Support Vector Machines: Eye movement classification in L1/L2 syntactic processing

Annelot de Rechteren van Hemert

Methodology & Statistics Linguistic Research

May 26th 2015

Outline

- Introduction
- Research Questions
- Experimental Design
- Support Vector Machines (SVM)
- Results
- Conclusion
- Discussion

Introduction

- L1 vs. L2 syntactic processing

  "Findings suggest that in the case of L2 sentence comprehension, syntactic processing may be fundamentally different from L1 processing, even when the learners perform like native speakers in offline comprehension tasks and even when the constructions under investigation are similar between the L1 and the target language."

  (Havik, 2009)

- Eye tracking

Research Questions

1. Can a comparison of sentence processing of structurally ambiguous Dutch which-questions by L1 German advanced learners of Dutch and Dutch native speakers provide insight into near-native syntactic L2 processing?
2. Can differences in 1) accuracy 2) reaction times in sentence processing be found between the L2 speakers and the native speakers?
3. Can differences in online sentence processing be found between the L2 speakers and the native speakers?

Experimental Design

- Eye tracking
- Picture selection task
- Six conditions:
  1. Simple SVO
  2. Simple OVS
  3. Complex SVO
  4. Complex OVS
  5. Embedded SVO
  6. Embedded OVS

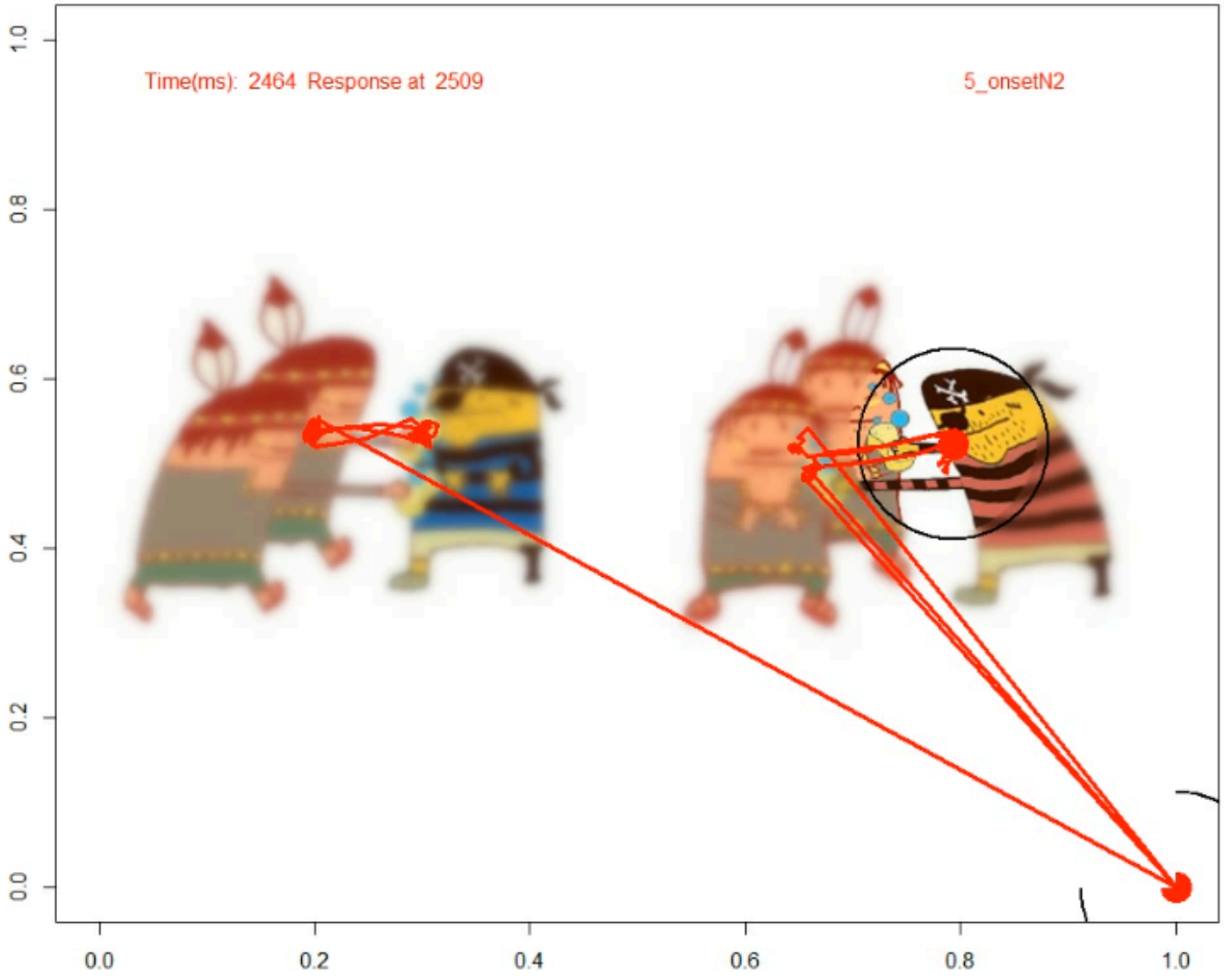
Examples

1. Welke boef wast de piloten? (Which robber$_{si}$ washes$_{si}$ the pilots$_{pl}$?)
2. Welke boef wassen de piloten? (Which robber$_{si}$ wash$_{pl}$ the pilots$_{pl}$? =which robber do the pilots wash?)
3. Jan vraagt welke boef de piloten wast. John asks which robber$_{sl}$ the pilots$_{pl}$ washes$_{sl}$ (=John asks which robber washes the pilots)
4. Jan vraagt welke boef de piloten wassen. (John asks which robber$_{sl}$ the pilots$_{pl}$ wash$_{pl}$)
5. Welke boef, vraagt Jan, wast de piloten? (Which robber$_{sl}$, asks John, washes$_{sl}$ the pilots$_{pl}$?)
6. Welke boef, vraagt Jan, wassen de piloten? (Which robber$_{sl}$, asks John, wash$_{pl}$ the pilots$_{pl}$? (=which robber, John asks, do the pilots wash?))

