Neural Data Science-- Clustering Neurons in High Dimensions

Maria Kesa

AstraZeneca Interview, June 9, 2020

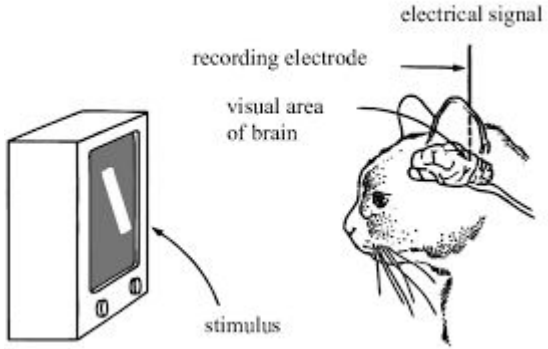
1. Agenda

New recording technologies may transform neuroscience.

- ➔ Simultaneous calcium imaging recordings from 19,000 neurons in mouse V1 and simultaneous behavioral videos
- ➔ Cool new algorithm for analyzing this data (it's fast too!)
- ➔ From algorithm to insights

History-- Neuroscience Classics

Hubel and Wiesel,
1960's and 1970's



Hebb,
1940's and 1950's

$$\frac{dw_{i,j}}{dt} = pre_i \cdot post_j$$

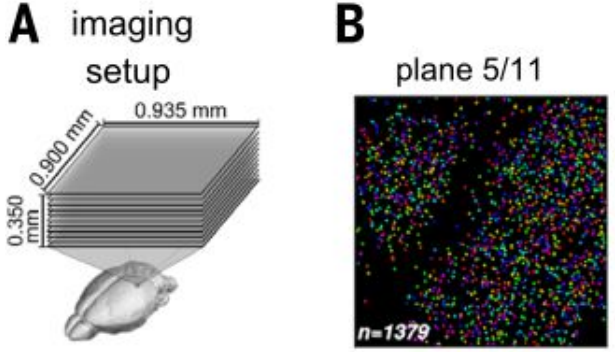
The experiment

The mice were viewing sparse noise stimuli without an explicit task or reward. Recordings were made in primary visual cortex.

Along with neural recordings, the movie of the mouse snout was recorded.

Data freely available online (google Stringer Figshare).

Preprocessing of raw data



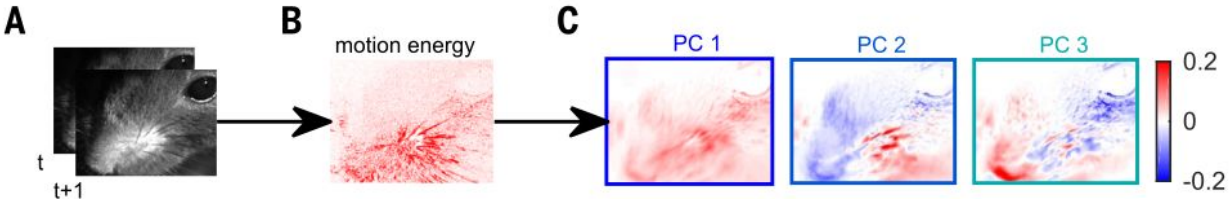
The data was recorded at 2.5 Hz (one time point is 400 ms).

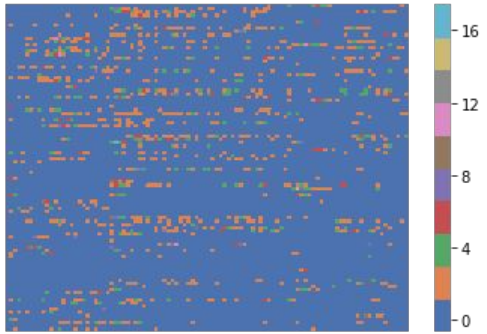
Fluorescence movie was processed using the suite2p library.

We used **deconvoloved data** (OASIS) for our analyses.

