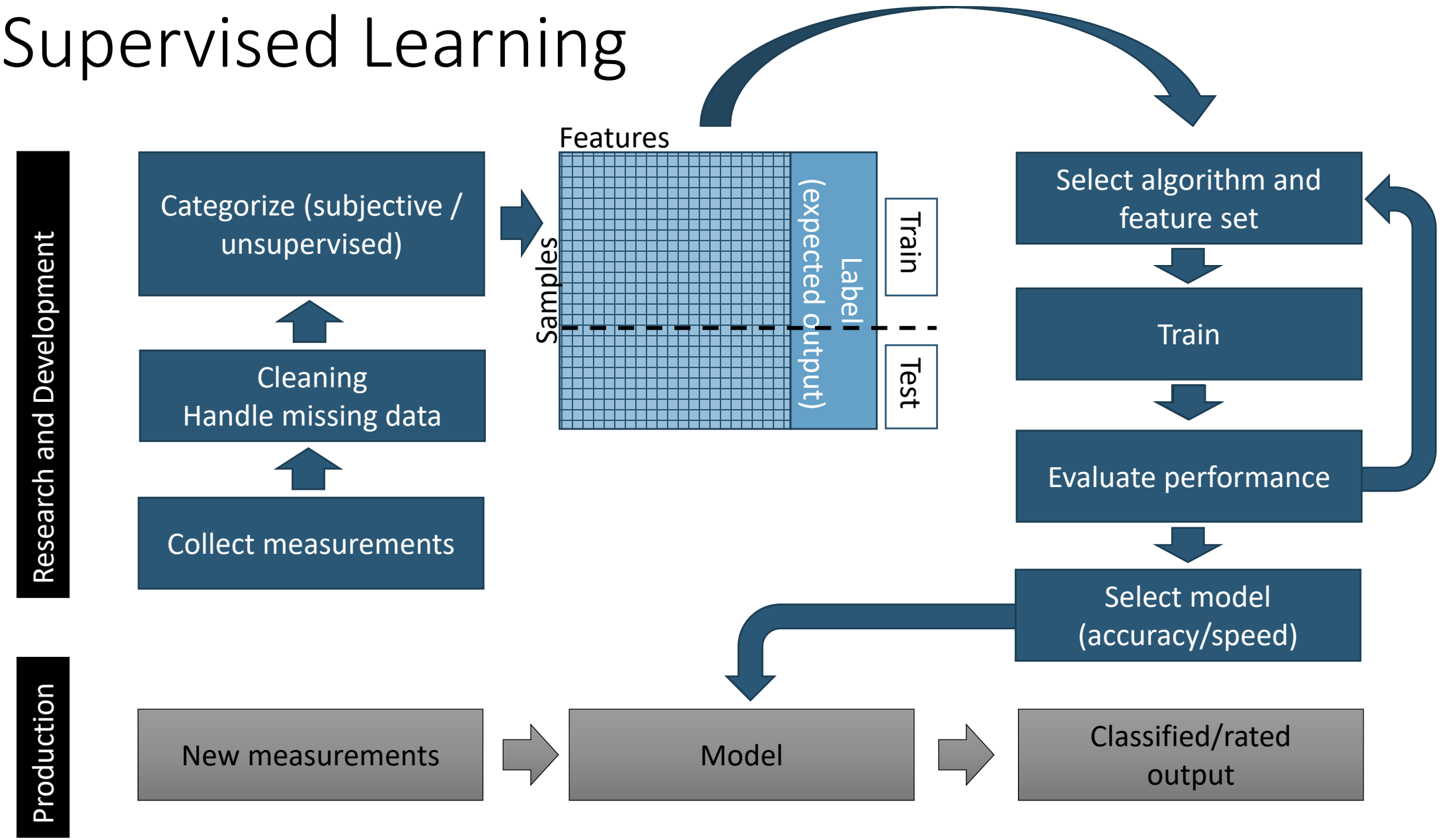
Annotation and Patterns in Sensor Data

Anders Hedlund
Analytics Specialist
anders@binordic.com

BI Nordic
Business Intelligence Nordic AB

Supervised Learning



Wine Quality Data Set*

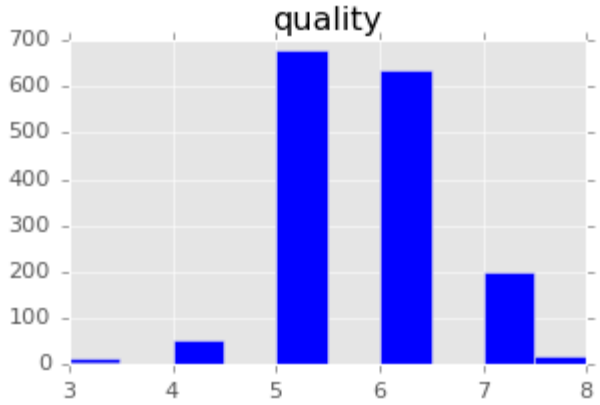
fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	type	quality
7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	red	5
7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	red	7
6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	white	7
8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	white	5
8.5	0.28	0.56	1.8	0.092	35	103	0.9969	3.3	0.75	10.5	red	7