Example (Simple SVO)

Welke boef wast de piloten? (Which robber_{si} washes_{si} the pilots_{pl}?)

In real time..



Experimental Design

- Participants:
  - 17 native speakers of Dutch
    - Age M = 36.2 years, SD = 9.3
  - 18 native speakers of German, highly proficient in Dutch
    - Age M = 44.6 years, SD = 16.2
- C-test score
- Left or right handedness
- Length of residence (LoR) in the Netherlands (L2)
- Whether or not they spoke Dutch prior to emigrating (L2)

Experimental Design

- Participants:
  - 17 native speakers of Dutch
    - ~~Age M = 36.2 years, SD = 9.3~~
  - 18 native speakers of German, highly proficient in Dutch
    - ~~Age M = 44.6 years, SD = 16.2~~
- ~~C-test score~~
- ~~Left or right handedness~~
- ~~Length of residence (LoR) in the Netherlands (L2)~~
- ~~Whether or not they spoke Dutch prior to emigrating (L2)~~

Research Question

Amber's research question: can differences be found in native and near native L2 syntactic processing?

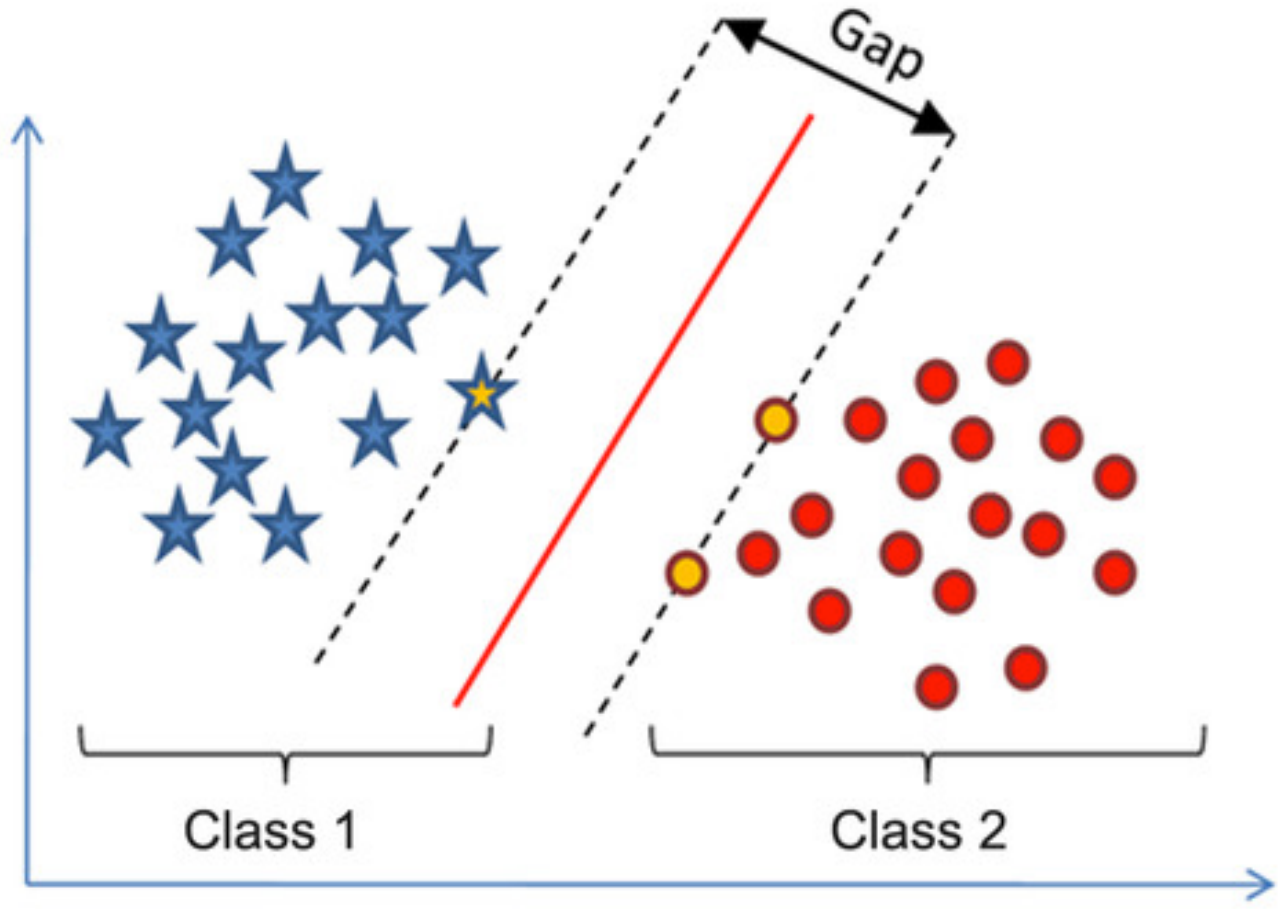
My research question: can we predict L1 and L2 syntactic processing purely on the basis of eye gaze data?