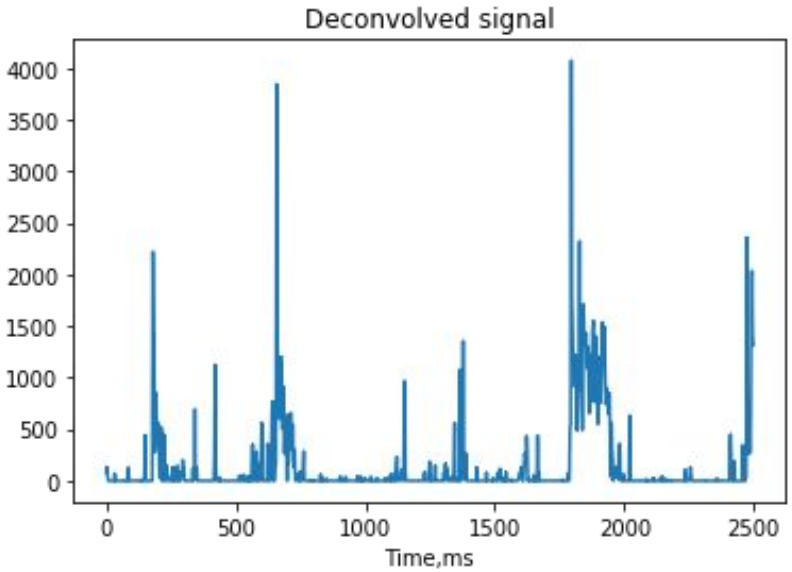
Deconvolution infers time-localized and noisy estimates of the calcium influx of the cells which is an indirect proxy for action potential firing (Stringer, Pachitariu, 2018).

We used 500 motion energy principal components from the snout movie as behavioral data.





What does the data look like?



Ensemble Pursuit-- A clustering algorithm where a neuron can belong to multiple clusters

EP learns clusters of cells that tend to co-activate, e.g. form ensembles.
https://github.com/MouseLand/EnsemblePursuit

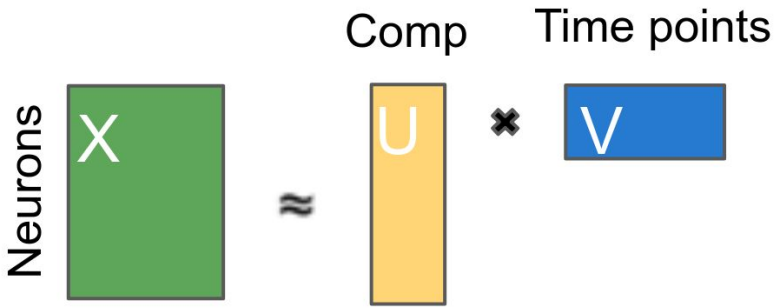
```
!pip install EnsemblePursuit
ep=EnsemblePursuit(n_components=100,lam=0.01,n_kmeans=100)
model=ep.fit(data)
V=model.components_
U=model.weights
```

Extracts 100 components of co-activating cells from 19,000 neurons x10,000 time points data in 4 minutes

