

Machine Learning Primer

Nick Record

Tandy Center for Ocean Forecasting

Colby Jan Plan: JP297Dj



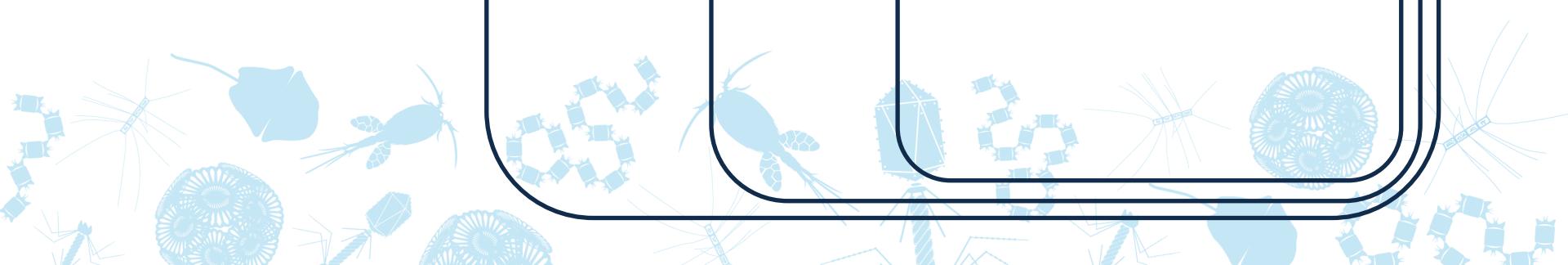
How Machine Learning Works

Machine Learning Algorithms

Artificial intelligence: algorithms learn sort of like humans:
“the effort to automate tasks normally performed by humans” -
Francois Chollet

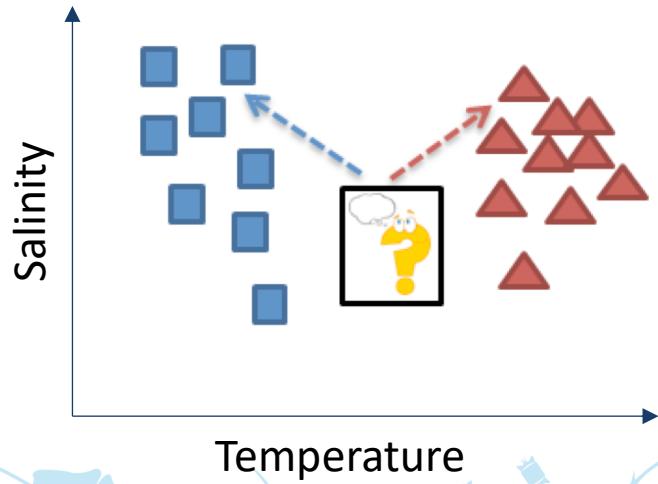
Machine learning: algorithms learn without
being programmed

Deep learning: Advanced
neural networks

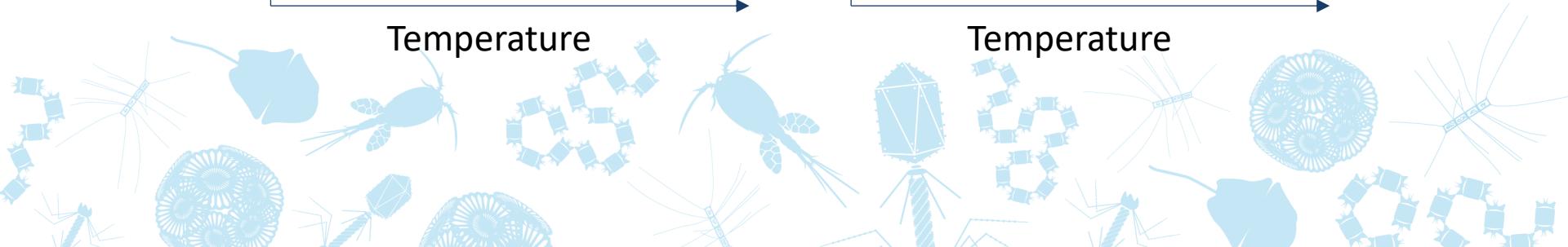
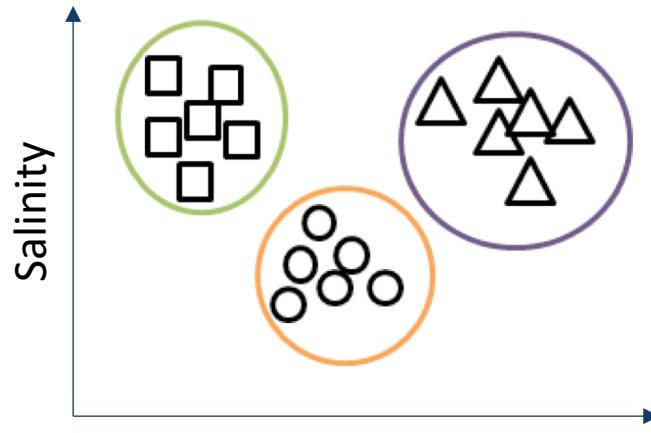


