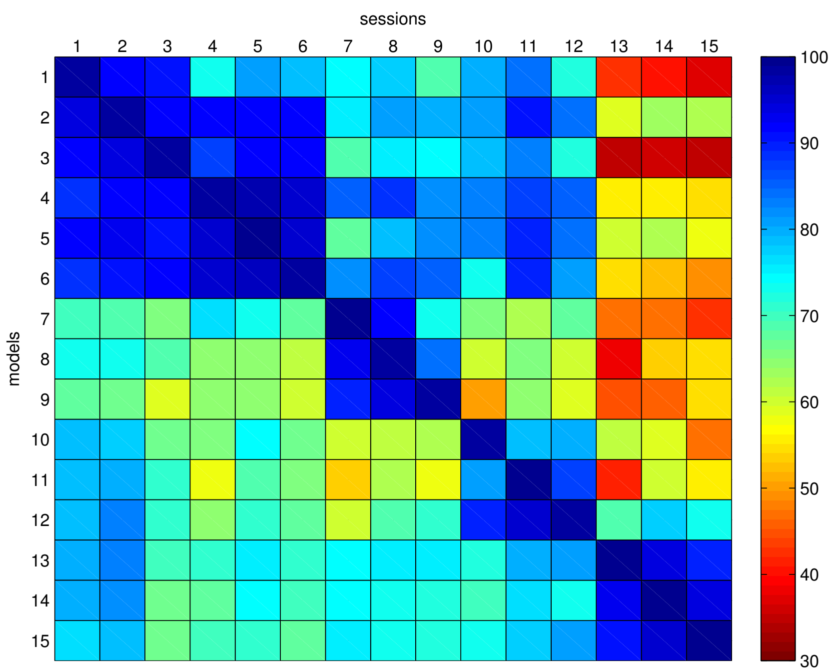
svc cross-session accuracy

$A_{ij} =$



A_{ij} is the classification accuracy of model trained upon session i, tested on session j. graphs refer to day 1 – day 2 is analogous.

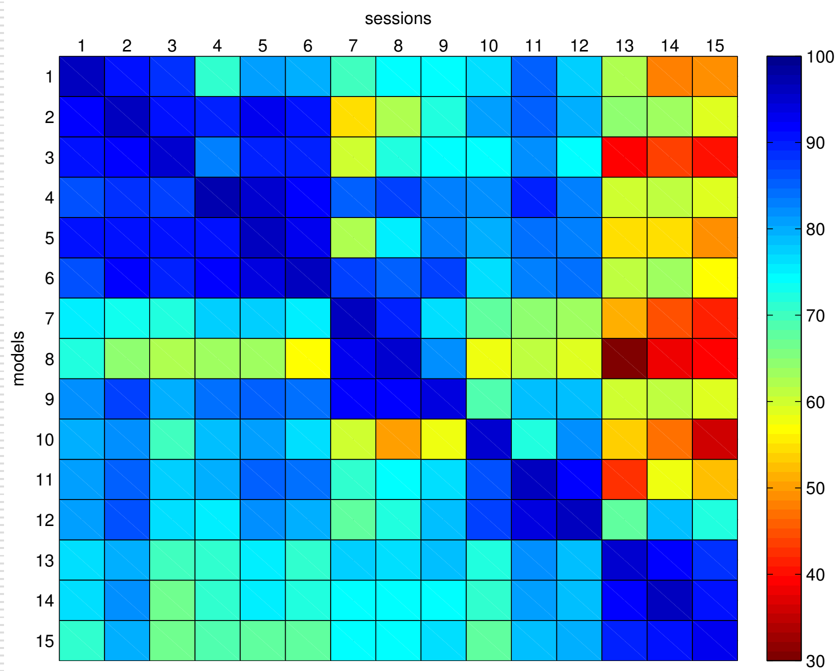
uniformisation



full training sets

diag: 98.73% ± 0.39%

non-diag: 73.23% ± 14.29%



uniform training sets

diag: 95.52% ± 1.21%

non-diag: 74.53% ± 13.70%

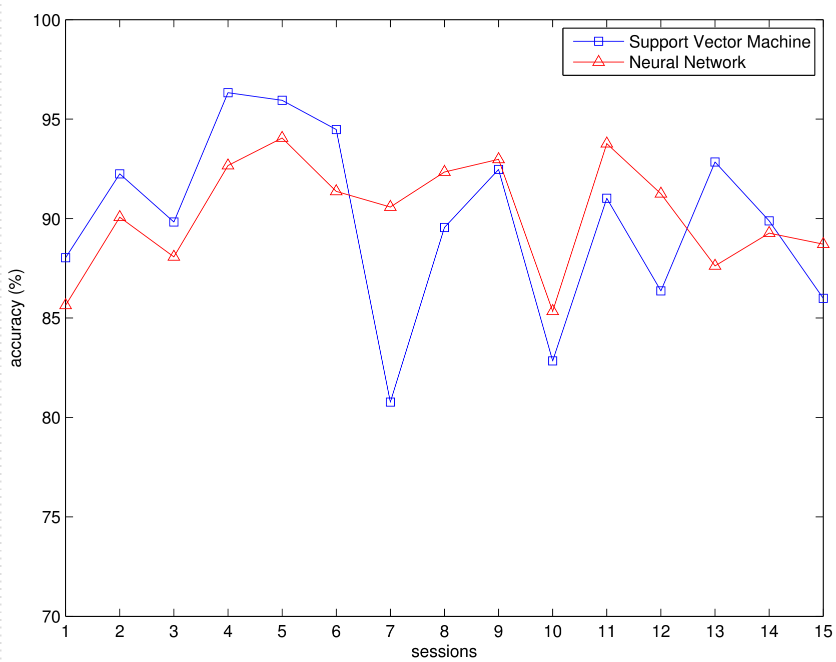
what makes it hard?