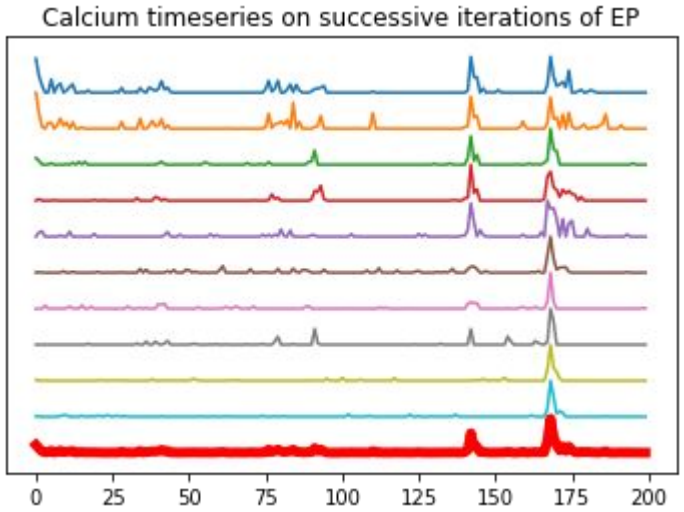
Ensemble Pursuit Algorithm

$$\text{Cost} = \|X - U \cdot V\|^2 + \lambda \|U\|_0$$

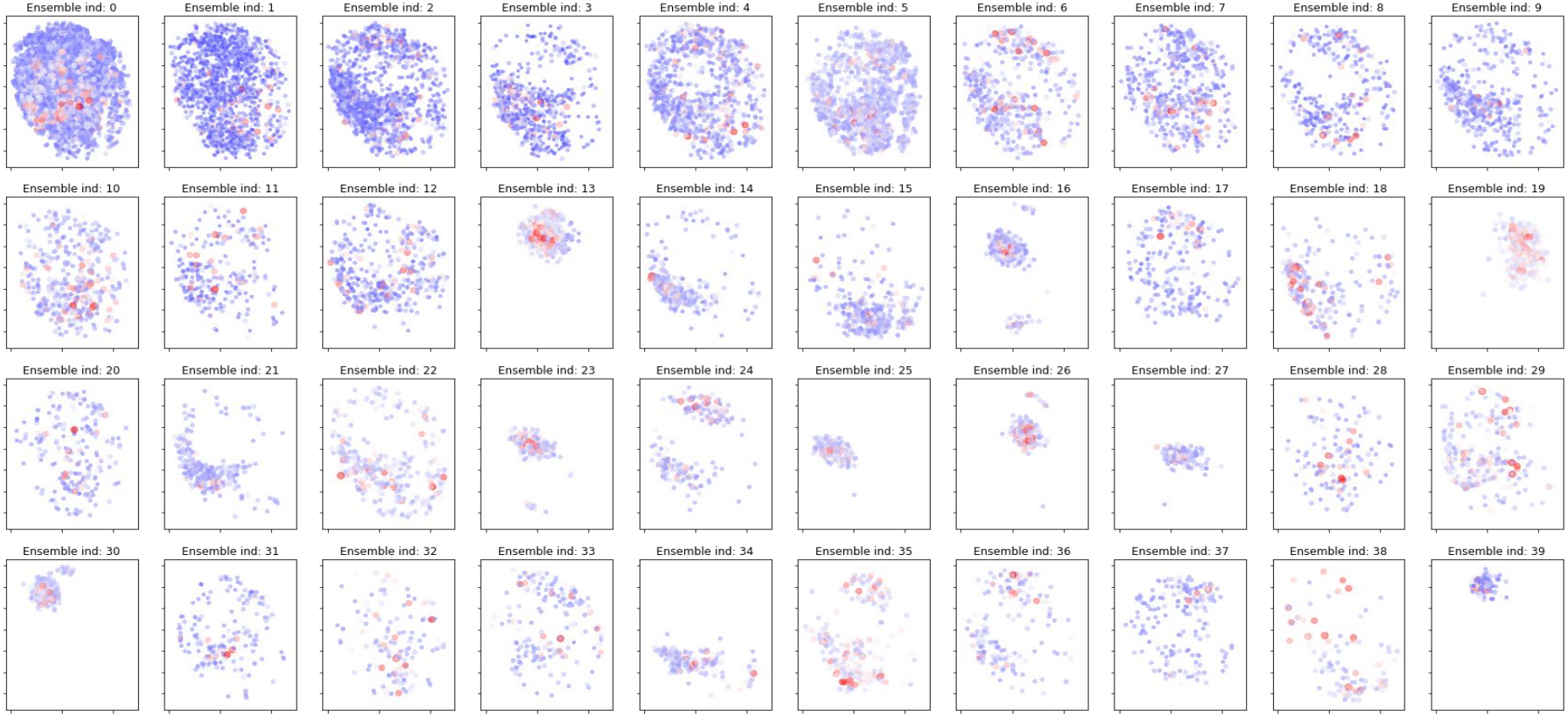
Pick lambda through supervised (KNN) cross-validation!