7.4	0.59	0.08	4.4	0.086	6	29	0.9974	3.38	0.5	9	red	4
6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	white	8

Vinho verde (Portugal)
White wine: 4898 samples
Red wine: 1599 samples

11 measured parameters

Subjective quality: median of at least 3 evaluations made by wine experts.
Range 0 (very bad) and 10 (very excellent)

Histogram Red Wine

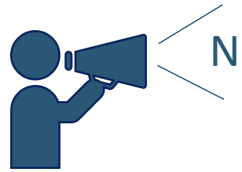


*P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.
Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

Correct Annotation is Key

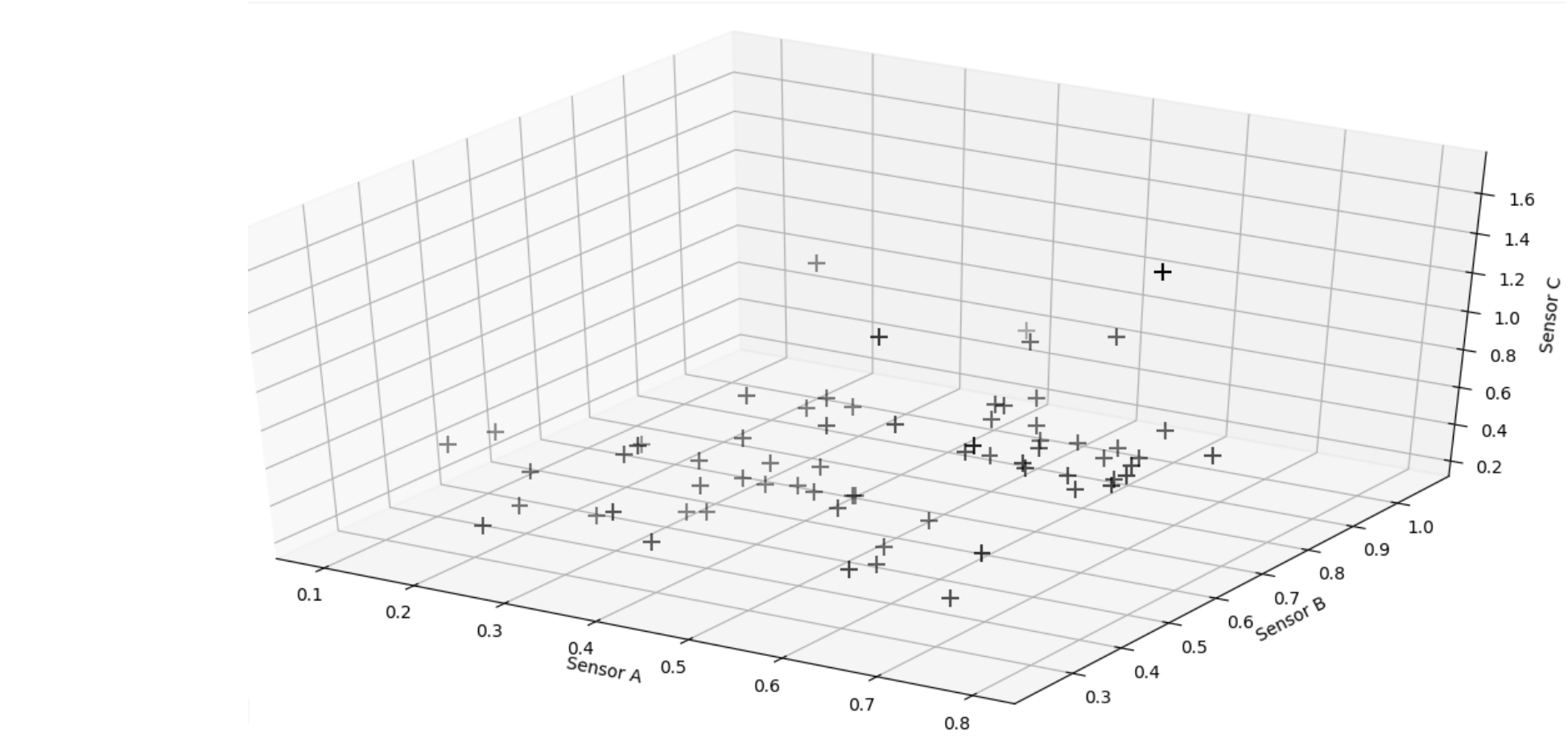
- ✓ MANY LINES WITH SIMPLE INFORMATION MUCH BETTER THAN FEW LINES WITH COMPLEX INFORMATION
- ✓ TIME FLIES