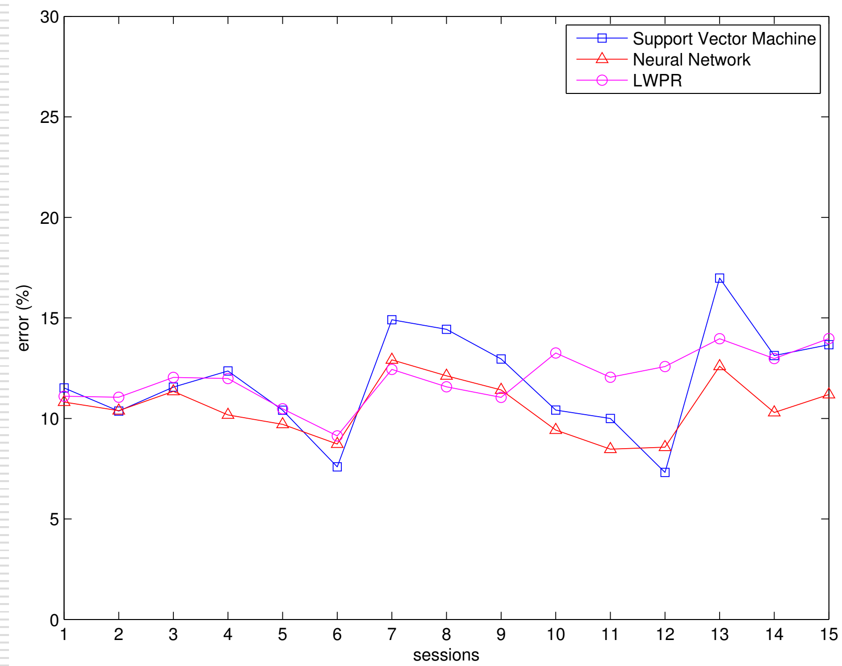
- average minimum sample set distance:

$$D_{ij} = \frac{1}{|S_j|} \sum_{s_j \in S_j} \min_{s_i \in S_i} ||s_j - s_i||^2$$

- D_{ij} is strongly correlated to A_{ij} (Pearson coefficient: -0.60)
- analogous results with nn and on regression (all approaches)
- what if we *adjoin good models* and then train on this new model?

best models



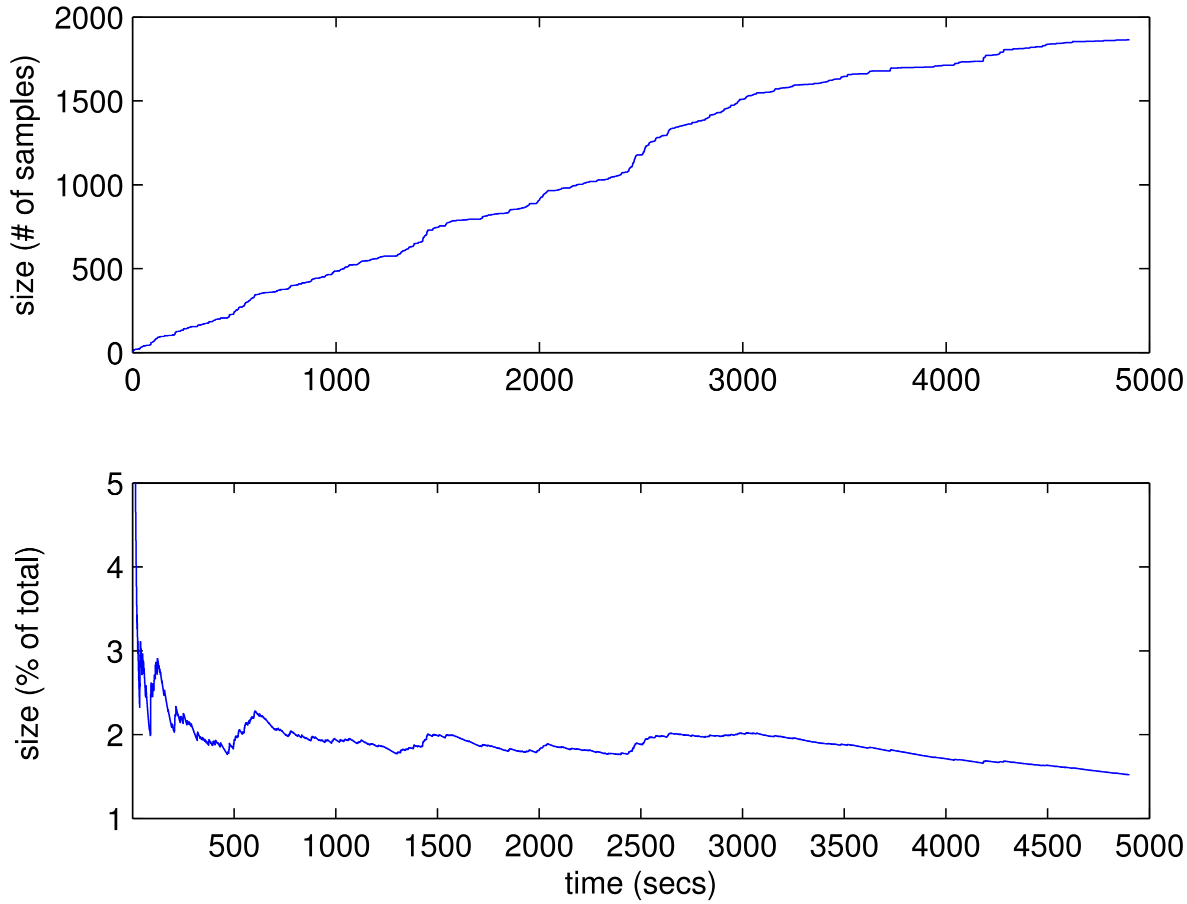


best models on day 1, classification accuracy (left) and regression NRMSE (right)

intermission (1)