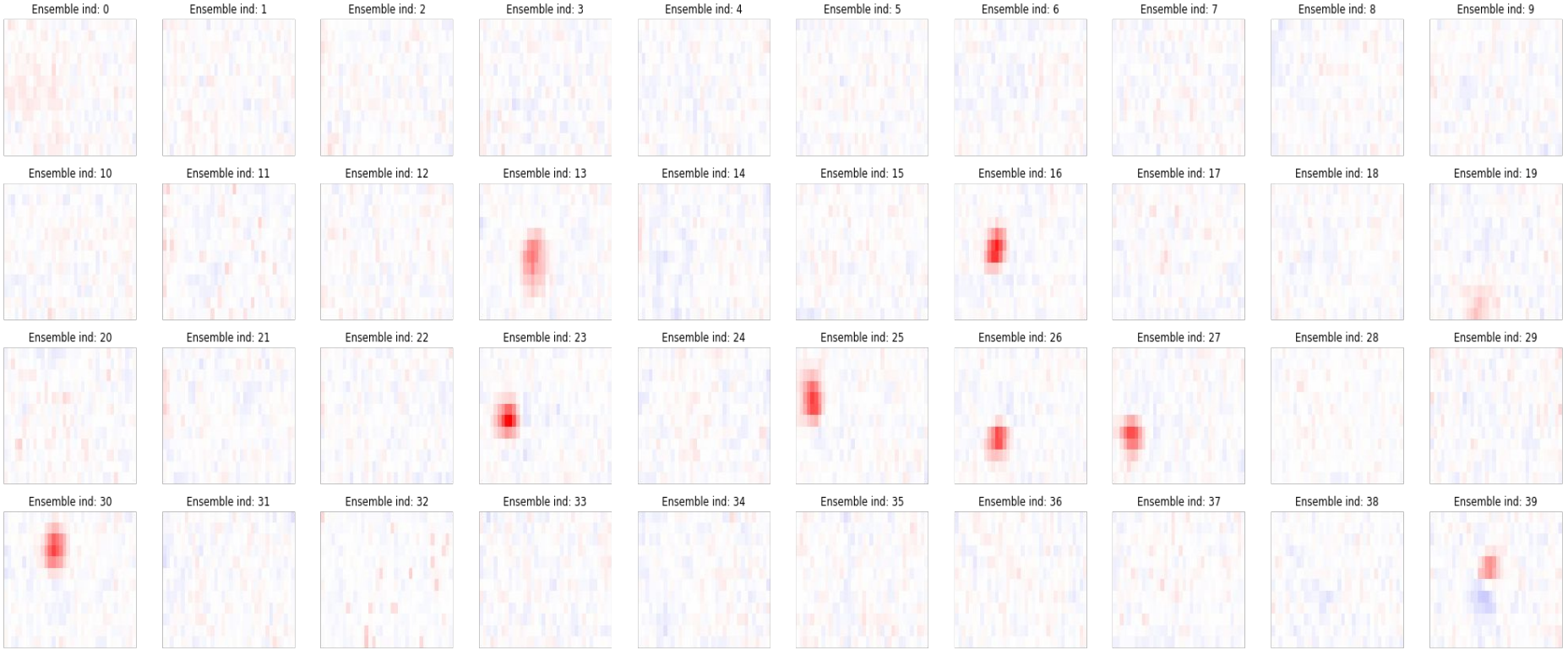
NP-Hard!
Greedy approximation!