Machine Learning Algorithms

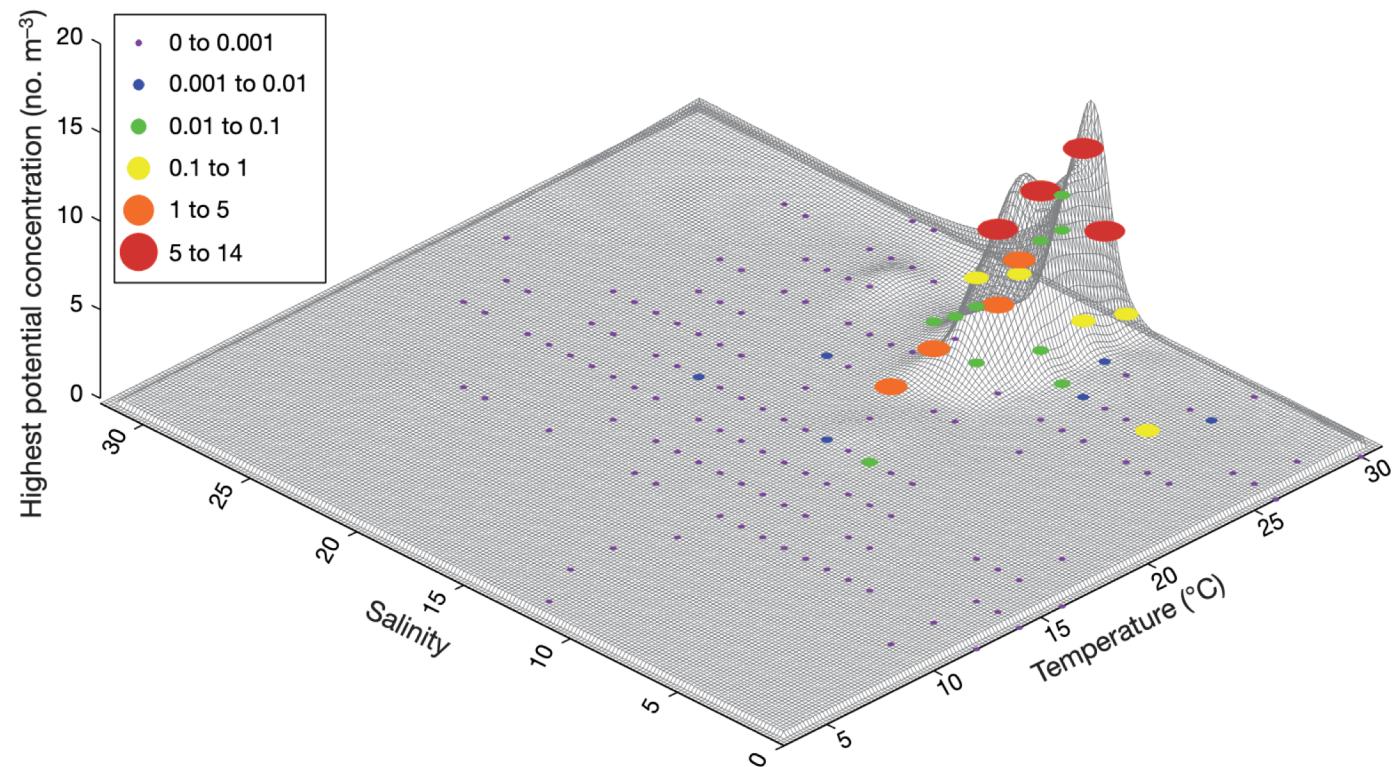
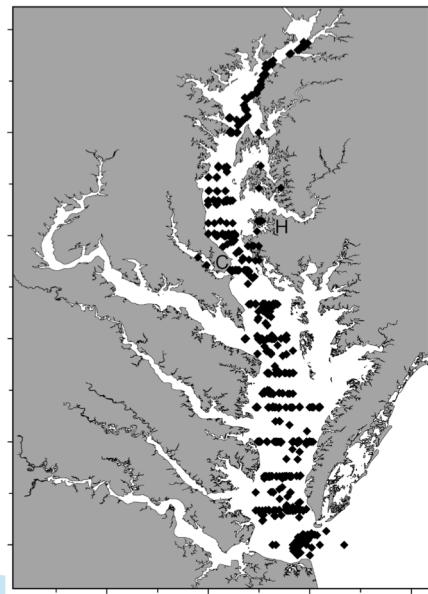
Supervised