Support Vector Machines

"In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that **analyze data and recognize patterns, used for classification and regression analysis.** Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier." (Wikipedia)

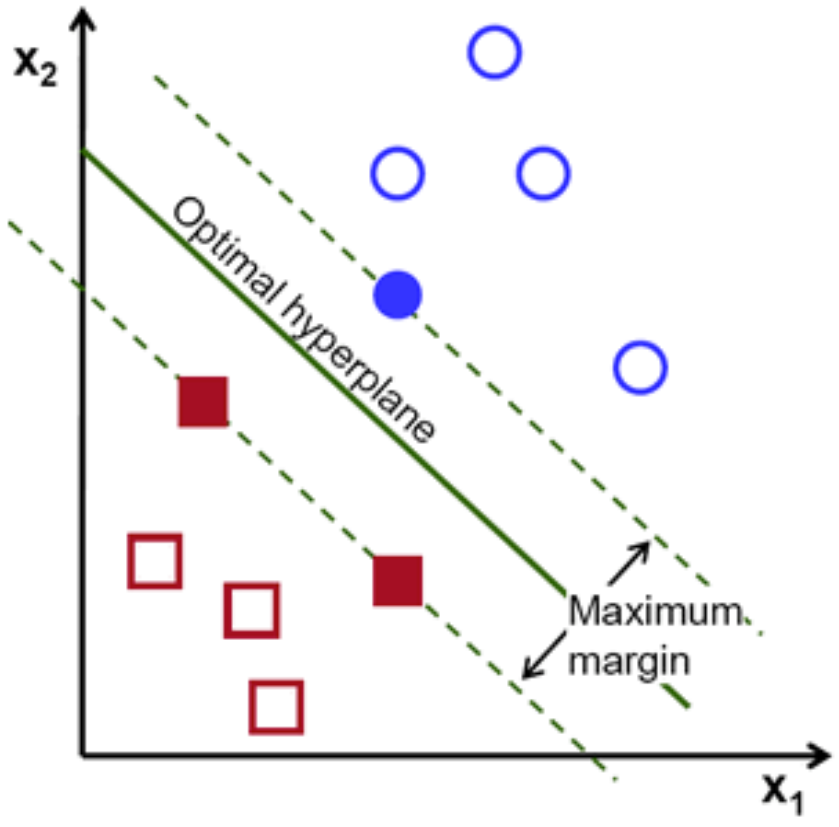
- Machine learning
- Supervised vs. unsupervised learning
- Training data, testing data

SVM: Most simple case



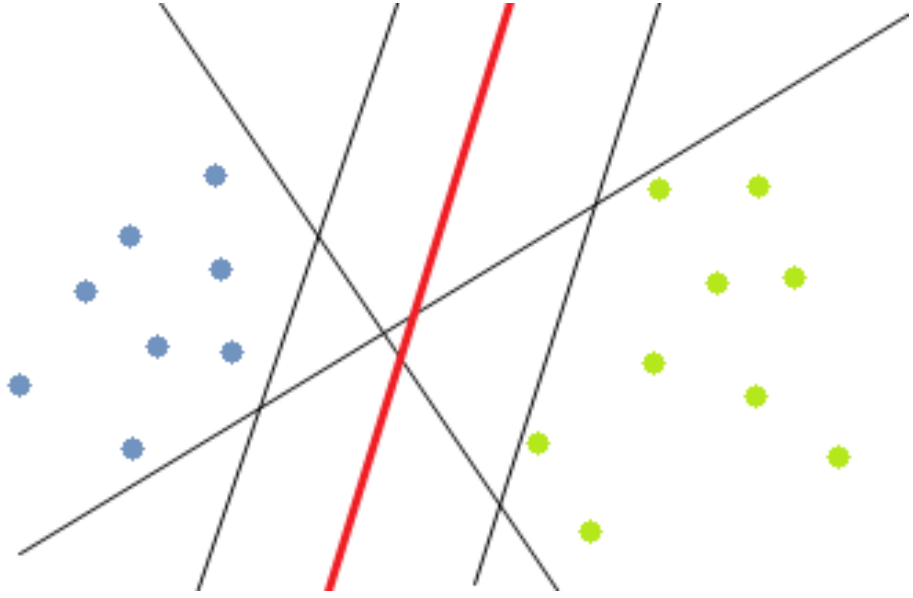
Picture from `http://diggdata.in/post/94066544971/support-vector-machine-without-tears`

SVM: Decision hyperplane, support vectors, margin



Picture from `http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html`

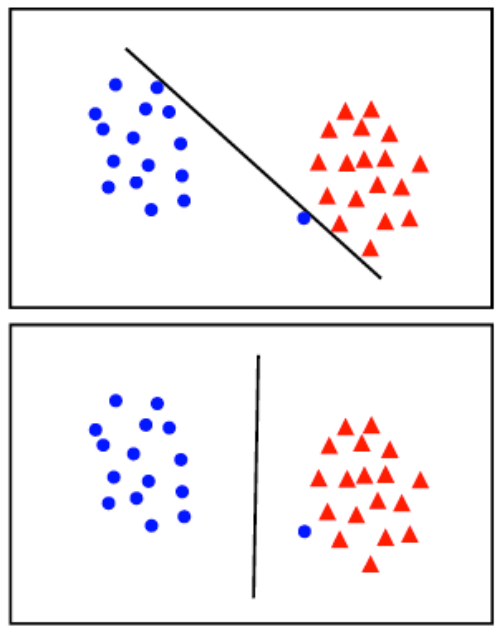
SVM: Hyperplane



Picture from http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Classification/SVM1

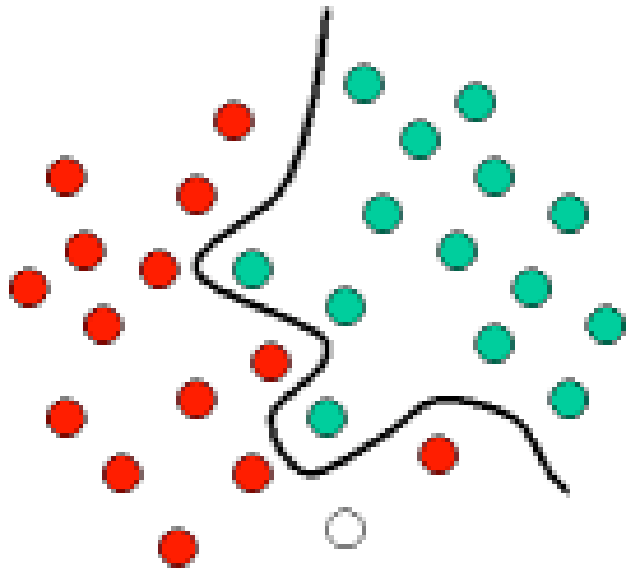
SVM: What about outliers?