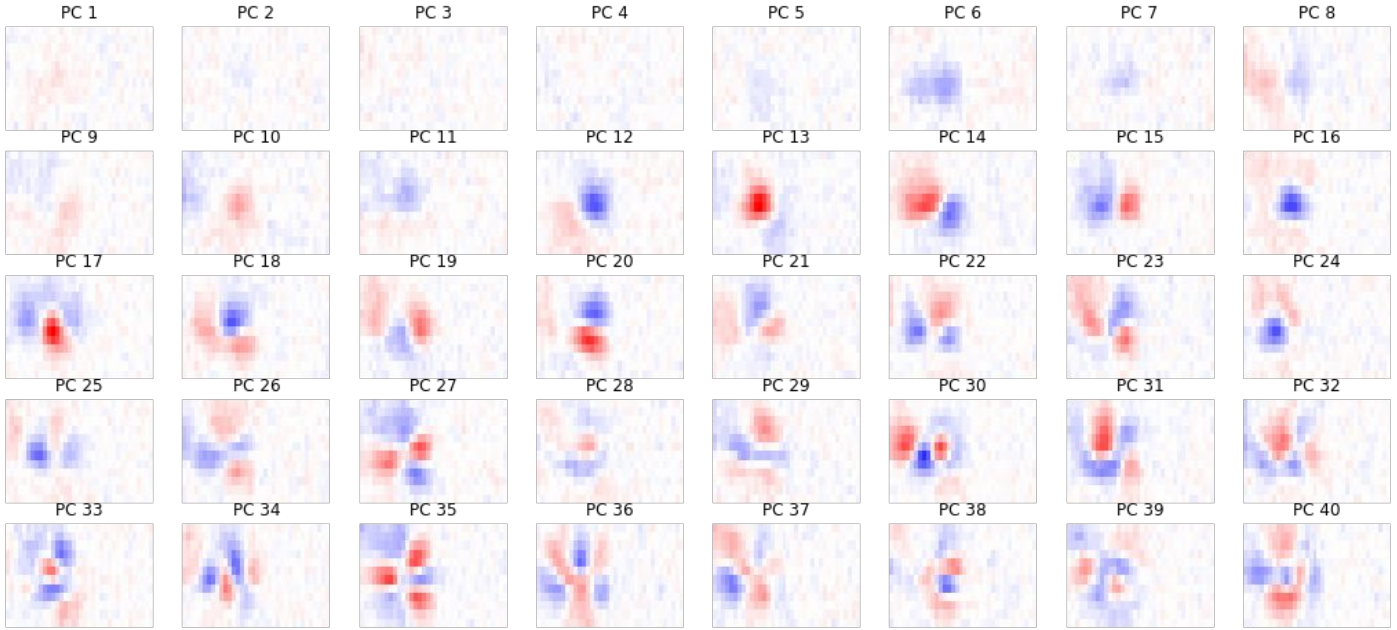
Cells in clusters



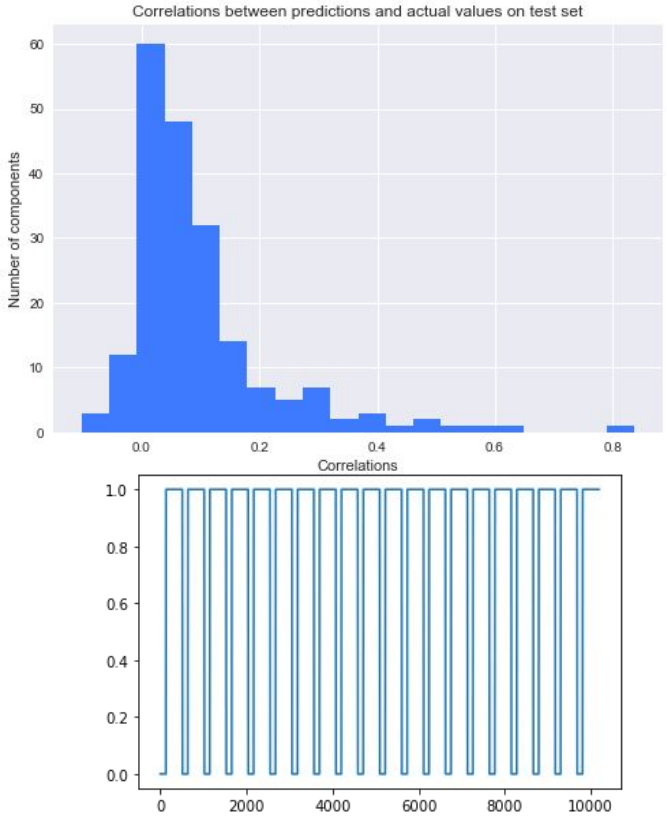
Ensemble receptive fields

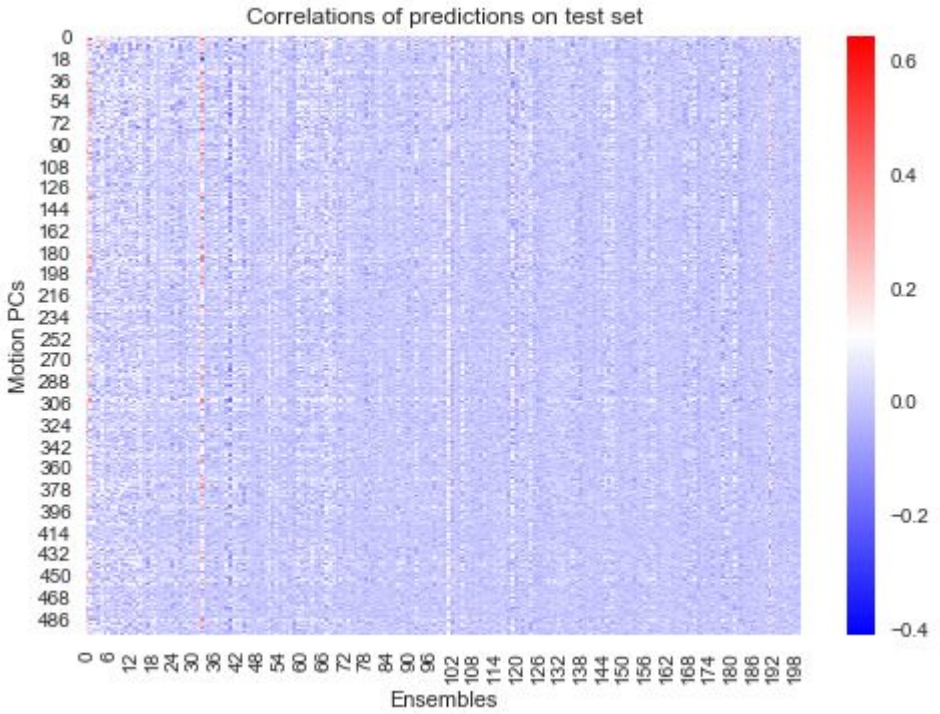


PCA receptive fields

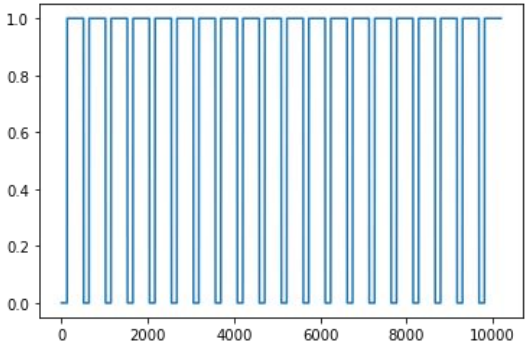