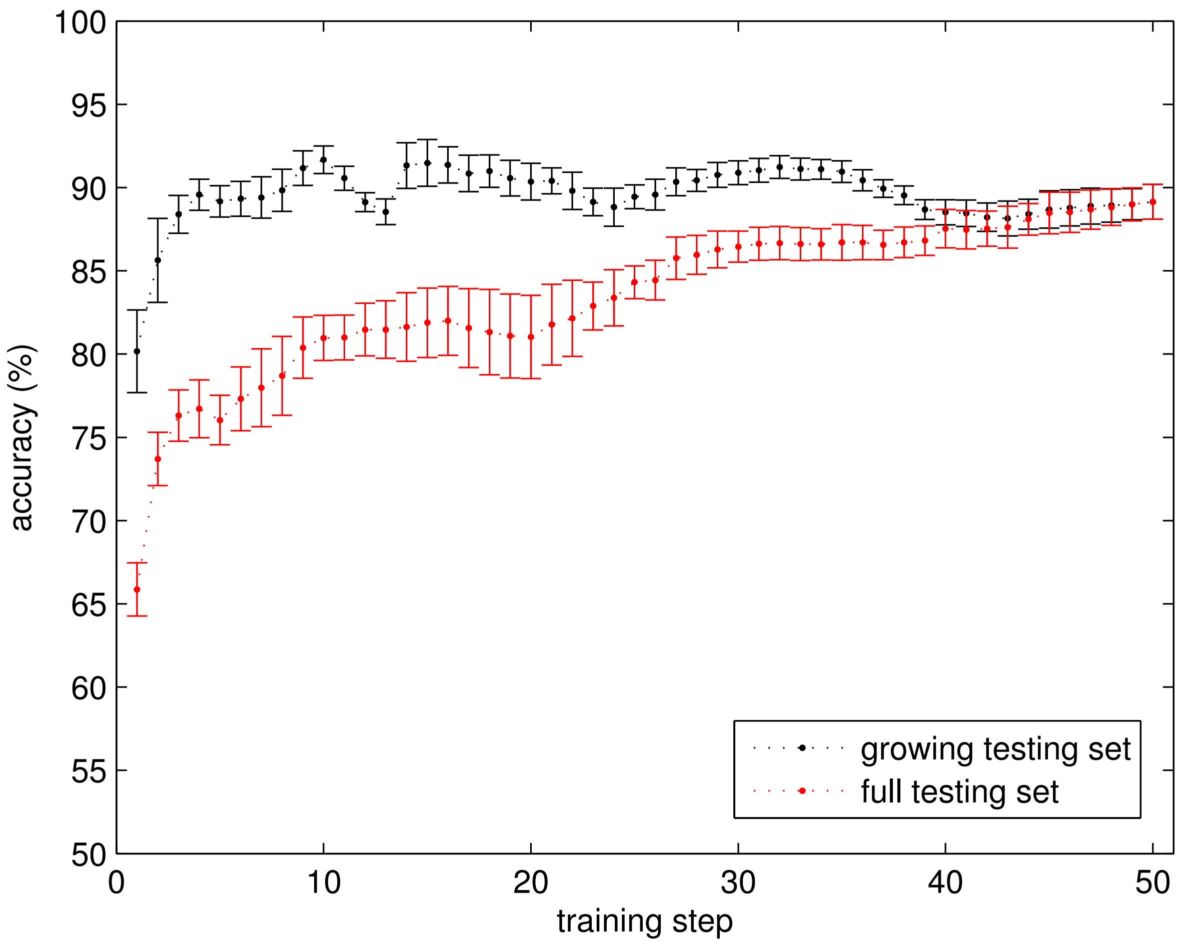
- information *is there,* one only has to dig and find it!
- distance is the key
- need a mechanical way of building good models
- on-line version of uniformisation:
  - check wheter a new sample is far away from the current training set
  - if so, use it
  - otherwise, ignore it

growth of a training set



growth is expected to come to a halt eventually...

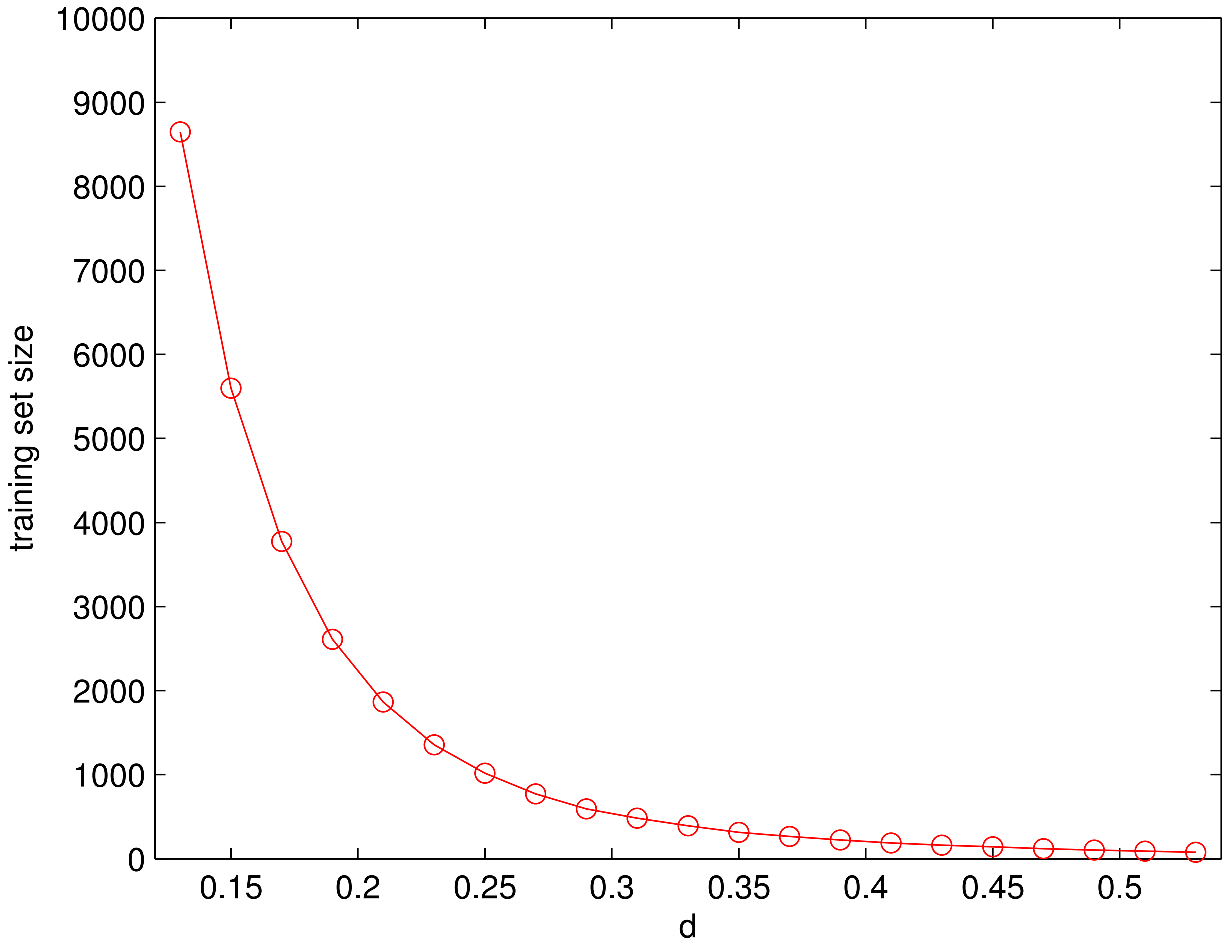
on-line accuracy



black line says: point by point, we're finding what you need.