- **C Parameter:**
  "In the support-vector networks algorithm one can control the trade-off between complexity of decision rule and frequency of error by changing the parameter C,..." (Cortes & Vapnik 1995)



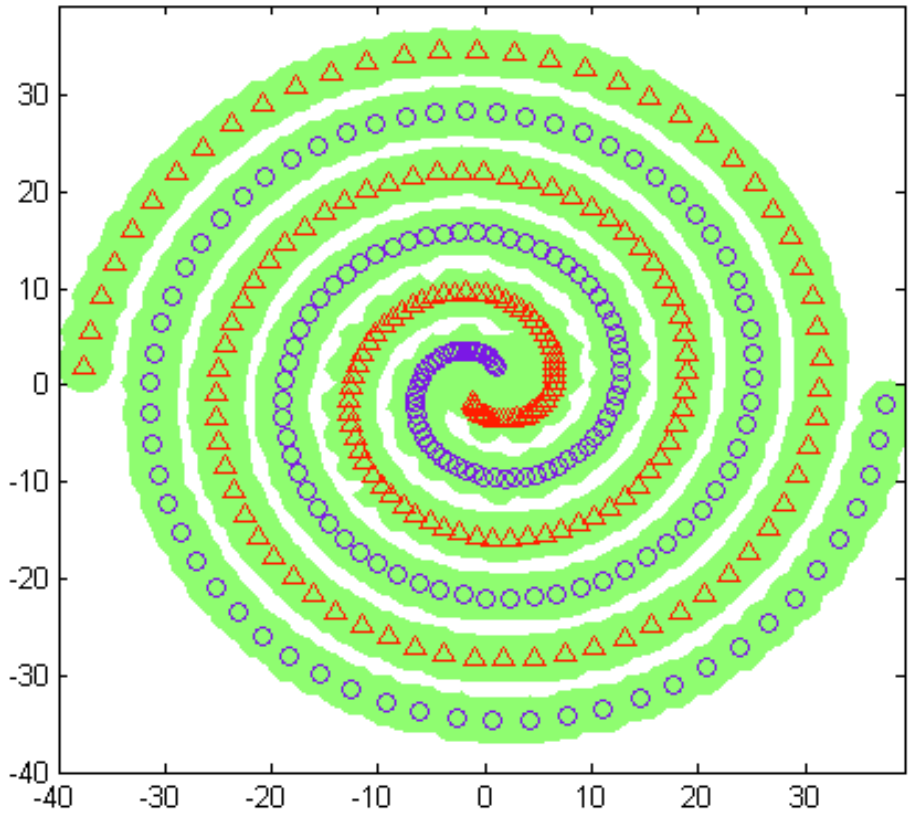
Picture from `http://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf`

SVM: Problematic?



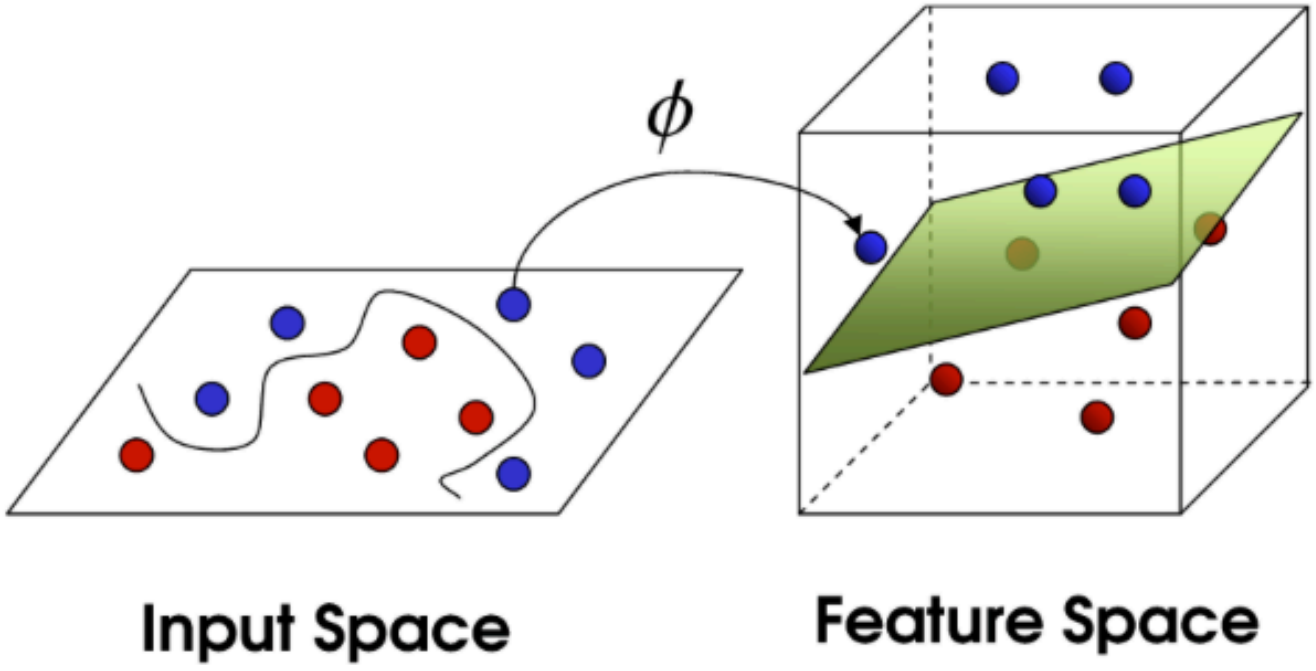
Picture from `http://www.statsoft.com/Textbook/Support-Vector-Machines`

SVM: Even more problematic?



