Agent-based Models
building reliable, intelligible classifiers

William La Cava

Kourosh Danai
Lee Spector*

Sara Silva[+]
Leonard Vanneschi[&]
Mauro Castelli[&]

[+]University of Lisbon, PT
[&]Nova University, PT
*Hampshire College, USA

Outline

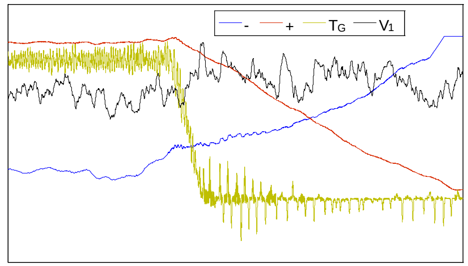
- My research - briefly
- Lisbon
- Multiclass Classification
- Bald Eagle Agent Modelling
  - Preliminary Results

My Research

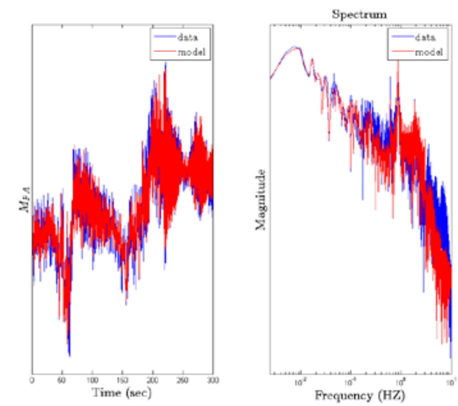
- Automatic model form development and adaptation
  - Machine learning techniques for understanding complex systems

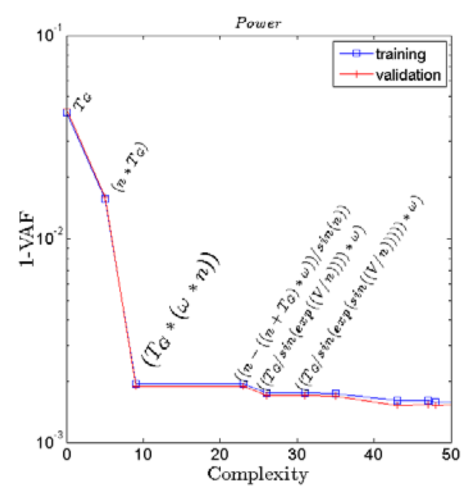
- Wind Energy Applications
  - Wind Turbine Dynamics
  - Vortex Induced Vibration
  - Bald Eagle Behavior

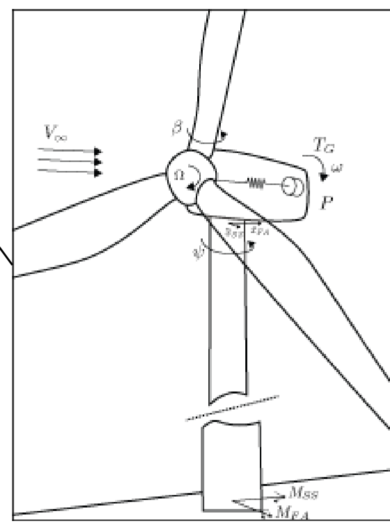
Wind Turbine Dynamics



Model Development







Control Design

NREL CART 3
Closed-Loop Data
Collection
Source: nrel.gov/wind

Wind Turbine Dynamics

	Training / Validation VAF (%)				
	ELGP	Multiple Regression	20^{th}-order ARX	ELGP	NARX-NN
Ω	98.7 / 98.7	91.9 / 91.9	71.0 / 71.0	100.0 / 99.9	99.9 / 99.8
ω	98.6 / 98.6	92.0 / 91.9	69.0 / 69.0	100.0 / 99.9	99.9 / 99.8
M_{FA}	74.2 / 74.4	31.5 / 32.2	25.6 / 25.6	98.7 / 94.9	98.6 / 94.9
M_{SS}	72.7 / 72.2	19.6 / 20.4	0.0 / 0.0	97.6 / 89.9	97.3 / 90.6
P	99.9 / 99.9	99.7 / 99.7	99.6 / 99.6	- / -	- / -

$$\omega = n_1(\omega_{k-1} + e^{(n_2\omega_{k-1})} T_G)$$

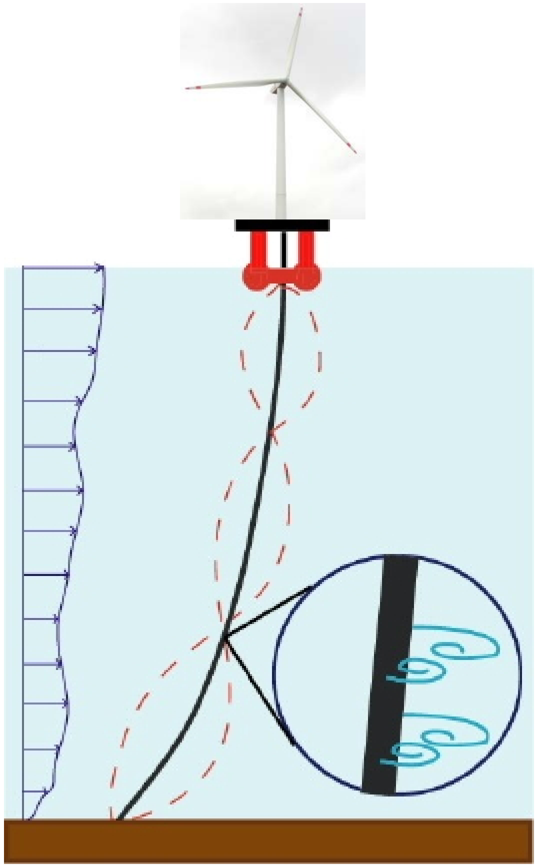
$$\Omega = n_1(\Omega_{k-1} - \sin(\frac{n_2}{t})\sin(n_3\frac{V_{k-1}T_{G_{k-1}}}{\Omega_{k-1}})$$

$$M_{FA} = n_1(M_{FA_{k-1}} + n_2\sin(T_G)(\ddot{x}_{FA} - \ddot{x}_{FA_{k-1}})/V)$$