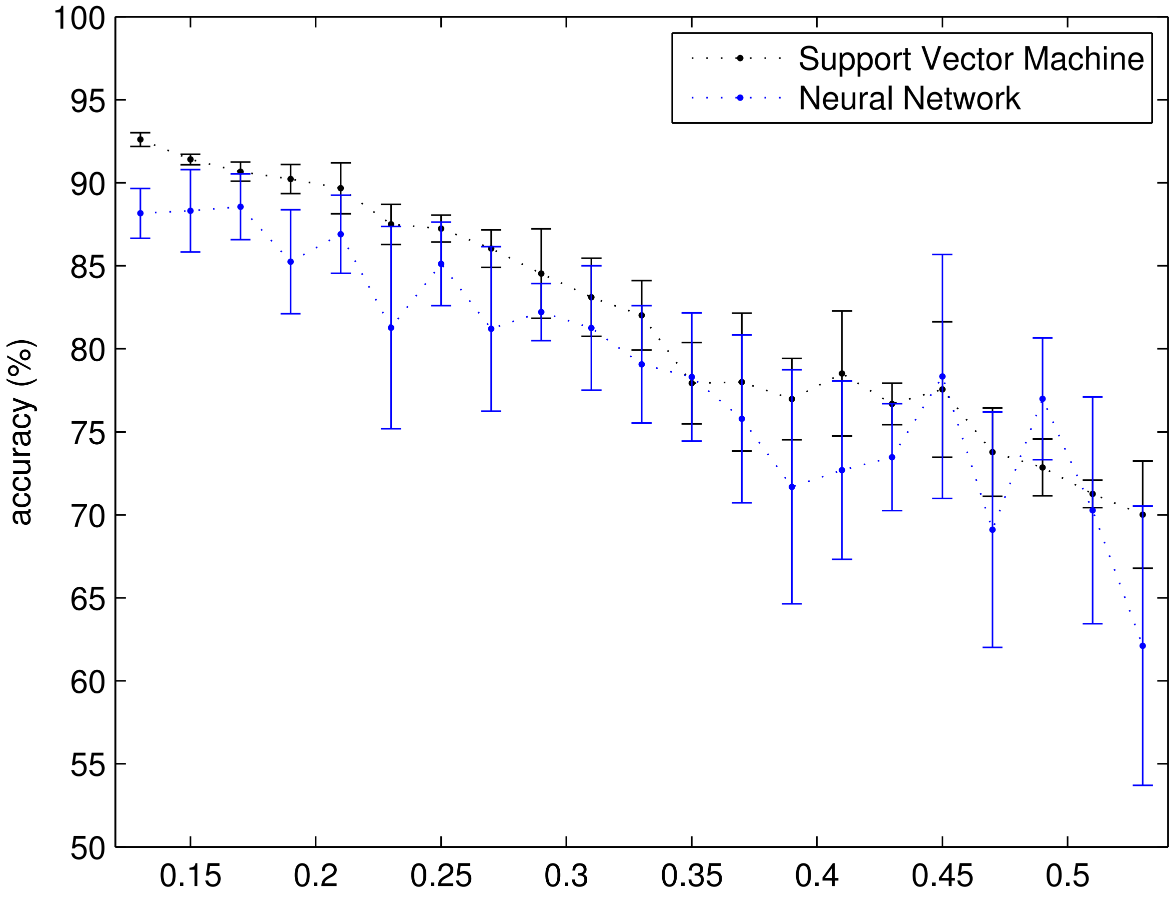
red line says: we aren't going to miss anything.

as the distance grows... (1)



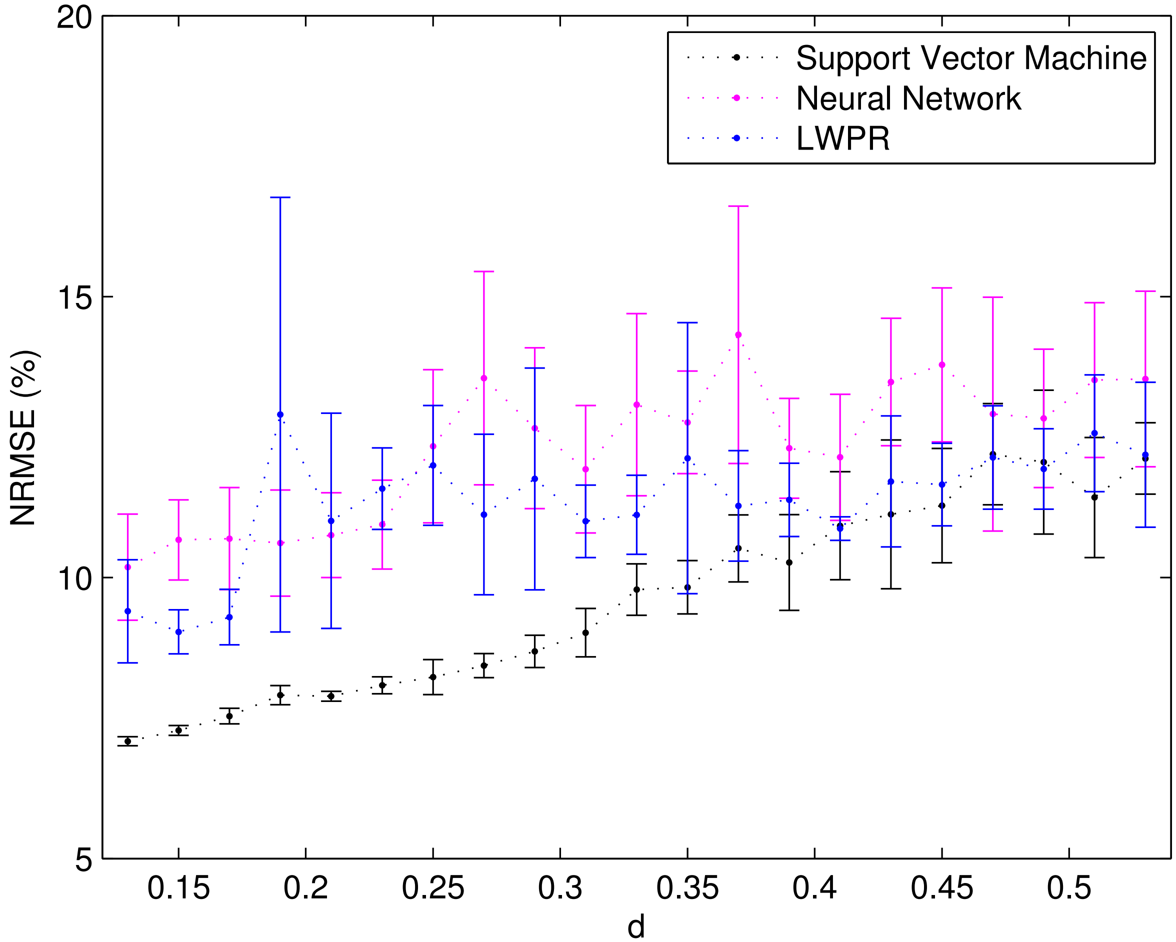
training set size decreases as d^{-10}

as the distance grows... (2)



accuracy decreases linearly.

as the distance grows... (3)



error increases linearly.

for d=0.53, error is 12.12%, and the training set has 77 samples over 153.000!