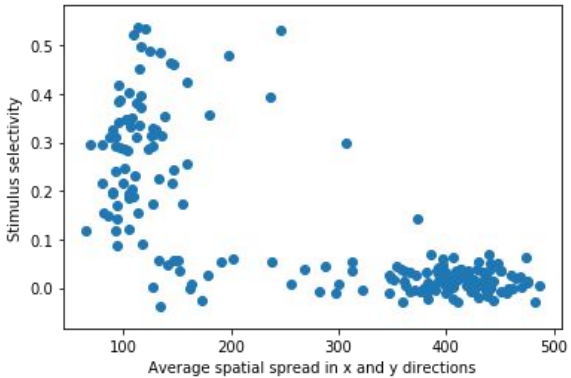
Signatures of sensorimotor integration

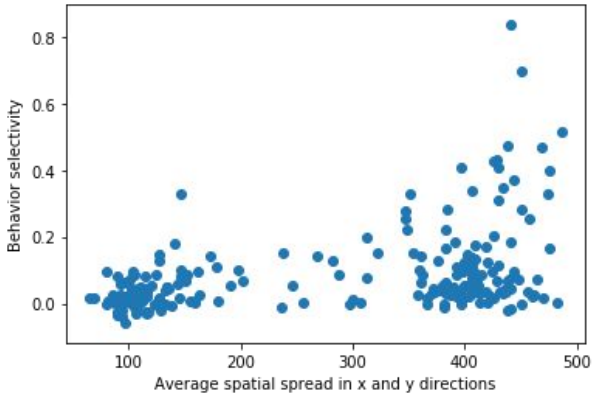




‘Behavior’ ensembles are spatially spread out, ‘Stimulus’ ensembles are localized