Unsupervised

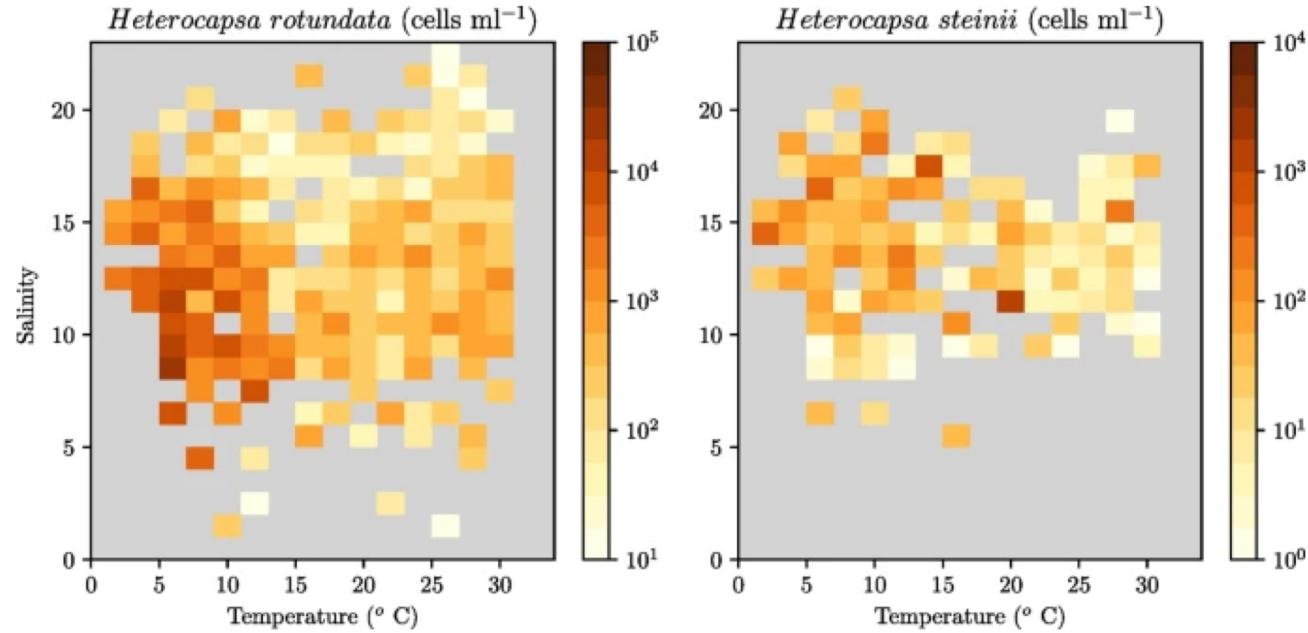


Supervised



Decker, M. B., et al. "Predicting the distribution of the scyphomedusa *Chrysaora quinquecirrha* in Chesapeake Bay." *Marine Ecology Progress Series* 329 (2007): 99-113.

Supervised



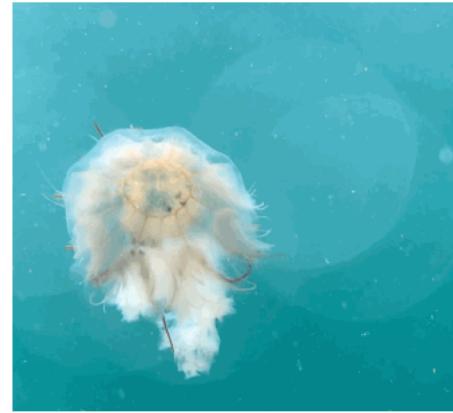
Supervised

Temperature	Salinity	Abundance
5.1	32.1	100
3.8	30.8	0
7.6	31.3	100
5.2	29.8	10
4.9	32.0	0
4.8	31.9	0
6.6	28.4	100



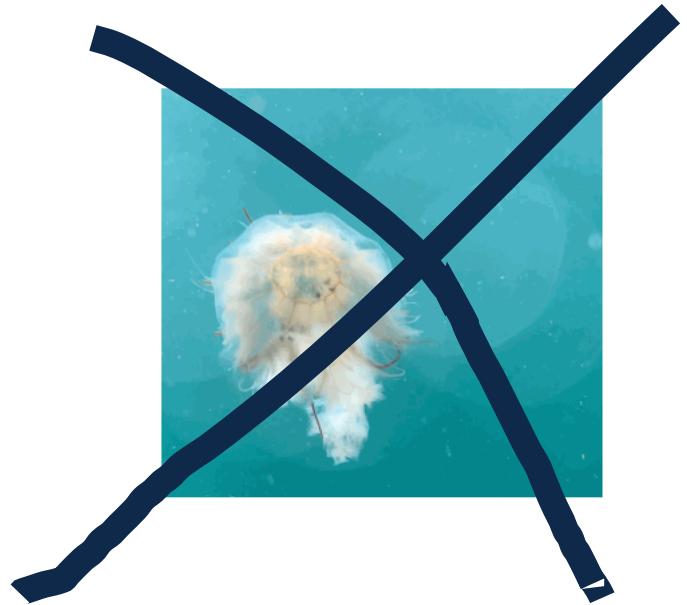
Supervised

Temperature	Salinity	Abundance
5.1	32.1	100
3.8	30.8	0
7.6	31.3	100
5.2	29.8	10
4.9	32.0	0
4.8	31.9	0
6.6	28.4	100



Supervised

Temperature	Salinity	Abundance
5.1	32.1	100
3.8	30.8	0
7.6	31.3	100
5.2	29.8	10
4.9	32.0	0
4.8	31.9	0
6.6	28.4	100



Supervised

Temperature	Salinity	Depth	Light	Turbidity	Oxygen	Abundance
5.1	32.1	100
3.8	30.8	0
7.6	31.3	100
5.2	29.8	10
4.9	32.0	0
4.8	31.9	0
6.6	28.4	100