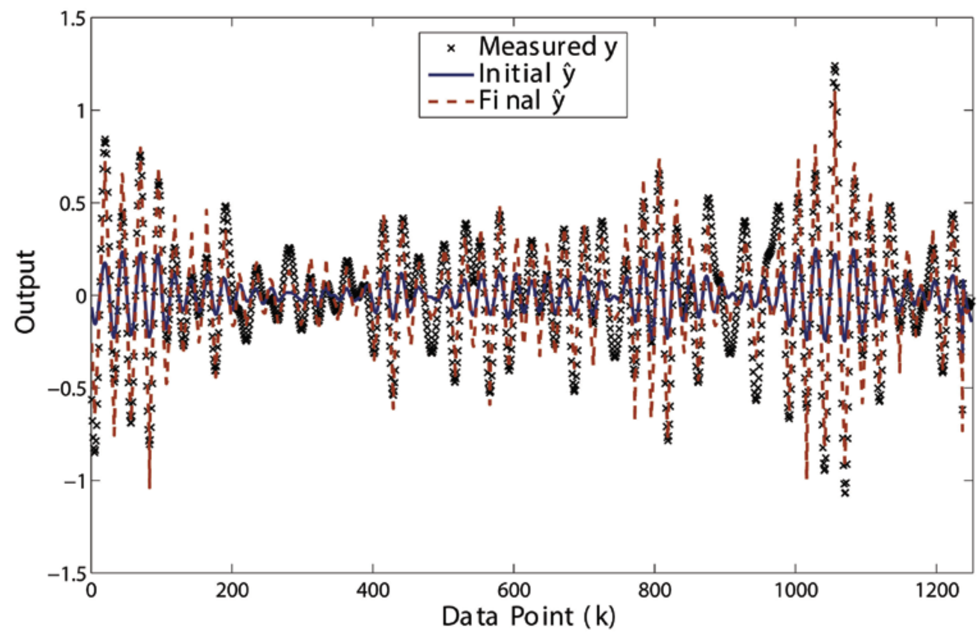
$$M_{SS} = n_1\left(M_{SS_{k-1}} + n_2\sin(n_3\psi)(\ddot{x}_{FA} - \ddot{x}_{FA_{k-1}})\right)$$

La Cava, William, Kourosh Danai, Lee Spector, Paul Fleming, Alan D. Wright, and Matthew Lackner (2015). “Automatic identication of wind turbine models using evolutionary multi-objective optimization". *Renewable Energy.*

Vortex Induced Vibration



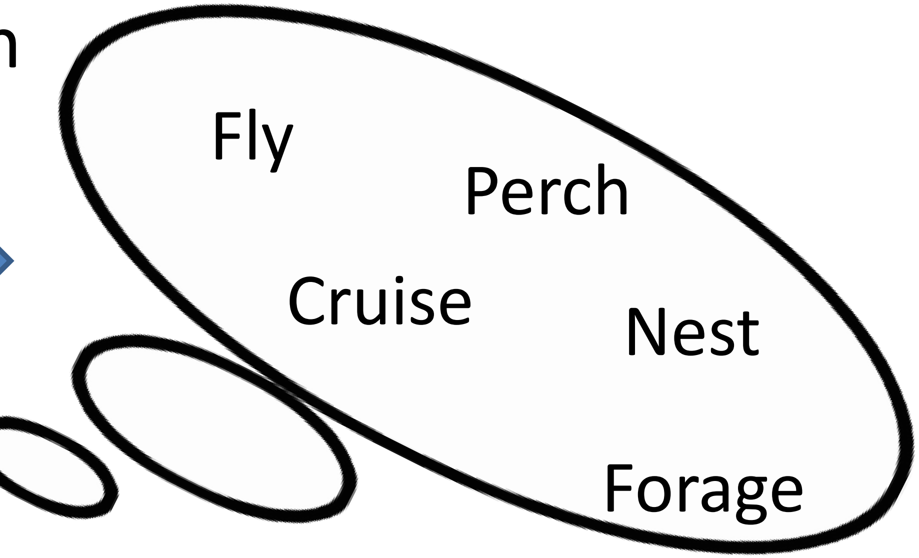
Model adaptation



La Cava, William G. and Kourosh Danai (2015b). “Gradient-based adaptation of continuous dynamic model structures". *International Journal of Systems Science* 47 (1), pp. 249-263. doi:10.1080/00207721.2015.1069905.

Bald Eagle Behavior

- Agent-based Modeling
- Discrete behaviors -> classes
- Multi-class classification




Toulouse
Carcassonne
Marseille
A Coruña
Gijón
Bilbao
San Sebastián
Oviedo
Santiago de Compostela
Vitoria-Gasteiz
León
Pamplona
Andorra
Girona
Pontevedra
Ourense
Vigo
Zaragoza
Valladolid
Barcelona
Porto
Salamanca
Madrid
Castellón de la Plana
Balearic Sea
Toledo
Portugal
Valencia
Palma
Spain
Lisbon
Badajoz
Benidorm
Alicante
Murcia
Córdoba
Algiers
الجزائر
Seville
Granada
Albufeira
Almería
Faro
Málaga
Gibraltar
Alboran Sea
Oran
وهران
Djelfa
الجلفة

LISBON

University of Lisbon

- LabMAg
  - Laboratory of agent modeling
  - Institute for Complexity Sciences
- Nova University

Breakfast

M2GP and M3GP

A Multi-dimensional Genetic Programming Approach for Multi-class Classification Problems

Vijay Ingalalli[1], Sara Silva[1,2,3], Mauro Castelli[4], and Leonardo Vanneschi[4]