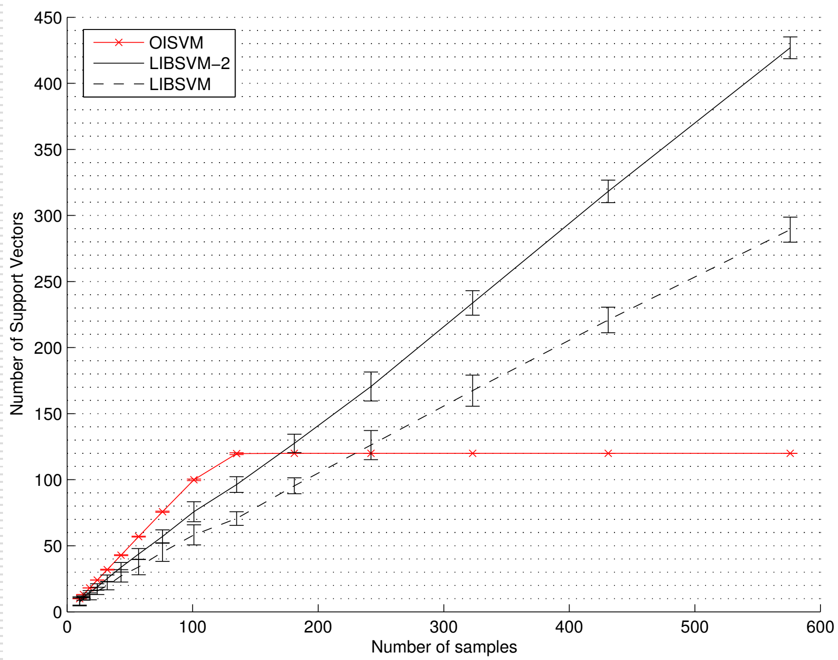
conclusions

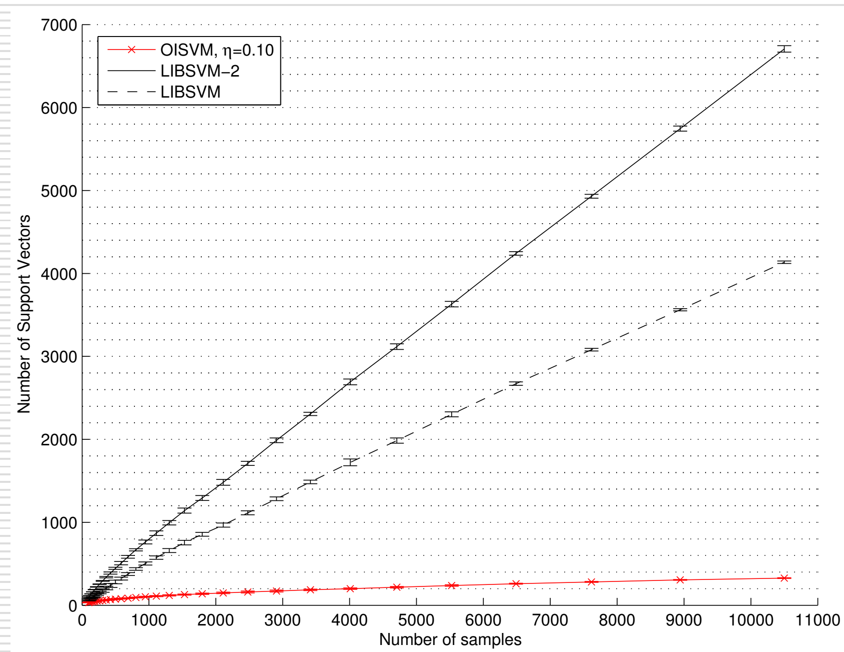
- problem is solved so far!
- now, gather more data to take arm movement into account
  - the online uniformisation seems a promising way to keep training sets small and effective
- then, control the hand, and lastly...
- ...try it on a patient.

learning to *recognise places*

- two mobile robots gather images of a complex building in different conditions of lighting, camera position, settings
- they are then asked to recognise the places they've been to