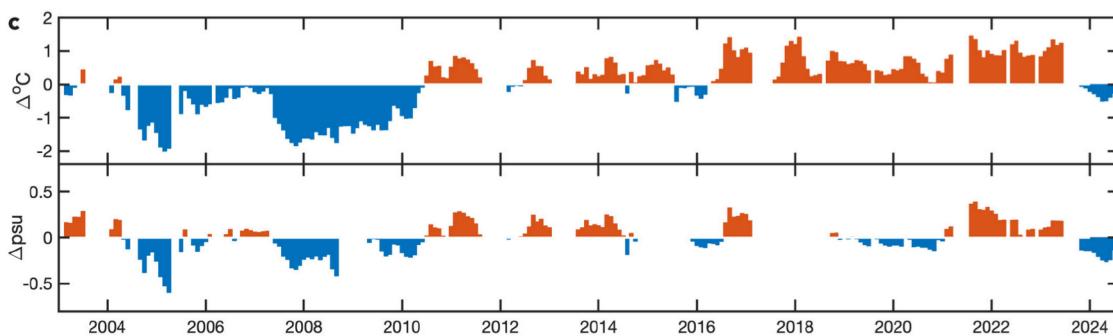


Components

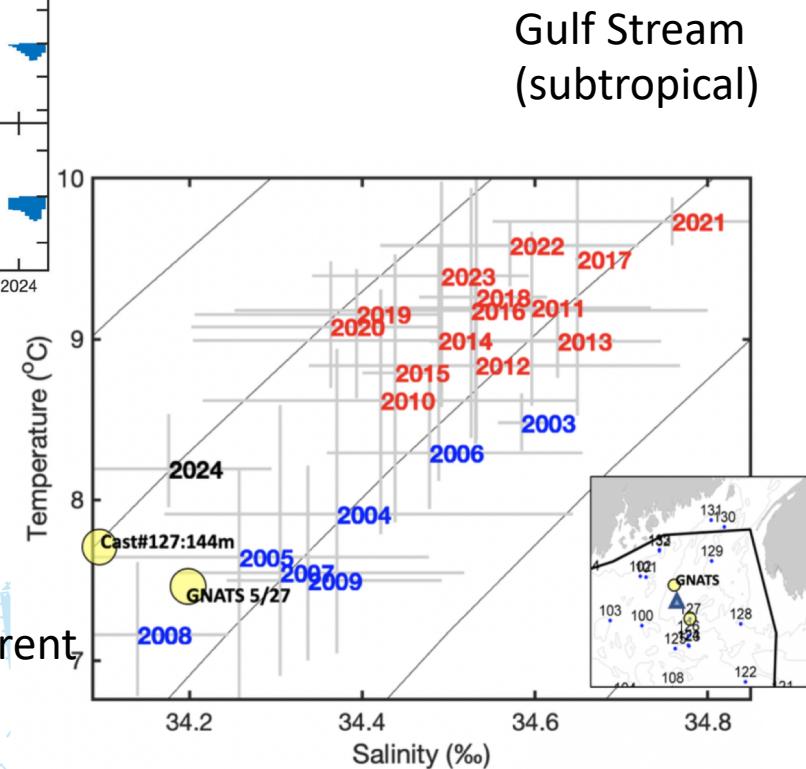
- Feature or covariate
 - E.g. temperature, salinity, etc.
 - Can be continuous or categorical
- Skill metric
 - Accuracy
 - True skill statistic
 - Correlation coefficient
 - Error/residual
 - Area under receiver-operator curve
- Prediction
- Variable importance

	Prediction	
Yes	True positive	False positive
No	False negative	True negative
	Yes	No
		Truth