[1] INESC-ID, IST, University of Lisbon, 1000-029 Lisbo
[2] LabMAg, FCUL, University of Lisbon, 1749-016 Lisb
[3] CISUC, Universidade de Coimbra, 3030-290 Coimbr
[4] ISEGI, Universidade Nova de Lisboa, 1070-312 Lisbo

Abstract. Classification problems are of profound interest for t
ing community as well as to an array of application fields. Ho
classification problems can be very complex, in particular wh
classes is high. Although very successful in so many applicatio
regarded as a good method to perform multi-class classification
present a novel algorithm for tree based GP, that incorporates
representation of the solution space in higher dimensions. Th
foundations on addressing multi-class classification problems
may lead to further research in this direction. We test the ne
large set of benchmark problems from several different source
competitiveness against the most successful state-of-the-art cla

M3GP – Multiclass Classification with GP

Luis Muñoz[1], Sara Silva[2,3,4], and Leonardo Trujillo[1]

[1] Tree-Lab, Instituto Tecnológico de Tijuana, B.C., México
[2] BioISI – Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, Portugal
[3] NOVA IMS, Universidade Nova de Lisboa, 1070-312 Lisboa, Portugal
[4] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
sara@fc.ul.pt, {lmunoz, leonardo.trujillo}@tectijuana.edu.mx

Abstract. Data classification is one of the most ubiquitous machine learning tasks in science and engineering. However, Genetic Programming is still not a popular classification methodology, partially due to its poor performance in multiclass problems. The recently proposed M2GP - Multidimensional Multiclass Genetic Programming algorithm achieved promising results in this area, by evolving mappings of the p-dimensional data into a d-dimensional space, and applying a minimum Mahalanobis distance classifier. Despite good performance, M2GP employs a greedy strategy to set the number of dimensions d for the transformed data, and fixes it at the start of the search, an approach that is prone to locally optimal solutions. This work presents the M3GP algorithm, that stands for M2GP with multidimensional populations. M3GP extends M2GP by allowing

- progress

Why use GP?

- Lots of classifier systems already exist (SVMs, MLPs, Decision trees, Random Forests)
- GP isn’t traditionally good at multi-class classification problems
- *Less assumptions*
- *Intelligibility*
- *Generalizability*
- *Feature selection*

Lunch

Insight

- Use GP to generate features
  - original attributes -> distinct distributions
- Mahalanobis distance

$$D_k = \sqrt{(X - M_k) C_k^{-1} (X - M_k)^T}$$

  - each transformed attribute vector to each class' attribute distribution in new space
- Assign class with lowest D_k
- *Classifier: Equations, M, C*

Challenges

- The results were good, but not great
  - Tied or worse than Random Forests & other methods
- *Not making use of more recent advances to other parts of GP*
- *Complex tree representation*

Lots of methods

Scheme	Name	Description
tournament selection	tourn	size 2 standard tournaments.
pareto survival	ps	multi-objective age-fitness pareto survival via SPEA2
lexicase selection	lex	parents are selected using lexicase selection, where the cases are each data sample
class-based pareto survival	ps5	multiple fitnesses are assigned, one for each class label. age is also used and pareto survival is conducted via SPEA2
lexicase selection with classes	lexc	parents are selected using lexicase selection, where the cases are the aggregate error on each class label

I also tried three different settings for the genetic operators:

Genetic Operators Name	Description
uniform	uniform alternation & point mutation
st	subtree crossover and mutation, specialized according to m3gp
stp	subtree crossover and mutation plus pruning of the best individual each generation