SIMPLE AND SQUARED

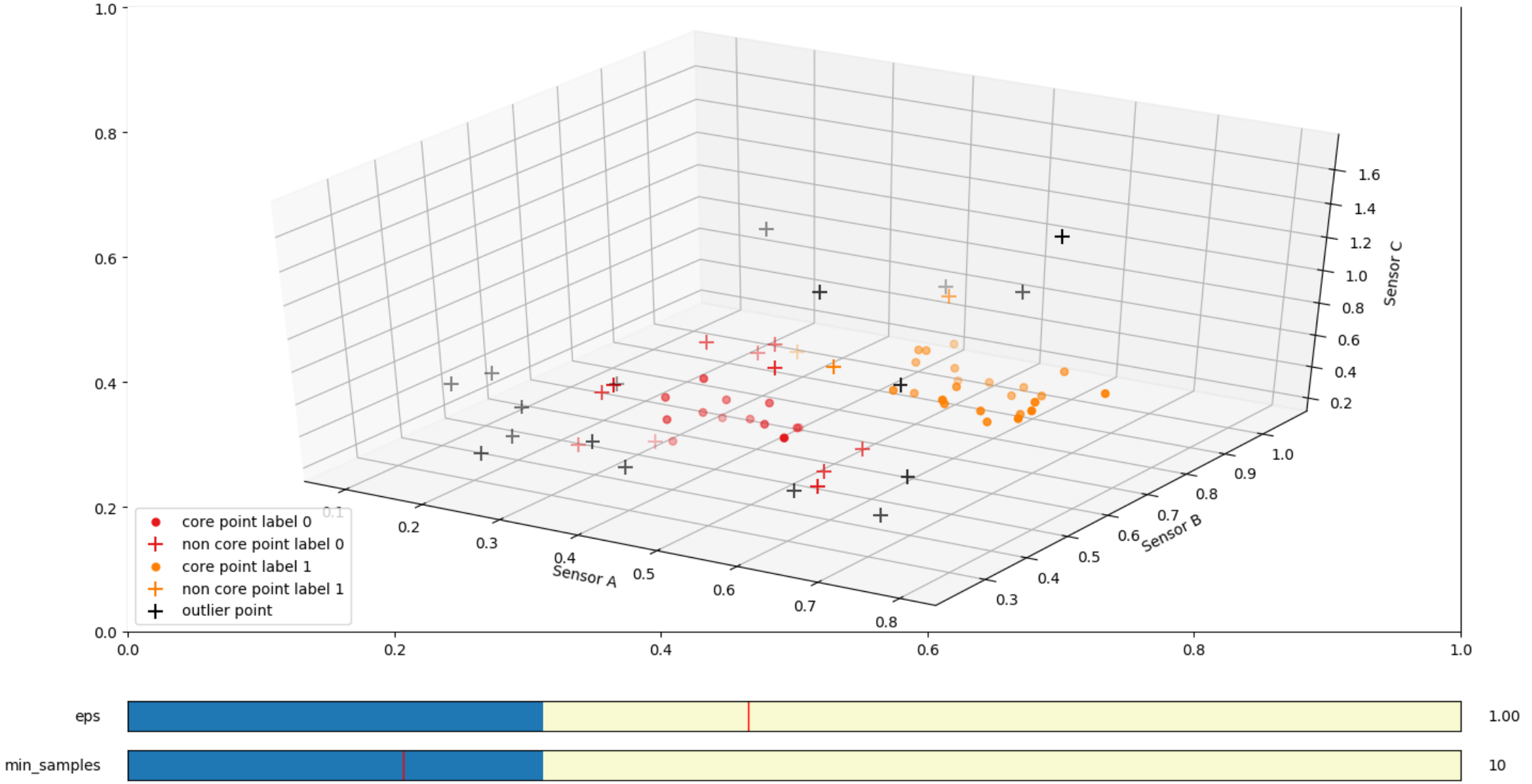


No Long Interviews Please

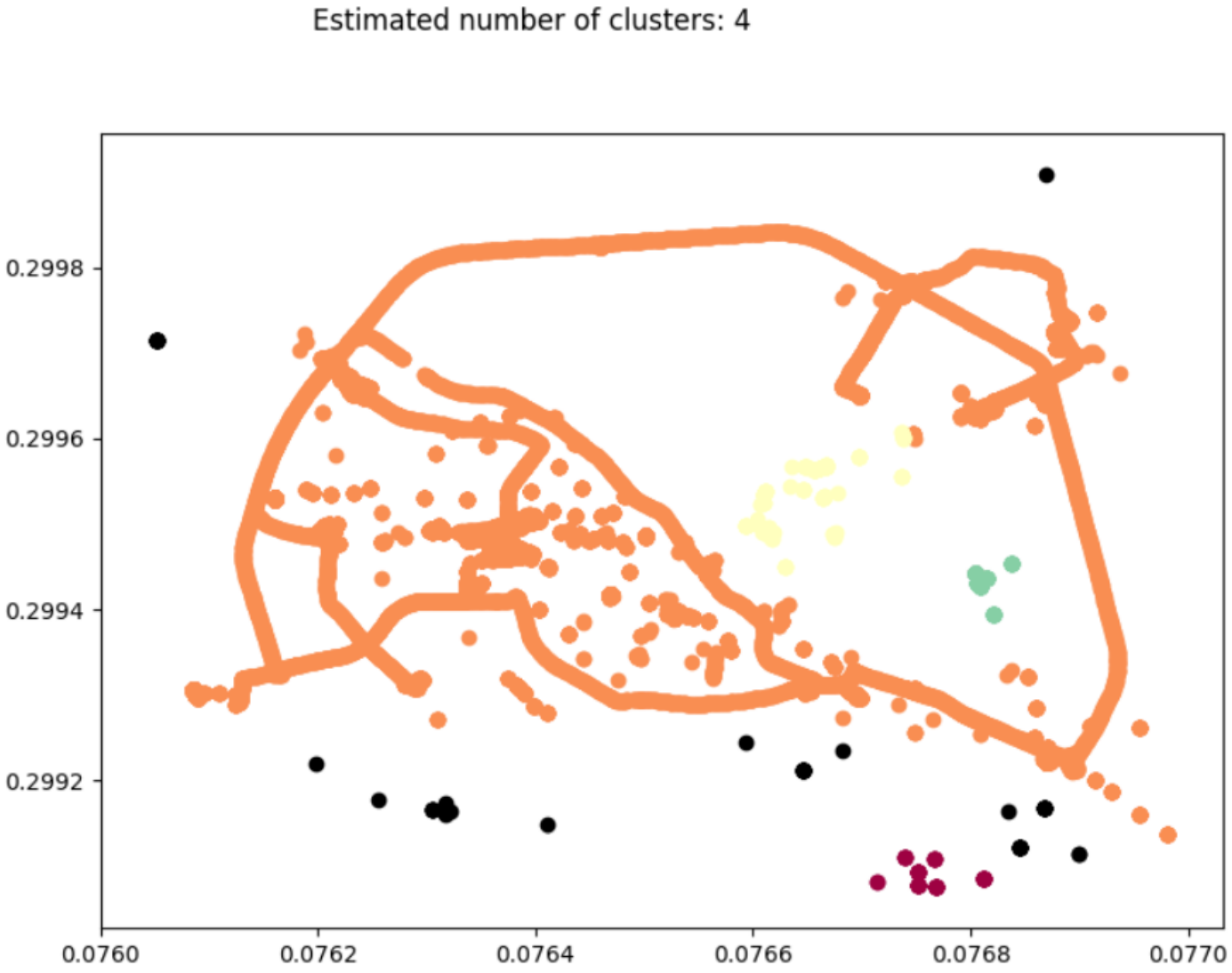
Unsupervised Learning



Clustering – A first step to Annotation and Finding Anomalies