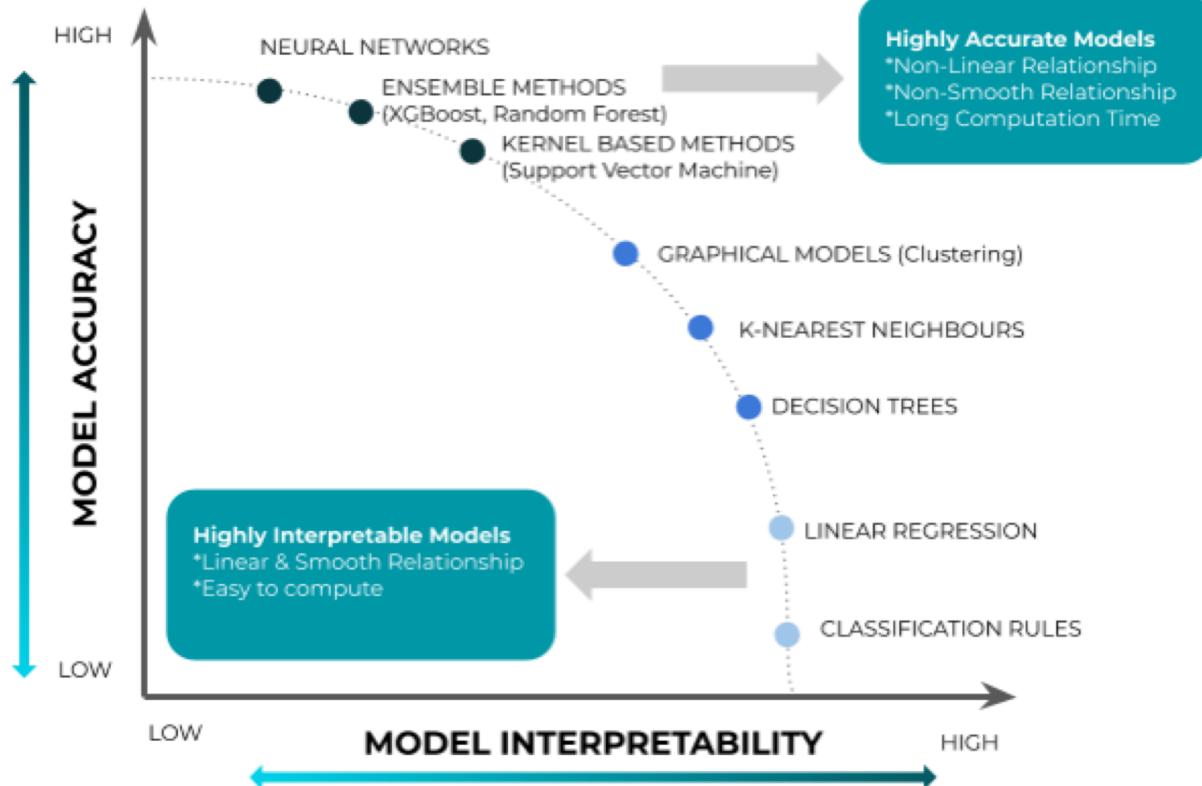
Unsupervised



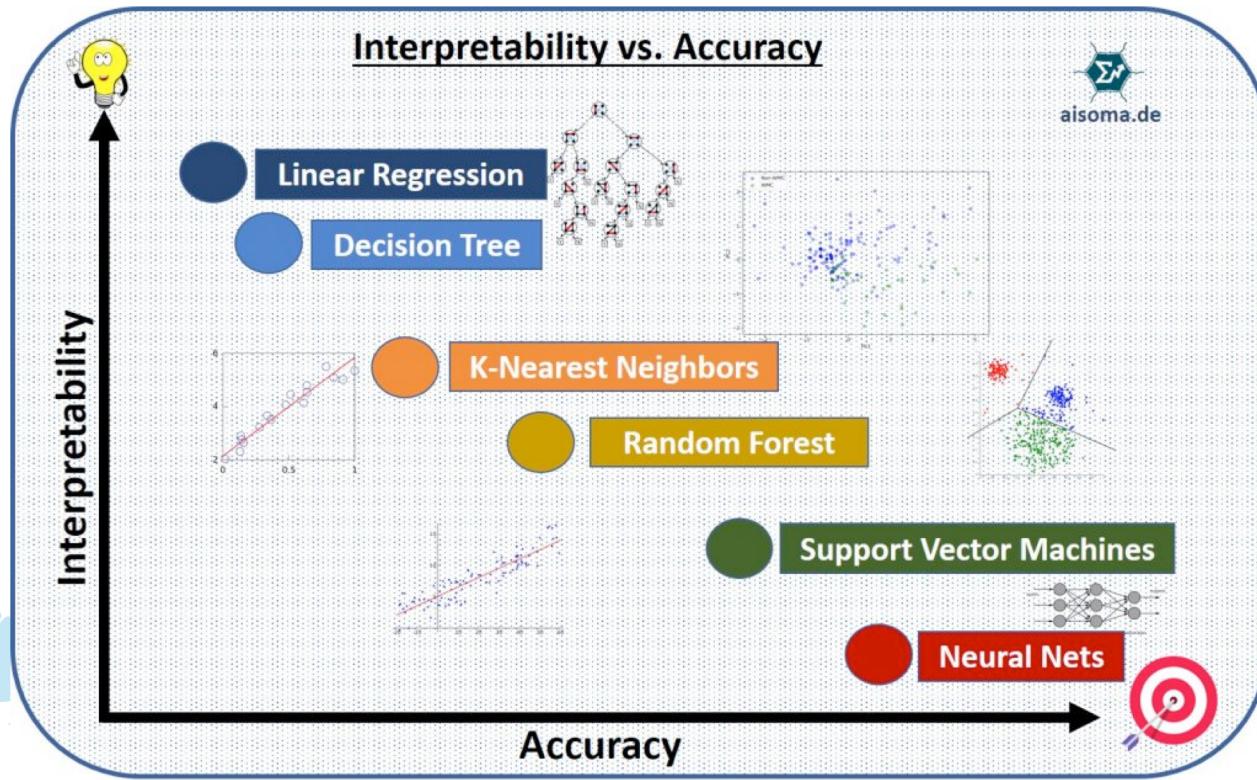
Labrador Current, (subarctic)