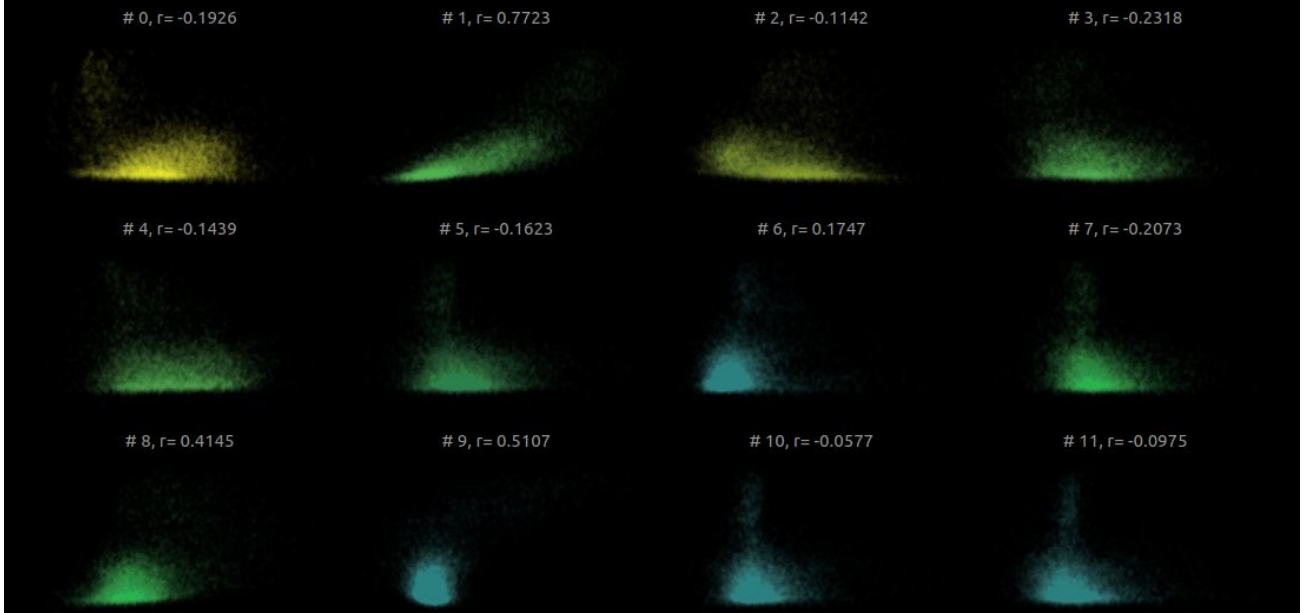


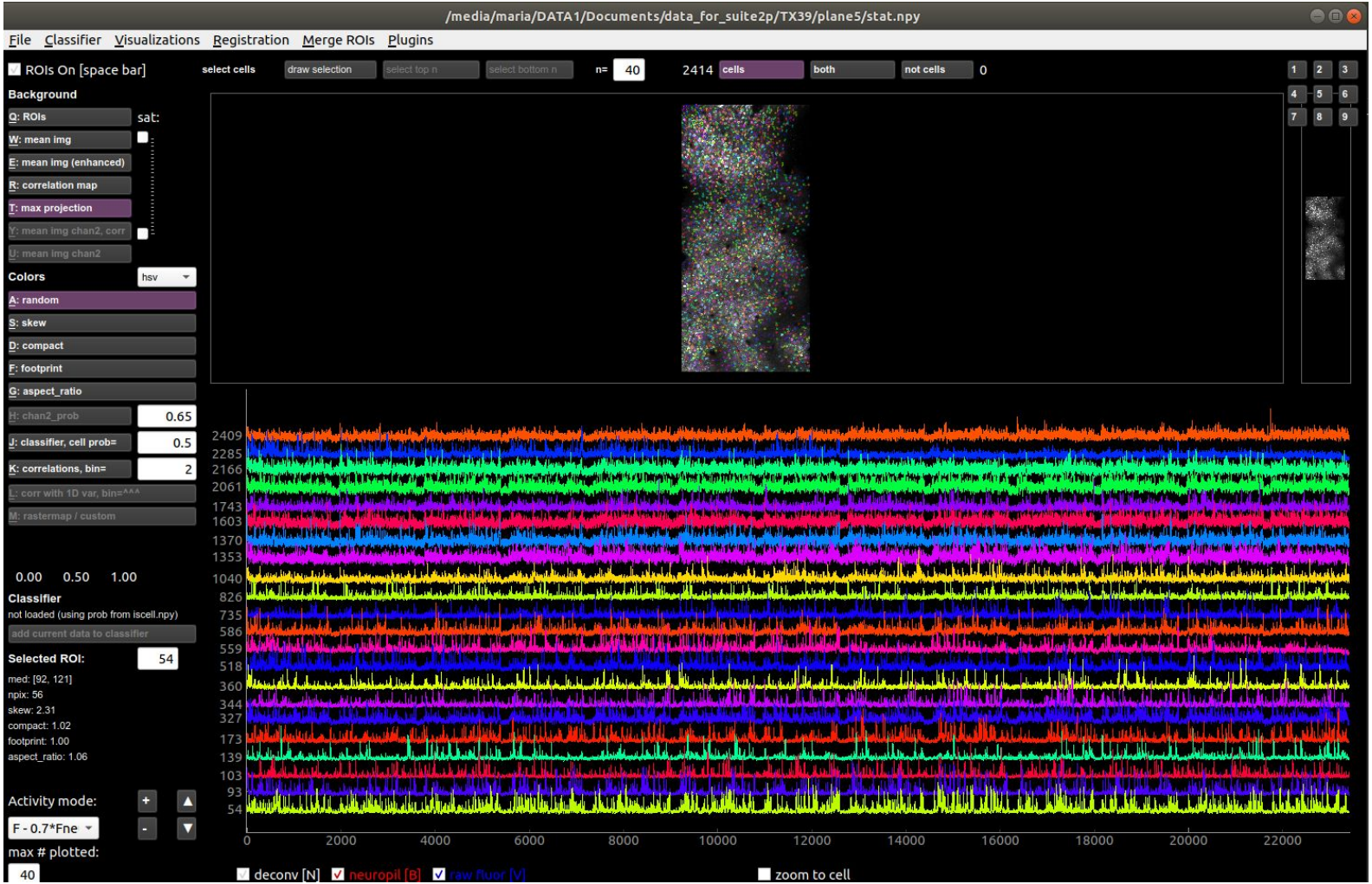
Visualize data

compute EnsemblePursuit | Nr of ensembles= 100 | Lambda: 0.01 | Range:0-69 | Select cells | Ensemble:0 | Neuron: 0

2168	416	630	416	420	178	137	204	210	53
87	64	136	143	56	131	65	100	100	60
68	91	86	61	67	55	70	38	92	62
47	96	44	91	42	47	52	36	28	14
31	48	36	5	26	34	20	31	46	54
40	37	18	19	23	26	33	24	25	24
25	35	24	22	28	18	16	15	17	24

Principal Components | Index= 0 | Range:0-11




/media/maria/DATA1/Documents/data_for_suite2p/TX39/plane5/stat.npy
File Classifier Visualizations Registration Merge ROIs Plugins
ROIs On [space bar]
select cells
draw selection
select top n
select bottom n
n= 40
2414 cells
both
not cells
0
Background
Q: ROIs
sat:
W: mean img
E: mean img (enhanced)
R: correlation map
T: max projection
Y: mean img chan2, corr
U: mean img chan2
Colors
hsv
A: random
S: skew
D: compact
F: footprint
G: aspect_ratio
H: chan2_prob
0.65
J: classifier, cell prob=
0.5
K: correlations, bin=
2
L: corr with 1D var, bin=^^^
M: rastermap / custom
0.00 0.50 1.00
Classifier
not loaded (using prob from iscell.npy)
add current data to classifier
Selected ROI:
54
med: [92, 121]
npix: 56
skew: 2.31
compact: 1.02
footprint: 1.00
aspect_ratio: 1.06
Activity mode:
F - 0.7*Fne
max # plotted:
40
2409
2285
2166
2061
1743
1603
1370
1353
1040
826
735
586
559
518
360
344
327
173
139
103
93
54
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
22000
deconv [N]
neuropil [B]
raw fluor [V]
zoom to cell

Acknowledgments!