Test Problems

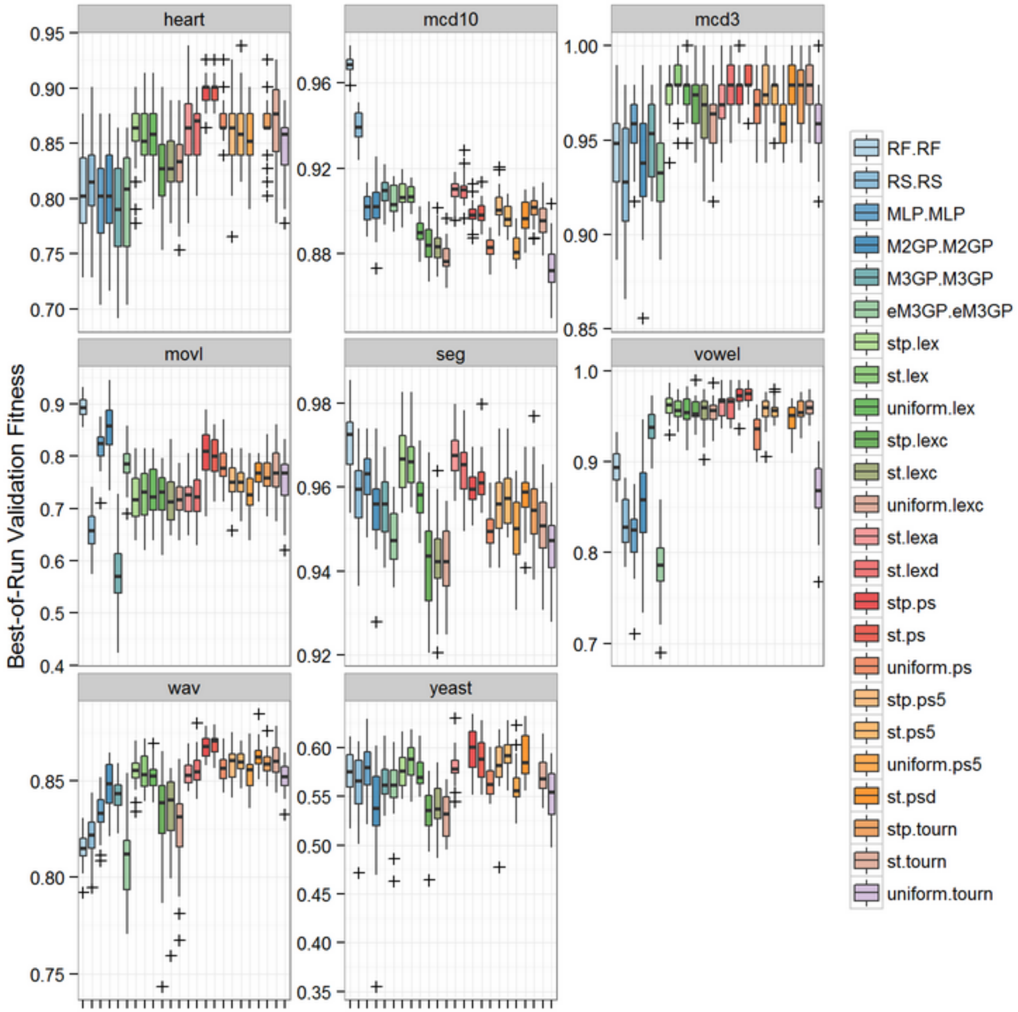
- UCI Repository data sets

Data set	heart	mcd10	mcd3	movl	seg	vowel	wav	yeast
Classes	2	10	3	15	7	11	3	10
Attributes	13	6	6	90	19	13	40	8
Cases	270	6798	322	360	2310	990	5000	1484