Machine Learning Algorithms

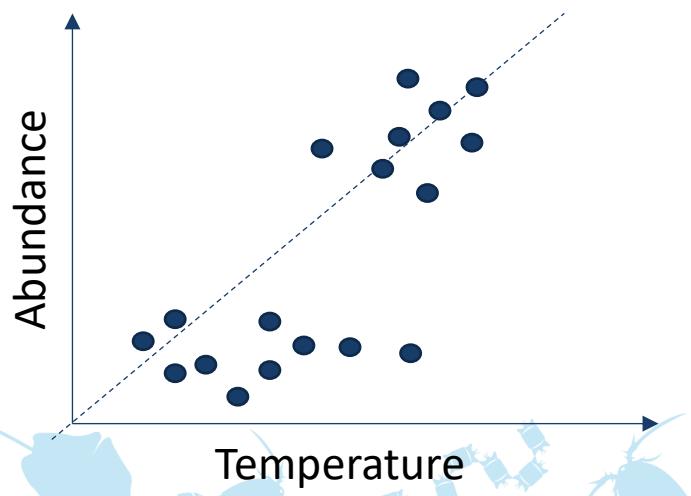


Machine Learning Algorithms

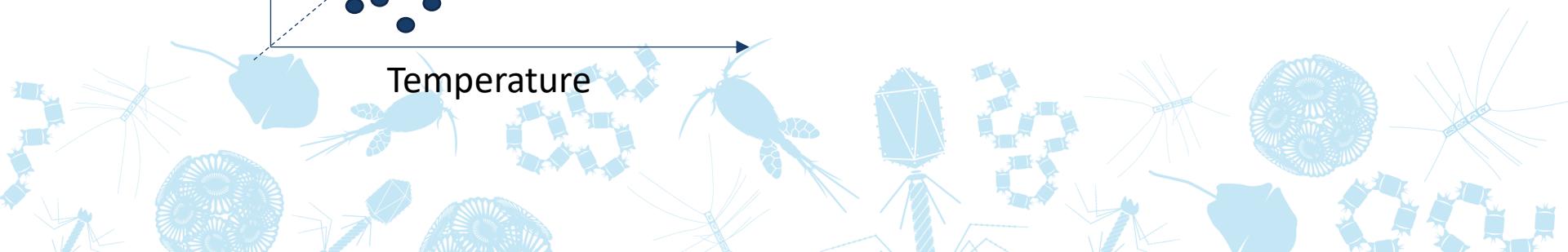


Machine Learning Algorithms

- Regression: Linear



Regression is for continuous output



Machine Learning Algorithms

- Regression: Nonlinear

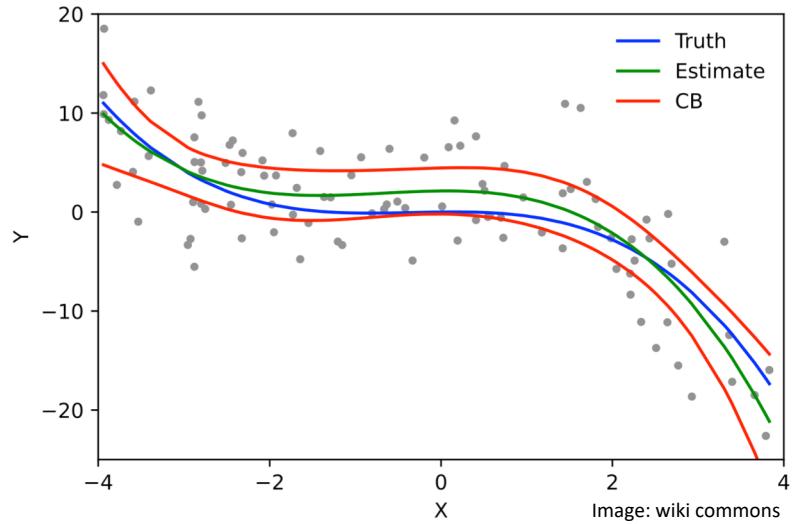
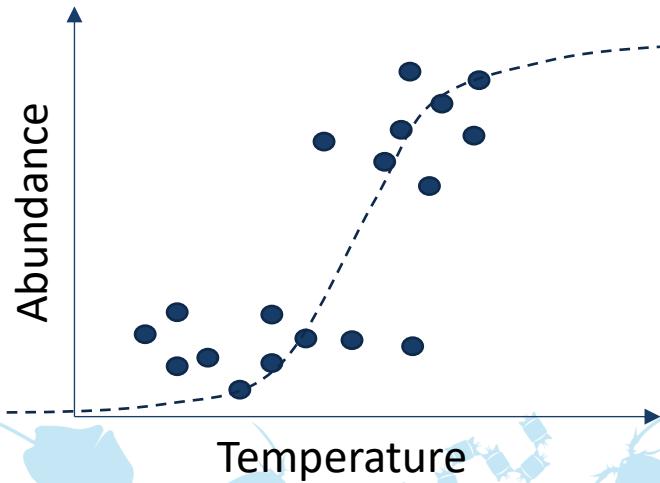