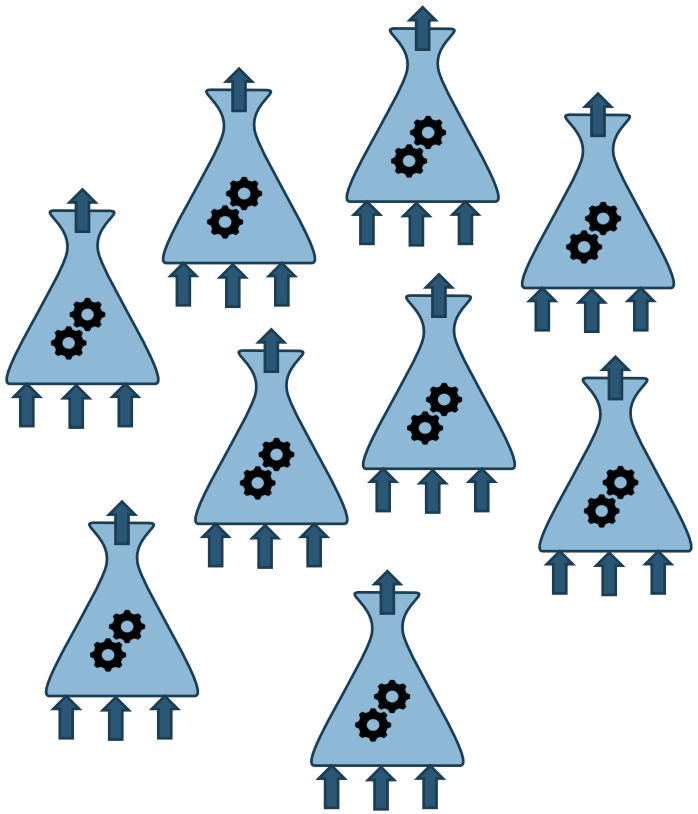


Complex Clusters – Enclosing



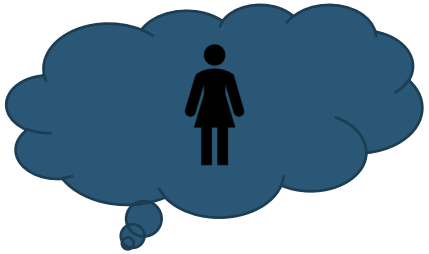
Reductionism

Artificial Specific Intelligence



Lack of glue
Very CPU intensive
The world changes behind our back

Reality



IMAGINATION
COMMUNICATION
CULTURE
ART
COGNITION (GENERATE NEW KNOWLEDGE FROM EXISTING)

Consciousness

Artificial General Intelligence