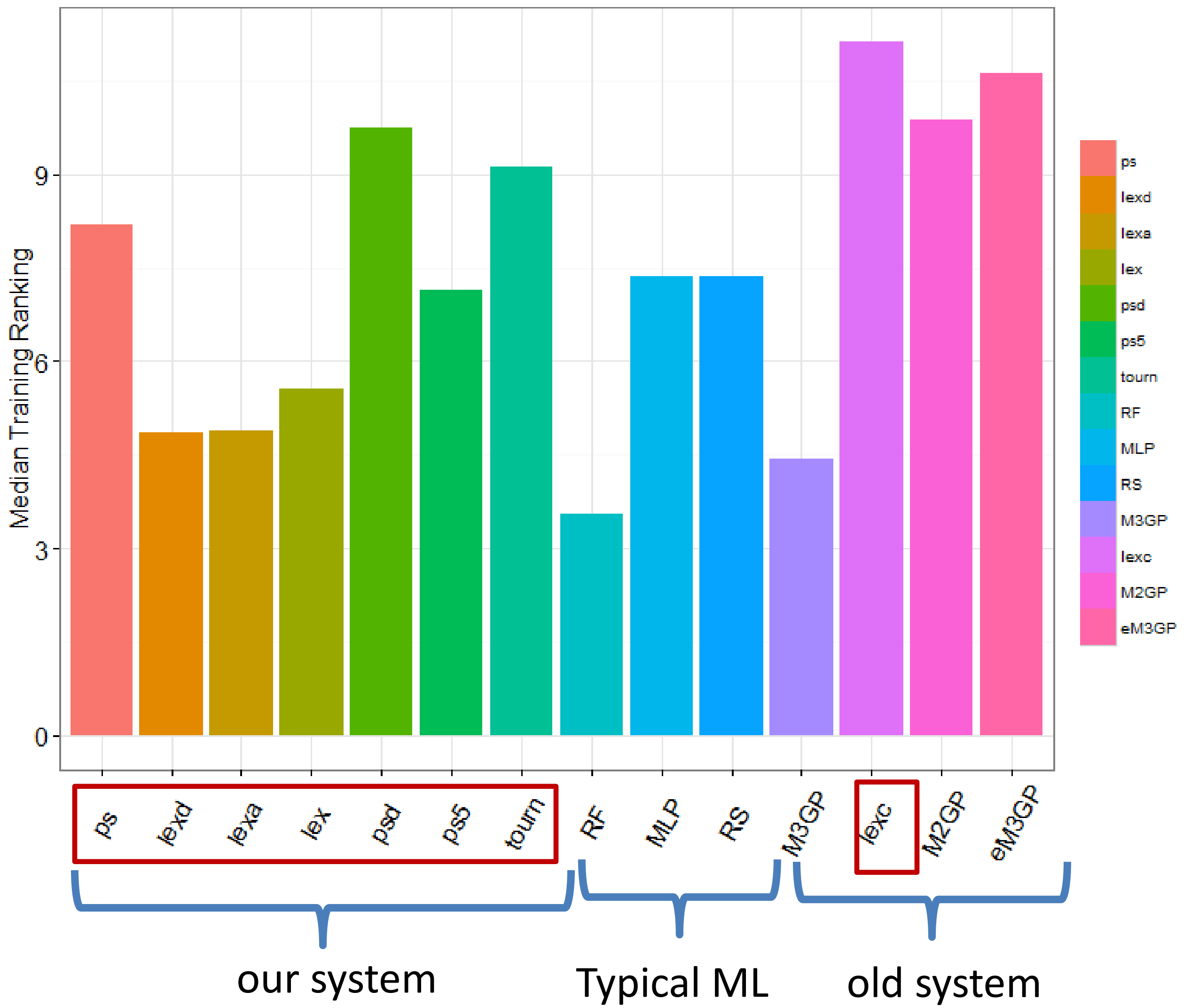
Lots of Results

Validation Fitness

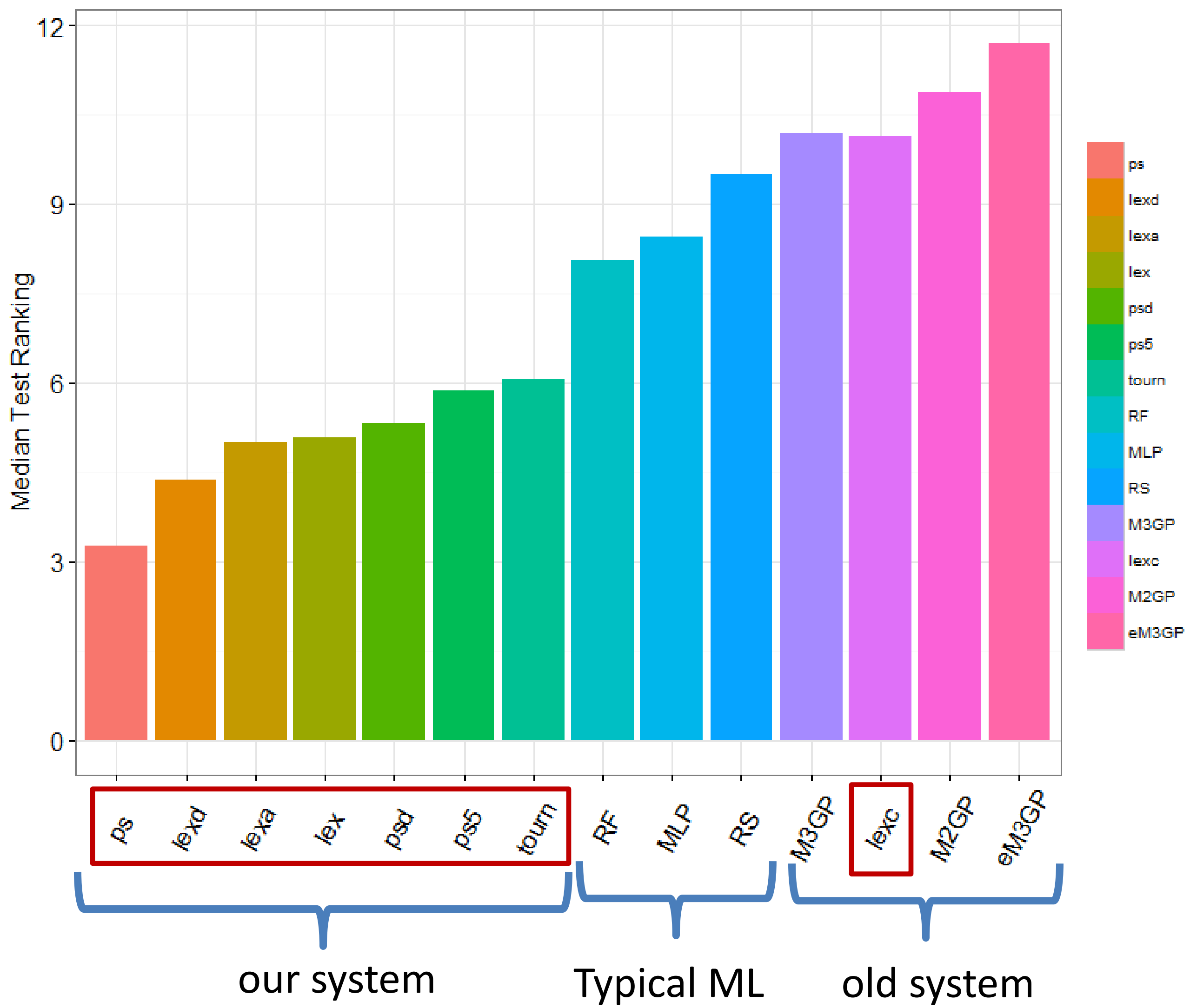
When we look at validation fitness, M3GP is often not the best:



Overall Results



Overall Results



Bald Eagle

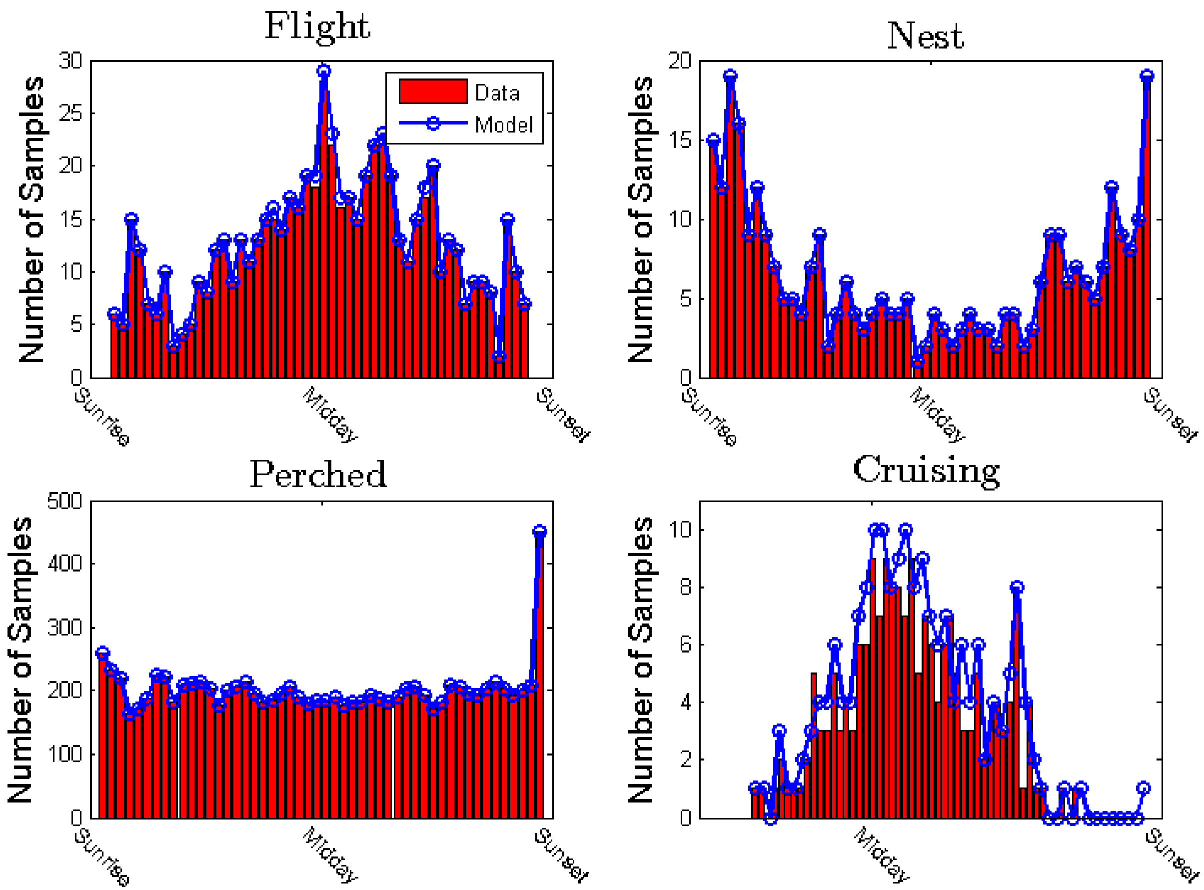
- 11,537 attributes
- 4 behaviors
  - Flight
  - Perched
  - Cruise
  - Nest