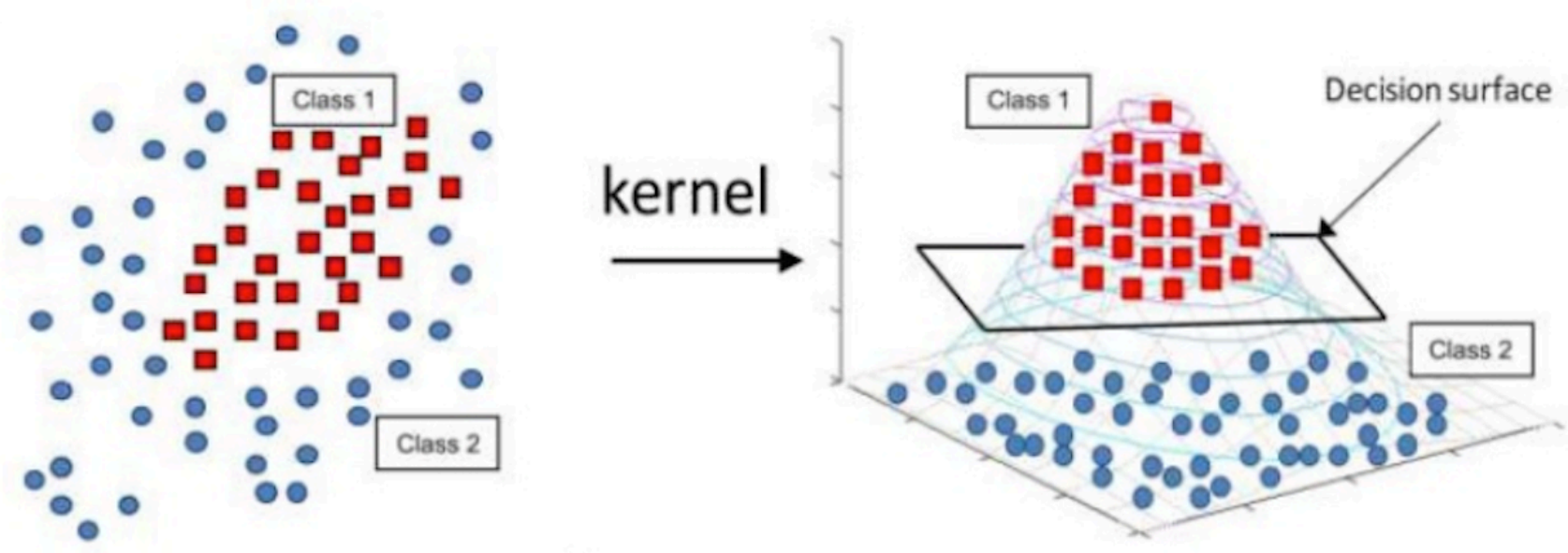
Picture from `http://en.wikipedia.org/wiki/Least_squares_support_vector_machine`

SVM: Transformation



Picture from http://www.reddit.com/r/MachineLearning/comments/15zrpp/please_explain_support_vector_machines_svm_like_i

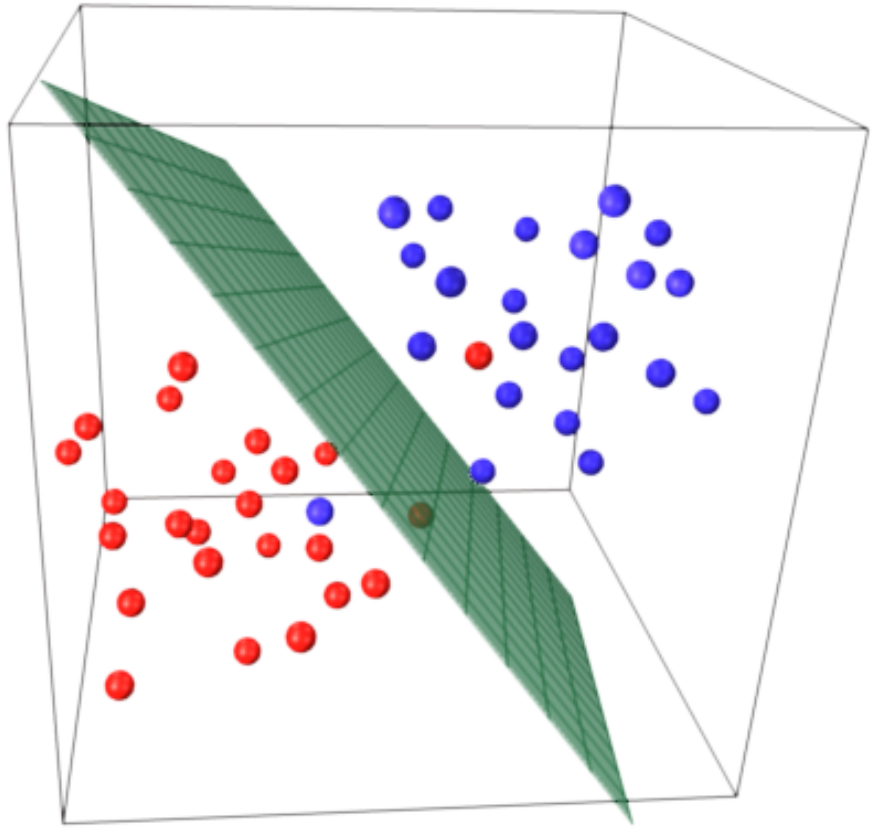
SVM: Solution



Picture from `http://www.slideshare.net/ankitksharma/svm-37753690`

Solution: RBF-Kernel

SVM



Picture from http://stackoverflow.com/questions/9480605/what-is-the-relation-between-the-number-of-support-vectors-and-training-data-and