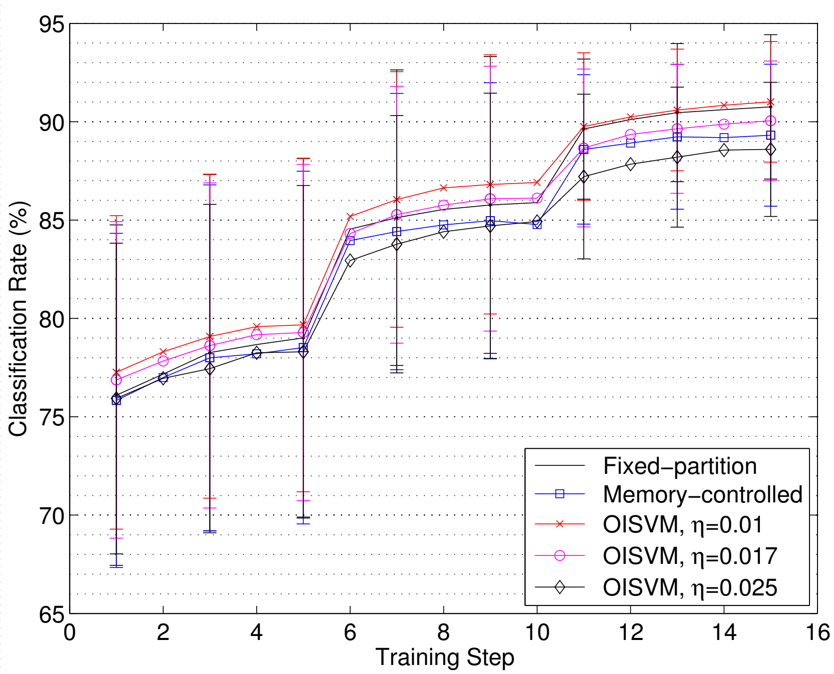
learning to *recognise places*

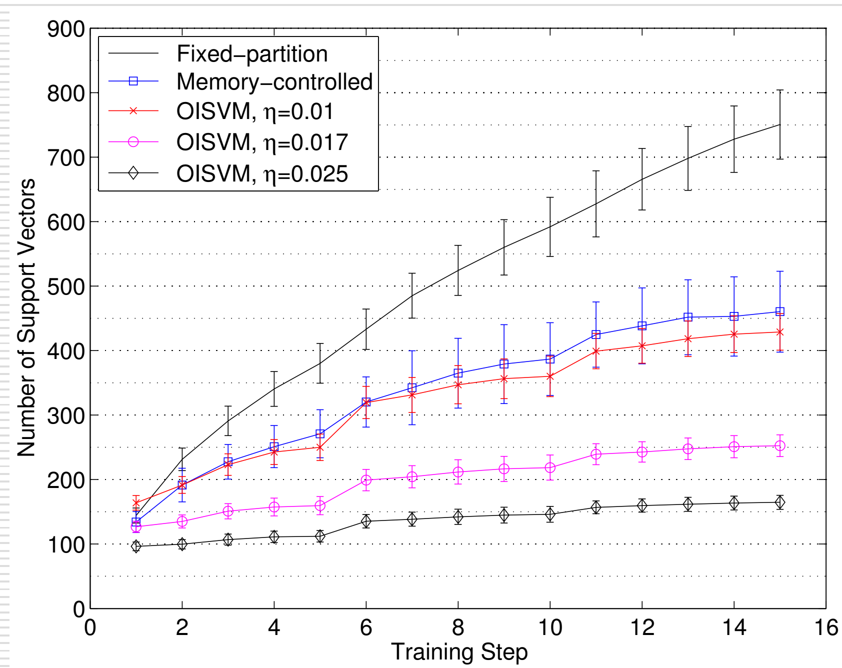




growth: the size of the models reaches a maximum value and then stops (left pane, finite kernel) or it grows much less (right pane, infinite kernel). the accuracy of the machines is *exactly the same*.

learning to *recognise places*





accuracy: one can obtain massively smaller models, if accuracy is allwed to degrade by a maximum of 4%.

learning *more efficiently*

- OISVM: *Online Independent Support Vector Machines* keep models small by exploiting *linear independence* in the *feature space:*
    - *check whether a new sample is*
    - *if so, use it*
    - *otherwise, ignore it.*
- suited for online learning, where
    - data appear out-of-the-blue with no possible reordering and/or statistics
    - data flow is potentially endless
- no approximation involved, or able to tune the approximation

what is l.i. in the feature space?

- recall your I year linear algebra:

$$\Delta = \min_{\mathbf{d}} \left\| \sum_{j \in \mathcal{B}} d_j \phi(\mathbf{x}_j) - \phi(\mathbf{x}_{l+1}) \right\|^2$$

- in real life, you need to check whether Δ is less than some *η* (linearly dependent) or greater than it (linearly independent)

what is l.i. in the feature space?

$$\begin{aligned} \Delta_i \leq \eta \quad & \Leftrightarrow \quad 1 - K(\mathbf{x}_{l+1}, \mathbf{x}_i)^2 \leq \eta \\ & \Leftrightarrow \quad \boxed{K(\mathbf{x}_{l+1}, \mathbf{x}_i) \geq \sqrt{1-\eta}} \\ & \Leftrightarrow \quad \exp\left(-\gamma \left|\left|\mathbf{x}_{l+1} - \mathbf{x}_i\right|\right|^2\right) \geq \sqrt{1-\eta} \\ & \Leftrightarrow \quad \boxed{\left|\left|\mathbf{x}_{l+1} - \mathbf{x}_i\right|\right|^2 \leq -\frac{1}{2\gamma} \log\left(1-\eta\right)} \end{aligned}$$

- in general, ignore x_{l+1} if K-close to x_i
- if K is gaussian, then ignore x_{l+1} if close to x_i... does this ring a bell?

intermission (2)

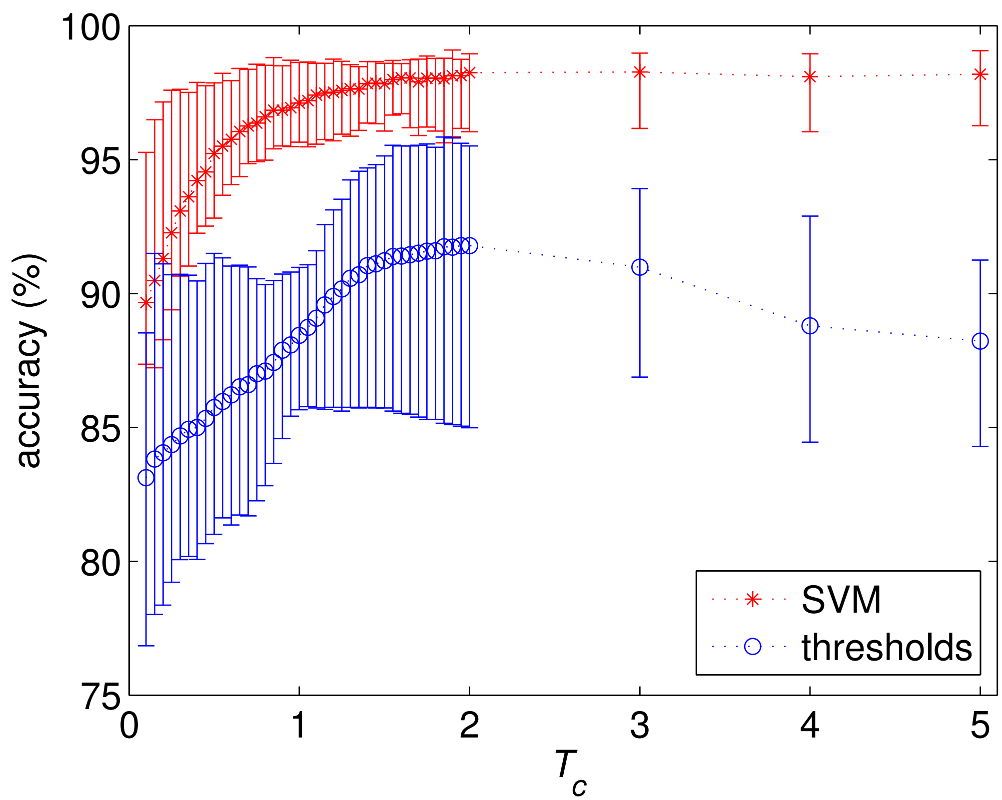
"When computers are applied to solve a practical problem it is usually the case that the method of deriving the required output from a set of inputs can be described explicitly. [...] As computers are applied *to solve more and more complex tasks,* however, situations can arise when *there is no known method for computing the desired output from a set of inputs,* or where that computation may be very expensive."