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	date	id	sex	datetime	serial	lat	long	hdop	fix	speed	alt	year	deploy_se	notes	long_utm	lat_utm	sunrise	solarnoon
2	9/23/2013	Crooked	male	9/23/2013 6:39	72187	45.34035	-68.4469	1.9	3	21.6	118	2013	1	NA	543330	5020909	9/23/2013 5:22	9/23/2013
3	9/23/2013	Crooked	male	9/23/2013 8:38	72187	45.35428	-68.4523	2.4	2	1.8	29	2013	1	NA	542900	5022453	9/23/2013 5:22	9/23/2013
4	9/23/2013	Crooked	male	9/23/2013 10:37	72187	45.33765	-68.4388	3.6	3	0.7	90	2013	1	NA	543970	5020614	9/23/2013 5:22	9/23/2013
5	9/23/2013	Crooked	male	9/23/2013 12:04	72187	45.35114	-68.477	2.5	3	16.9	128	2013	1	NA	540964	5022092	9/23/2013 5:22	9/23/2013
6	9/23/2013	Crooked	male	9/23/2013 12:07	72187	45.351	-68.483	1.6	3	20.5	110	2013	1	NA	540495	5022073	9/23/2013 5:22	9/23/2013
7	9/23/2013	Crooked	male	9/23/2013 12:24	72187	45.34049	-68.4469	1.2	3	0.5	109	2013	1	NA	543335	5020924	9/23/2013 5:22	9/23/2013
8	9/23/2013	Crooked	male	9/23/2013 12:39	72187	45.34194	-68.4439	1.3	3	16.5	184	2013	1	NA	543565	5021087	9/23/2013 5:22	9/23/2013
9	9/23/2013	Crooked	male	9/23/2013 12:55	72187	45.35523	-68.4742	3	3	0.3	100	2013	1	NA	541184	5022548	9/23/2013 5:22	9/23/2013
10	9/23/2013	Crooked	male	9/23/2013 13:10	72187	45.34717	-68.4861	2.6	3	0.4	65	2013	1	NA	540255	5021646	9/23/2013 5:22	9/23/2013
11	9/23/2013	Crooked	male	9/23/2013 13:26	72187	45.34716	-68.4857	1.9	3	0	89	2013	1	NA	540291	5021645	9/23/2013 5:22	9/23/2013
12	9/23/2013	Crooked	male	9/23/2013 13:40	72187	45.3471	-68.4856	1.1	3	0	96	2013	1	NA	540299	5021639	9/23/2013 5:22	9/23/2013
13	9/23/2013	Crooked	male	9/23/2013 13:56	72187	45.34705	-68.4856	2.1	3	0	78	2013	1	NA	540296	5021633	9/23/2013 5:22	9/23/2013
14	9/23/2013	Crooked	male	9/23/2013 14:11	72187	45.34716	-68.4856	1.5	3	0	101	2013	1	NA	540301	5021645	9/23/2013 5:22	9/23/2013
15	9/23/2013	Crooked	male	9/23/2013 14:27	72187	45.34697	-68.4857	1.4	3	0	82	2013	1	NA	540289	5021624	9/23/2013 5:22	9/23/2013
16	9/23/2013	Crooked	male	9/23/2013 14:43	72187	45.3471	-68.4857	2.4	3	0	108	2013	1	NA	540291	5021638	9/23/2013 5:22	9/23/2013
17	9/23/2013	Crooked	male	9/23/2013 14:59	72187	45.34716	-68.4856	2	3	0	105	2013	1	NA	540293	5021645	9/23/2013 5:22	9/23/2013
18	9/23/2013	Crooked	male	9/23/2013 15:14	72187	45.34703	-68.4856	2.5	3	0	90	2013	1	NA	540300	5021631	9/23/2013 5:22	9/23/2013
19	9/23/2013	Crooked	male	9/23/2013 15:29	72187	45.34732	-68.4855	2.9	3	0	121	2013	1	NA	540304	5021663	9/23/2013 5:22	9/23/2013
20	9/23/2013	Crooked	male	9/23/2013 15:45	72187	45.34678	-68.485	1.4	3	0	69	2013	1	NA	540342	5021603	9/23/2013 5:22	9/23/2013
21	9/23/2013	Crooked	male	9/23/2013 16:01	72187	45.34657	-68.4852	1.6	3	0	78	2013	1	NA	540331	5021579	9/23/2013 5:22	9/23/2013
22	9/23/2013	Crooked	male	9/23/2013 16:17	72187	45.34663	-68.4851	1.4	3	0	92	2013	1	NA	540335	5021586	9/23/2013 5:22	9/23/2013
23	9/23/2013	Crooked	male	9/23/2013 16:34	72187	45.34633	-68.4849	3.7	3	0.8	134	2013	1	NA	540355	5021553	9/23/2013 5:22	9/23/2013
24	9/23/2013	Crooked	male	9/23/2013 16:48	72187	45.34629	-68.4853	1.1	3	0	84	2013	1	NA	540325	5021549	9/23/2013 5:22	9/23/2013
25	9/23/2013	Crooked	male	9/23/2013 17:03	72187	45.34622	-68.4851	1.4	3	0	97	2013	1	NA	540337	5021541	9/23/2013 5:22	9/23/2013
26	9/23/2013	Crooked	male	9/23/2013 17:19	72187	45.34635	-68.4852	1.3	3	0.4	94	2013	1	NA	540331	5021555	9/23/2013 5:22	9/23/2013
27	9/23/2013	Crooked	male	9/23/2013 17:35	72187	45.34627	-68.4851	1.4	3	0	102	2013	1	NA	540339	5021547	9/23/2013 5:22	9/23/2013