SVM: Eye tracking classification

1. Initial data
2. Clustering
3. Training/testing
4. Results

SVM: Initial data

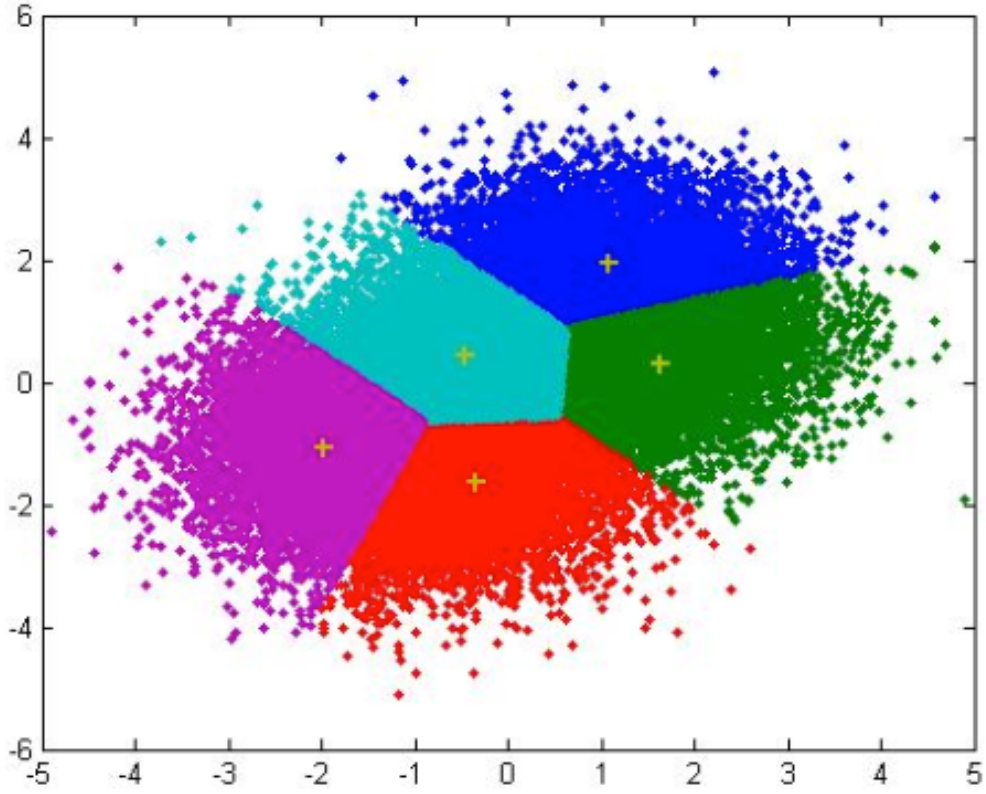
Subject	TrialId	Time(ms)	XGazePosition	YGazePosition
3	1	25	0.5052596	0.5360690
3	1	34	0.5090059	0.5263739
3	1	42	0.5049760	0.5401105
3	1	50	0.5001656	0.5444035
3	1	59	0.4986031	0.5377644
3	1	67	0.5035046	0.5335898
3	1	75	0.5074065	0.5296473
3	1	84	0.5083644	0.5333189
3	1	92	0.5116778	0.5260236
3	1	100	0.5088111	0.5170164
3	1	108	0.4979734	0.5211922
3	1	117	0.4978812	0.5260236
3	1	125	0.4995985	0.5203860
3	1	133	0.5026995	0.5250191
3	1	142	0.5161150	0.5389937
3	1	150	0.5162154	0.5281928
3	1	158	0.5055349	0.5212978
3	1	167	0.5062645	0.5274062
3	1	175	0.5000217	0.5353294
3	1	183	0.4923982	0.5300978
3	1	192	0.5012991	0.5222415
3	1	200	0.5076168	0.5260516
3	1	208	0.5026959	0.5333123
3	1	217	0.4994553	0.5336850
3	1	225	0.5041866	0.5280103
3	1	233	0.5027142	0.5328726
3	1	242	0.5029675	0.5346212
3	1	250	0.5013313	0.5354313
3	1	258	0.4982771	0.5449564
3	1	267	0.4965771	0.5432884
3	1	275	0.5031588	0.5277694
3	1	283	0.5141580	0.5251240
3	1	292	0.5123628	0.5301459
3	1	300	0.5075457	0.5371867
3	1	308	0.5070092	0.5377107
3	1	317	0.5087959	0.5320474

Dimensionality reduction: "is the mapping of data to a lower dimensional space such that uninformative variance in the data is discarded, or such that a subspace in which the data lives is detected." (Burges 2009)

- **Time:** *Onset sentence, Onset Noun 1, Onset Verb, Onset Noun 2*
- **Space:** *K-means clustering*