Marius Pachitariu, Howard Hughes Medical Institute, Janelia

Carsen Stringer, Howard Hughes Medical Institute, Janelia

VERIFIED
CERTIFICATE of ACHIEVEMENT

HarvardX

This is to certify that

Maria Kesa

successfully completed and received a passing grade in

PH559x: Causal Diagrams: Draw Your Assumptions Before Your Conclusions

a course of study offered by HarvardX, an online learning initiative of Harvard University.

Miguel Hernán
Kolokotrones Professor of Biostatistics and Epidemiology
Harvard T.H. Chan School of Public Health

edX

VERIFIED CERTIFICATE
Issued June 9, 2020

VALID CERTIFICATE ID
58fd5a12b55d4686b6a347ba1c6cacbf

VERIFIED
CERTIFICATE of ACHIEVEMENT

Stanford | ONLINE

This is to certify that

Maria Kesa

successfully completed and received a passing grade in

STATSX0001: Statistical Learning

a course of study offered by StanfordOnline, an online learning initiative of Stanford University.

Trevor Hastie
John A. Overdeck Professor, Professor of Statistics and Biomedical Data Sciences
Stanford University

Robert Tibshirani
Professor of Biomedical Data Science and Statistics
Stanford University

edX

VERIFIED CERTIFICATE
Issued June 9, 2020

VALID CERTIFICATE ID
0592bfa1880b42b5beac1ce35b812998

PLEASE NOTE: SOME ONLINE COURSES MAY DRAW ON MATERIAL FROM COURSES TAUGHT ON-CAMPUS BUT THEY ARE NOT EQUIVALENT TO ON-CAMPUS COURSES. THIS STATEMENT DOES NOT AFFIRM THAT THIS PARTICIPANT WAS ENROLLED AS A STUDENT AT STANFORD UNIVERSITY IN ANY WAY. IT DOES NOT CONFER A STANFORD UNIVERSITY GRADE, COURSE CREDIT OR DEGREE.