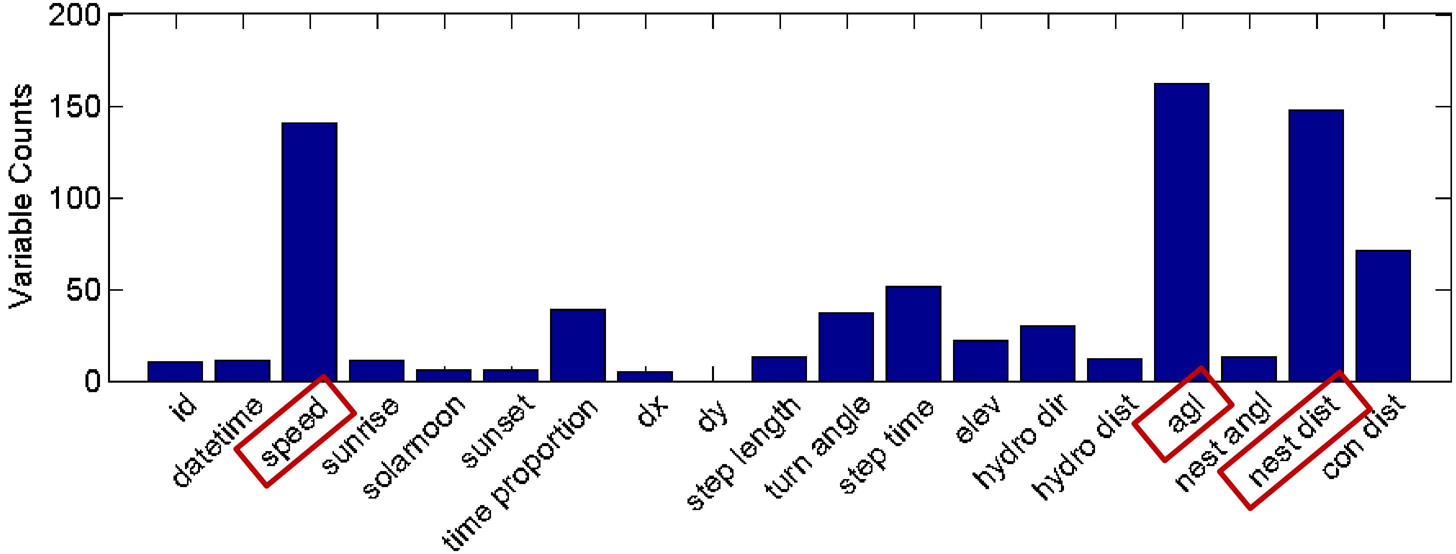
Results

Median classifier accuracy: 98.87%

Best classifier accuracy : 99.82%

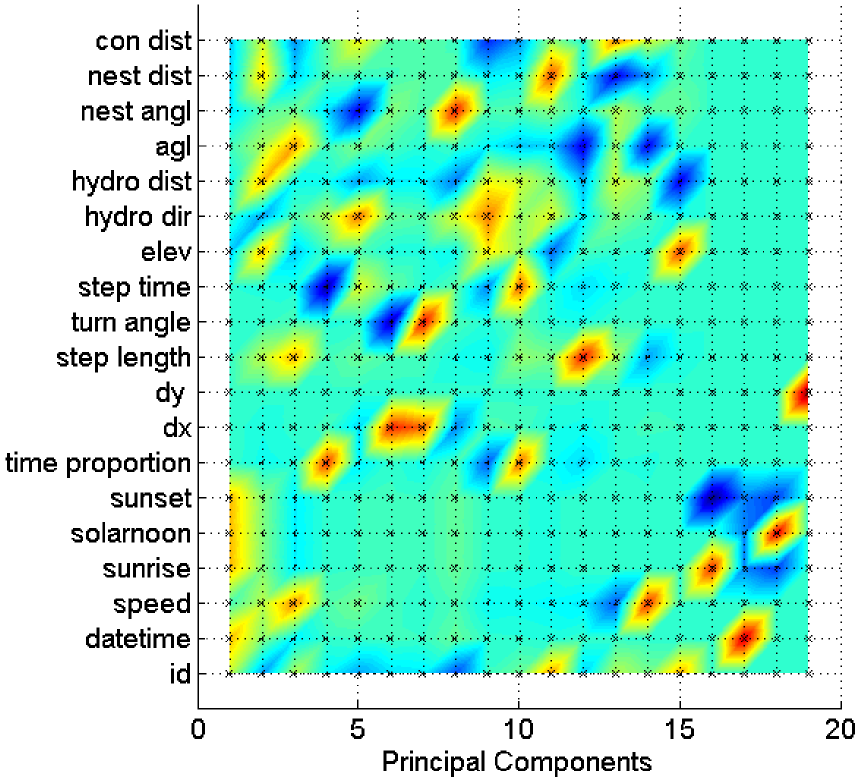


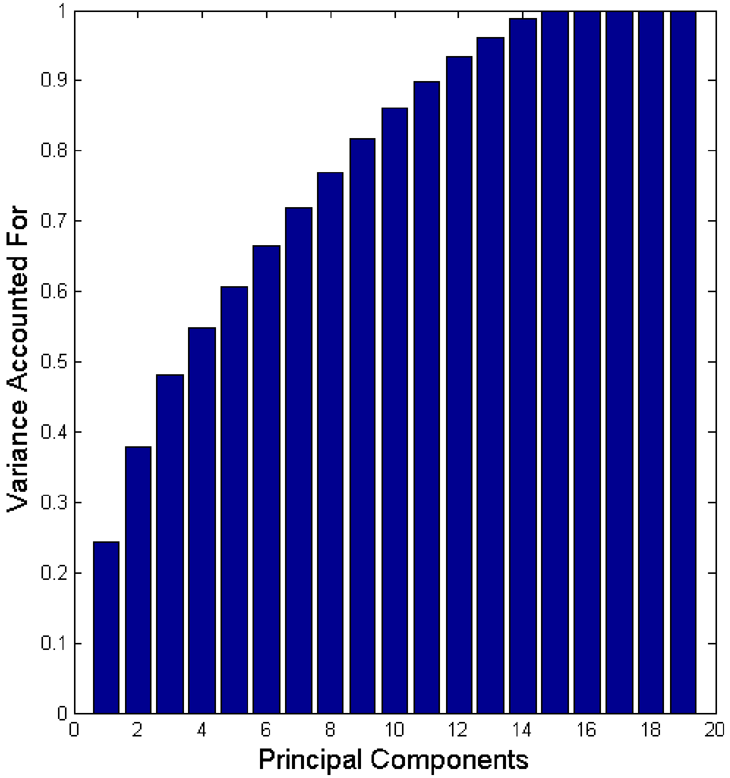
Features



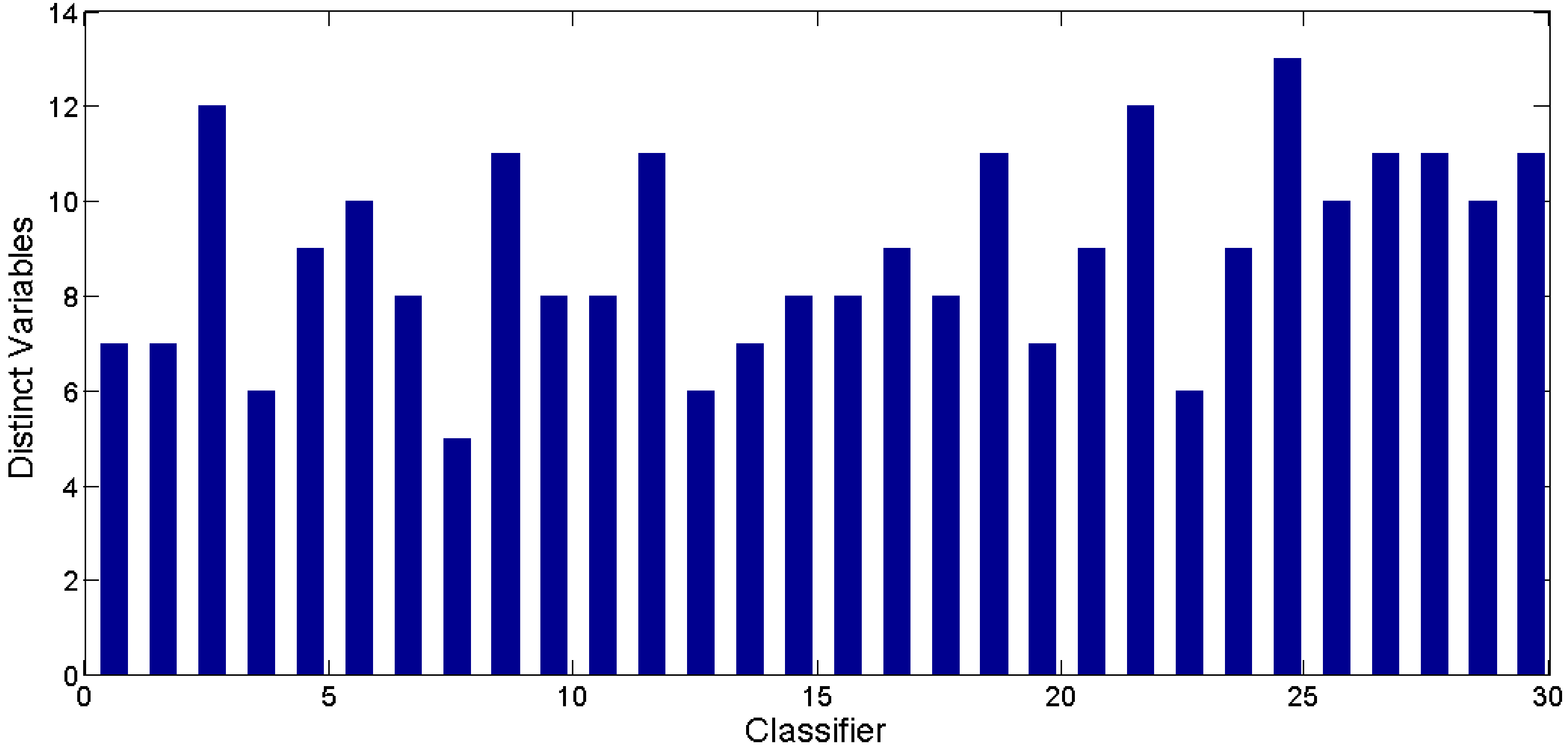
speed: separates [*flight*, *cruise*] from [*nest*, *perched*]
agl: separates [*flight*] from [*cruise*]
nest dist: separates [*nest*] from [*perched*]

Principal Component Analysis





Number of features



Future Work

- Try simpler classification schemes for Bald Eagle data
- New data
  - More specific behaviors
  - Classify new data automatically
- Interpretation of results
- Test scalability of GP method

Thank you!