K-means clustering

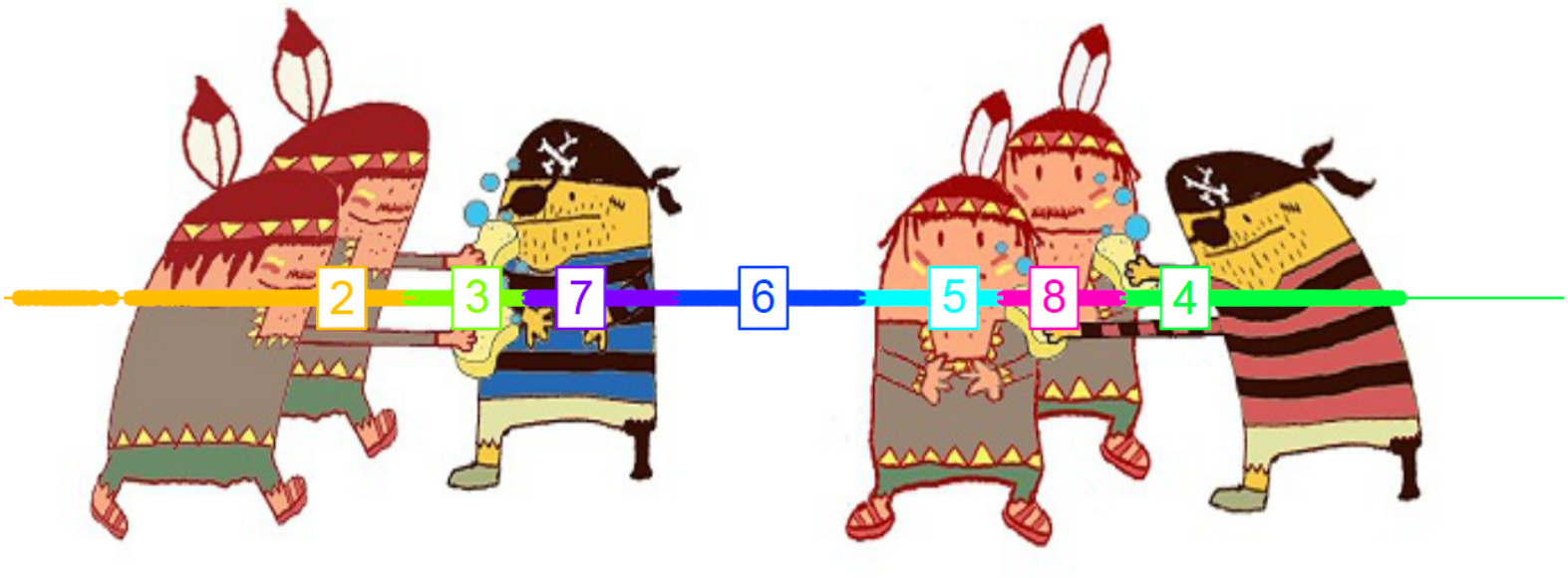
"K-means algorithm is for clustering, that is, for finding groups in the data, where the groups are represented by their centers, which are the typical representatives of the groups." Alpaydin (2010)



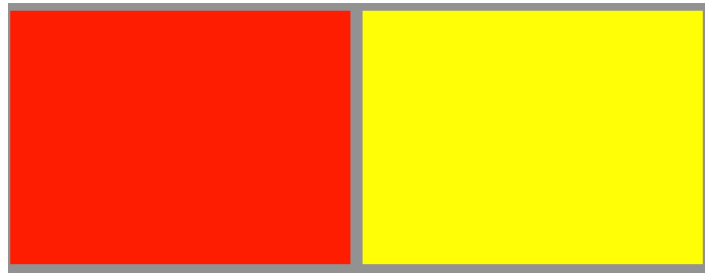
Picture from `http:`

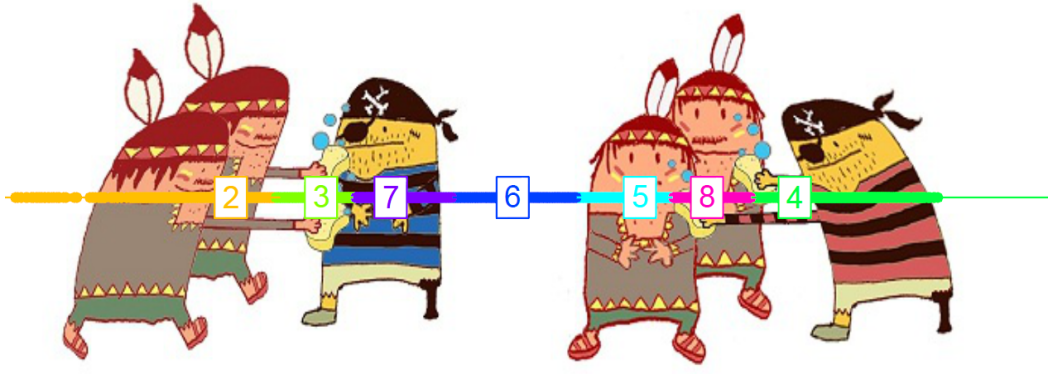
`//www.mathworks.com/matlabcentral/fileexchange/19344-efficient-k-means-clustering-using-jit`

K-means clustering: Eye gaze clusters



K-means clustering: Eye gaze clusters





Input data for SVM

		Noun 1				Verb				Noun 2	Label
		R_1	R_2	..	R_n	R_1	R_2	..	R_n	..	
S1	Trial1	0,1	0,3	..	0,1	0,2	0,1	..	0,2	..	L1
	Trial2	0,3	0,2	..	0,3	0,3	0,1	..	0,1	..	L1
	Trial3	0,2	0,1	..	0,2	0,1	0,2	..	0,3	..	L1
	TrialN	..	..	..	..	..	..	..	..	..	L1
S2	Trial1	0,2	0,2	..	0,1	0,2	0,2	..	0,1	..	L2
	Trial2	0,1	0,2	..	0,3	0,1	0,2	..	0,1	..	L2
	Trial3	0,2	0,3	..	0,2	0,2	0,3	..	0,2	..	L2
	TrialN	..	..	..	..	..	..	..	..	..	L2

Training/testing

- **Leave-One-Out-Cross-Validation (LOOCV)**

Training/testing: Results

- **Leave-One-Out-Cross-Validation (LOOCV)**
  one L1 speaker vs. one L2 speaker: 98%
  two L1 speakers vs. two L2 speakers: 90%
  four L1 speakers vs. four L2 speakers: 85%
  fifteen L1 speakers vs. fifteen L2 speakers: 70%

Training/testing

- **Leave one subject out**

Training/testing: Results

- **Leave one subject out**
  Only 58%..