[Cristianini & Shawe-Taylor, An introduction to Support Vector Machines, 2000]

learning *when* to grasp

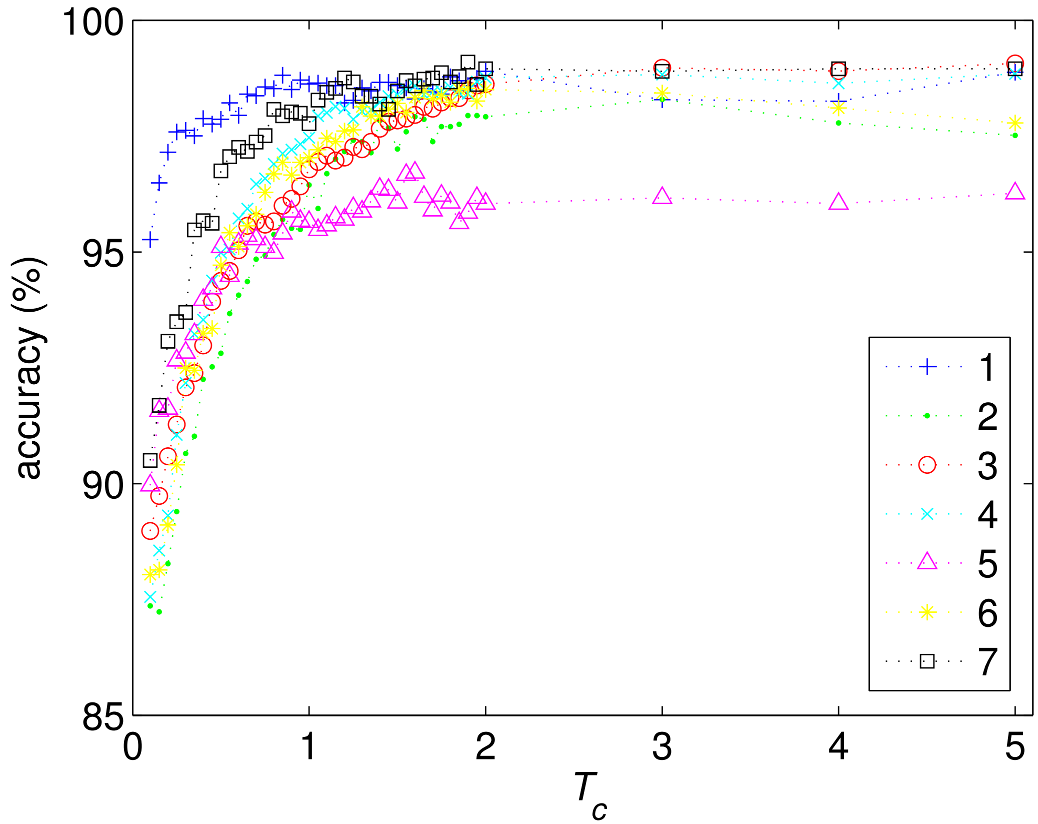
- human subjects look at objects on a monitor and simulate the act of grasping
- by monitoring the user's *gaze* and *arm motion*, guess when he wants to grasp
- consider an adequate *time window* of *gaze variance* and *arm velocity*

learning *when* to grasp



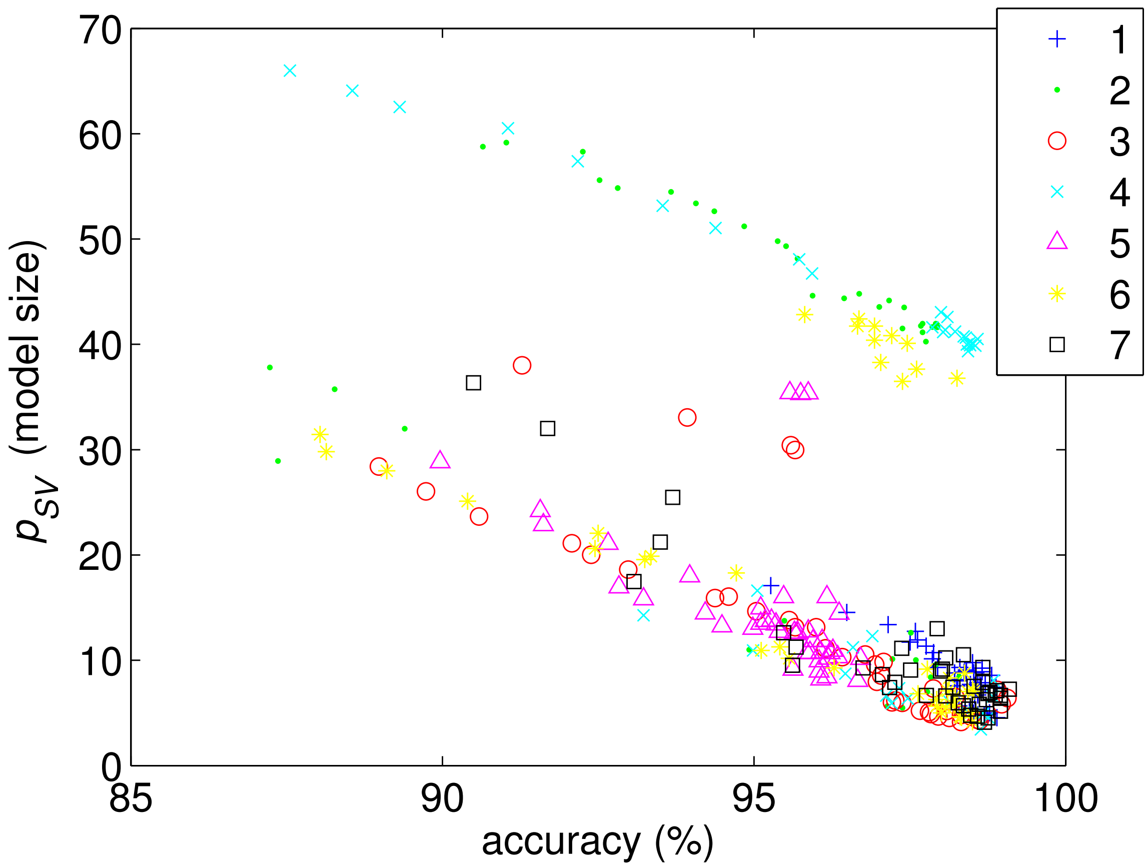
comparison: Support Vector Machines vs. a simple decision tree

learning *when* to grasp



accuracy: how well do we guess the subject's intention?

learning *when* to grasp



compactness: how large are the models obtained?