Image: wiki commons

Machine Learning Algorithms

- Regression: Logistic



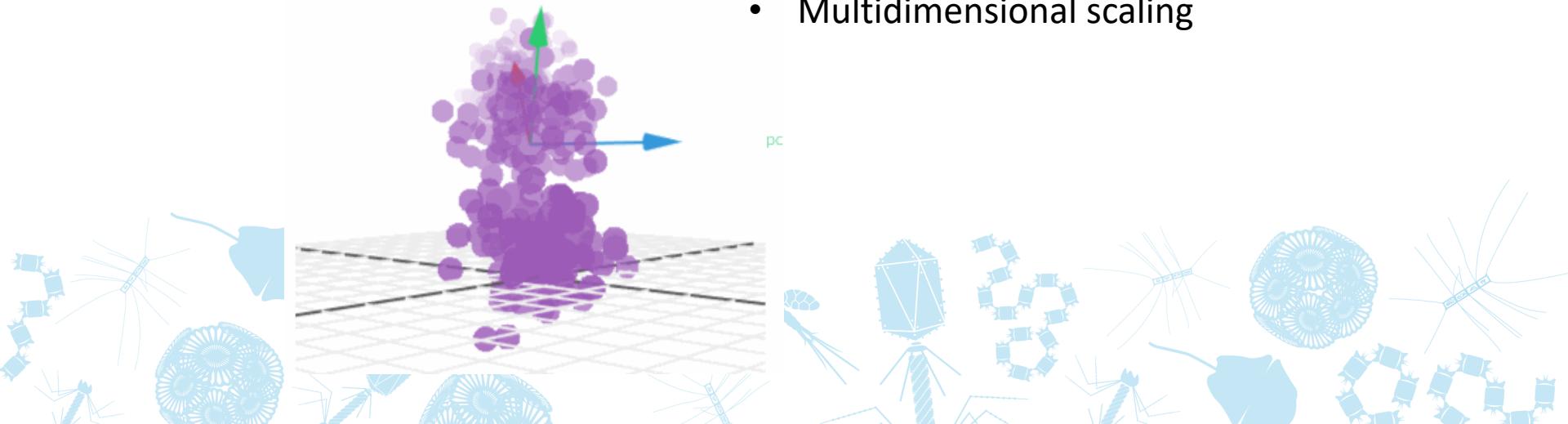
Getting toward classification



Machine Learning Algorithms

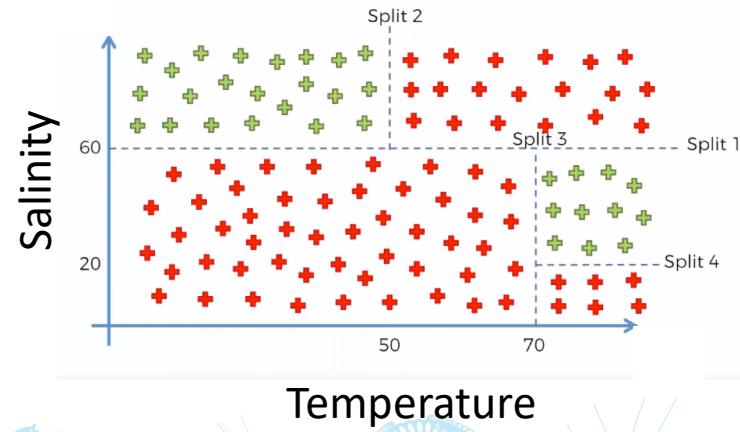
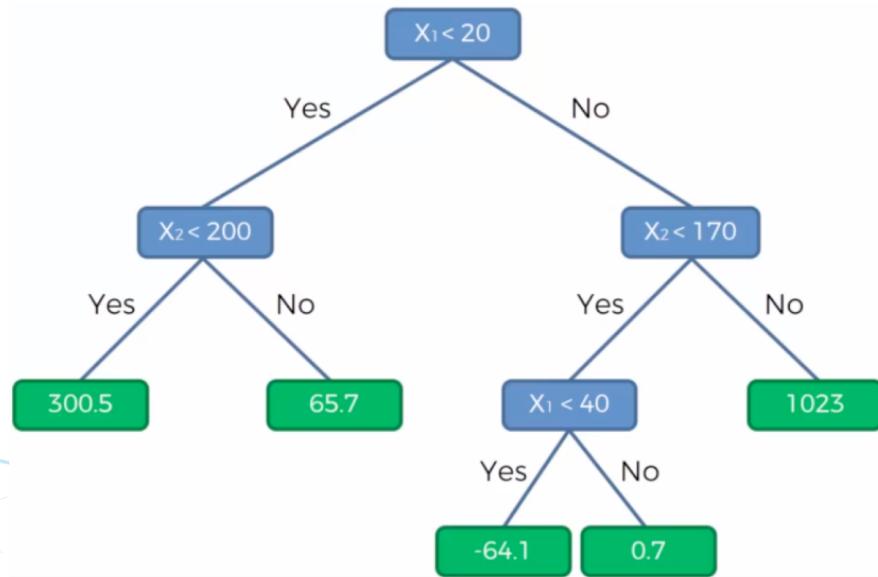
- Regression: Multiple dimensions

- Principle components analysis
- Empirical orthogonal function
- Multidimensional scaling