Conclusion?

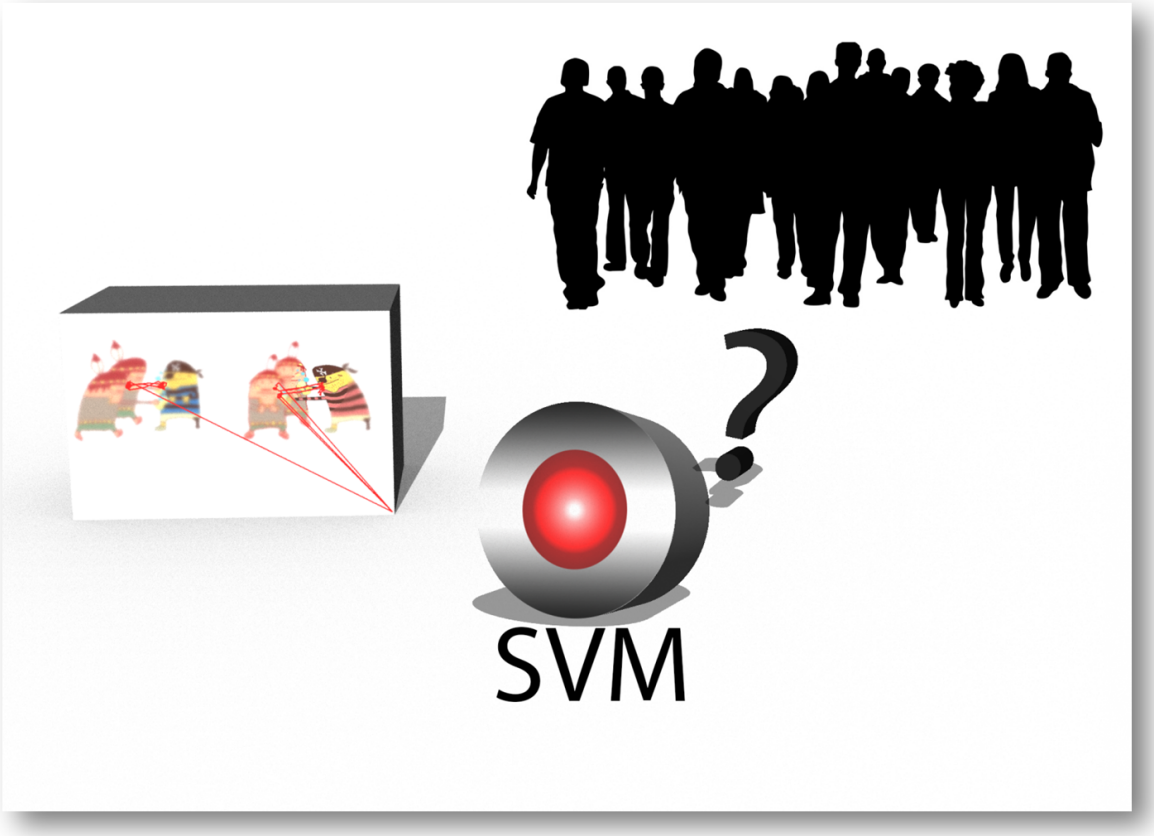
- Speakers may have unique eye movements during syntactic processing.

Does the SVM learn to recognize specific people (of which the label is known) rather than general differences between L1 and L2 speakers?

Training/testing

- Label ~~L1 / L2~~ Subject ID
- Two methods:
  1) 1 vs. all
  2) Multiclass SVM

1 vs. all



Is it subject x or not?

Accuracy: 80%

Multiclass SVM

To which of the 30 participants does this gaze data belong?

Accuracy: 40%
(Actually more impressive than the previous method..)

Conclusion

- Eye movements (alone) do not provide a lot of information in L1/L2 syntactic processing: every person has a remarkably unique eye gaze signature.
- Machine learning is a powerful tool for bottom up understanding of our data.

Questions?


SUPPORT
VECTOR
MACHINES
END
DUALITY
GAP
REPEA
POWE
LAWS
BAYSIANS
AGAINST
DISCRIMINATION
FREE
BAN
GENETIC
ALGORITHMS

Discussion

Significance testing vs. prediction:
What are the differences, what method would you prefer?

References

Alpaydin, E. (2010). Introduction to machine learning. MIT press.

Burges, C. (2009). Dimension reduction: A guided tour. Machine Learning, 2(4), 275-365.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing SubjectObject Ambiguities in the L2: A SelfPaced Reading Study With German L2 Learners of Dutch. Language Learning, 59(1), 73-112.

Some useful material

- An Idiots guide to Support vector machines (SVMs) by R. Berwick (2003):
  `http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf`
- K-means clustering simplified by A. Swartz (2013):
  `http://www.webzeest.com/article/715/kmeans-clustering-simplified`
- Kernels Part 1: What is an RBF Kernel? Really? by C. Martin (2012):
  `https://charlesmartin14.wordpress.com/2012/02/06/kernels_part_1/`

Many thanks to Amber Nota and Atty Schouwenaars for letting me use their data.