learning *when* to grasp

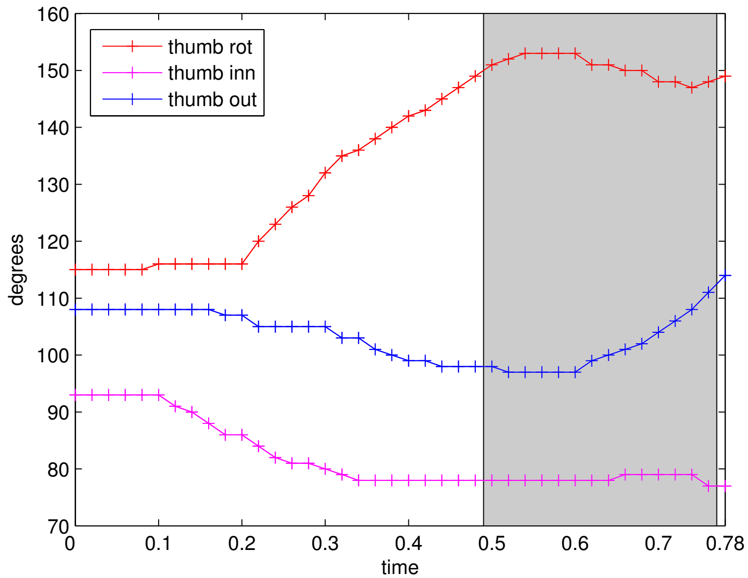
- SVMs build excellent models of the user’s intention to grasp:
  - accurate
  - small (i.e., fast and suitable for online applications)
- the problem is easy *in the feature space*, not necessarily in the input space!
- research about (functional) biological plausibility is ongoing

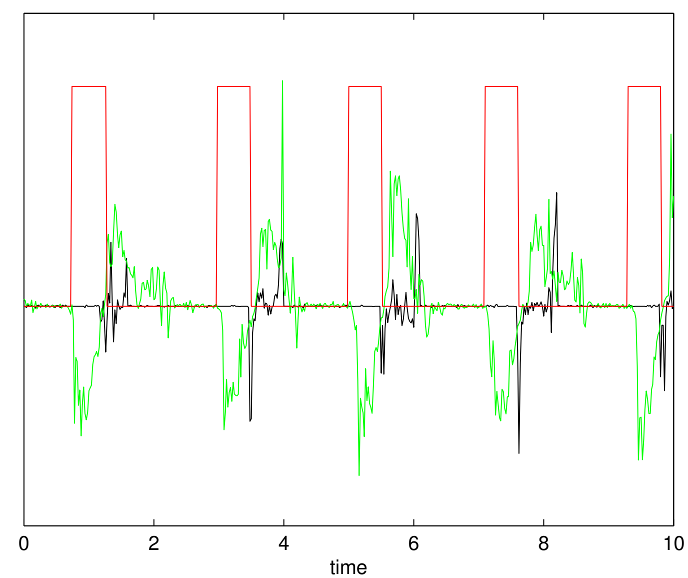
learning *how* to grasp





learning *how* to grasp



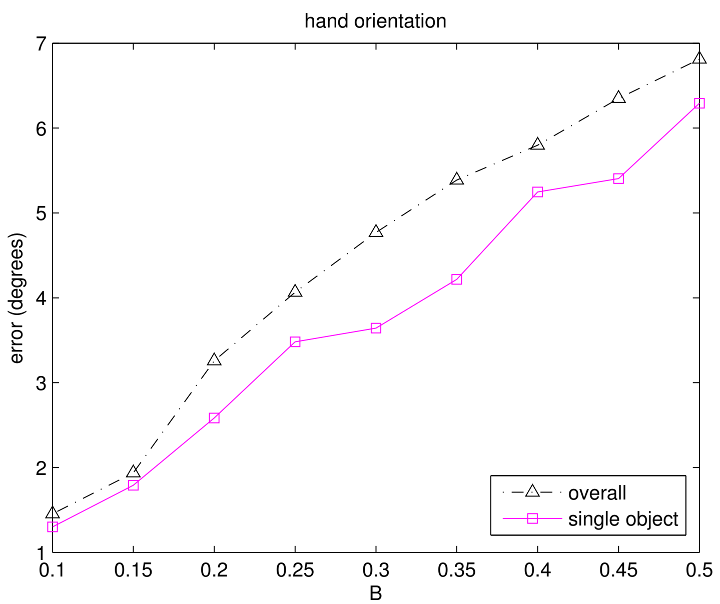


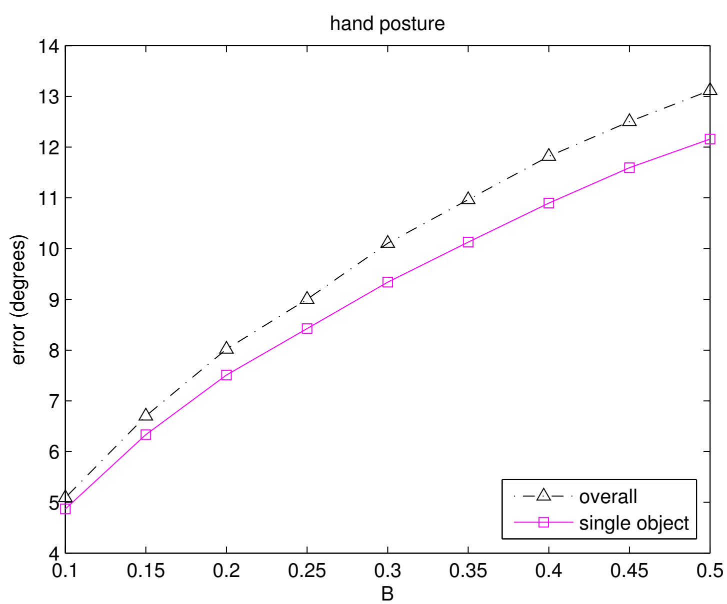
accuracy: how well do we guess the subject's grasping style? show the system the beginning of the trajectories (22 joints + 6 position/orientation) and figure out the final position

learning *how* to grasp

- allows for a precise prediction of
  - hand position: 0.5in
  - hand orientation: 2.5°
  - hand posture: 7.5°
- with a reasonable advance:
  - hand position: 200msecs.
  - hand orientation: 120msecs.
  - hand posture: 90-200msecs. (depending on the object)

learning *how* to grasp





affordance: knowing *a priori* the object will uniformly improve the error, at no additional computational overhead.

conclusions

- ml will teach your machines useful concepts such as human grasping, reaching, intentions
- those models can be trasferred to humanoid platforms (but not only to them)
- they can be employed in aiding the disabled
- lots of lovely maths behind it!

thank you!