Machine Learning Algorithms

- Decision Tree

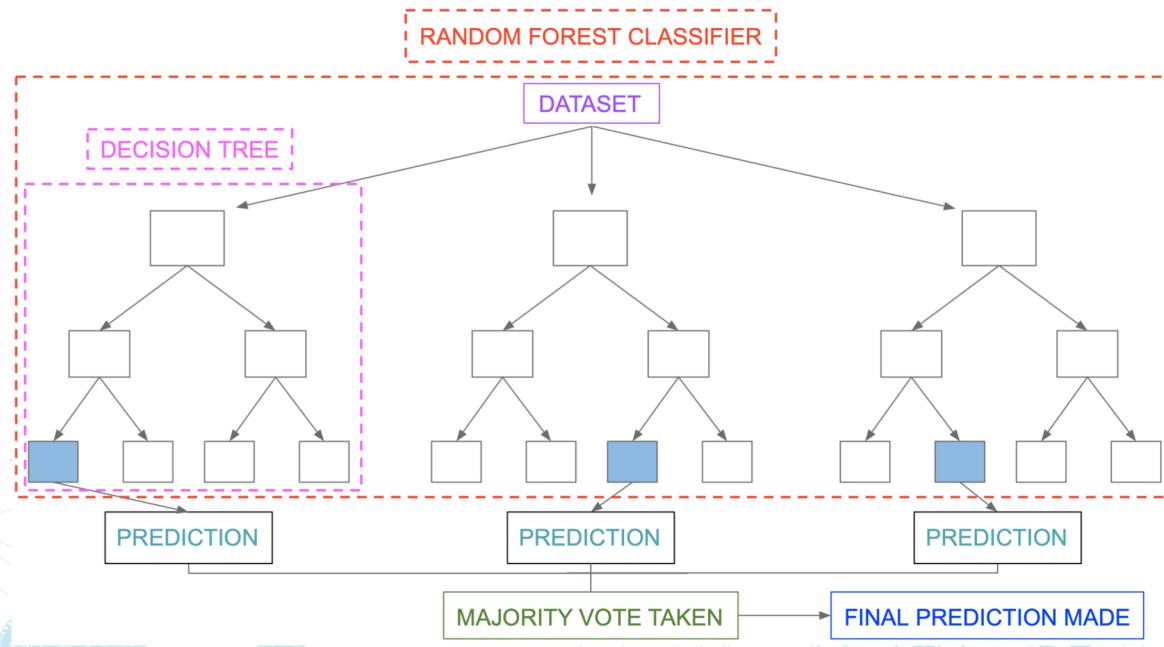


Temperature

Image: <https://medium.com/pursuitnotes/decision-tree-regression-in-6-steps-with-python-1a1c5aa2ee16>

Machine Learning Algorithms

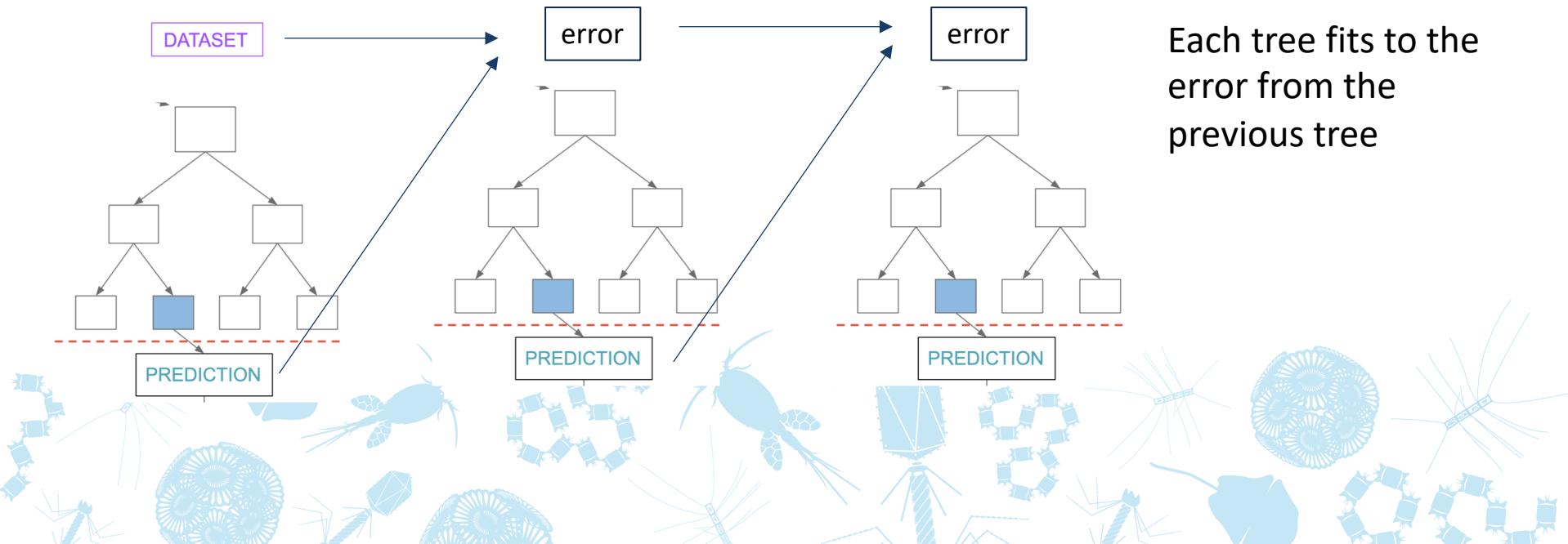
- Random forest



Each tree uses a subset of data and a subset of features

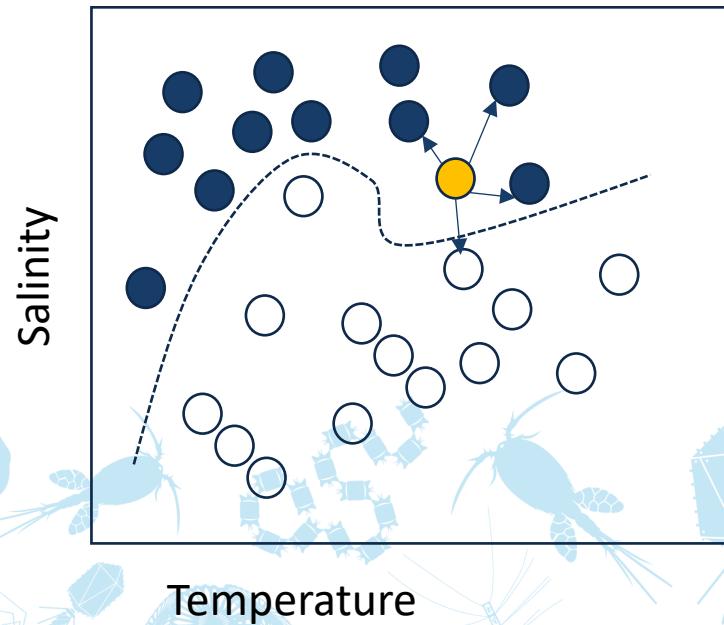
Machine Learning Algorithms

- Boosted regression tree



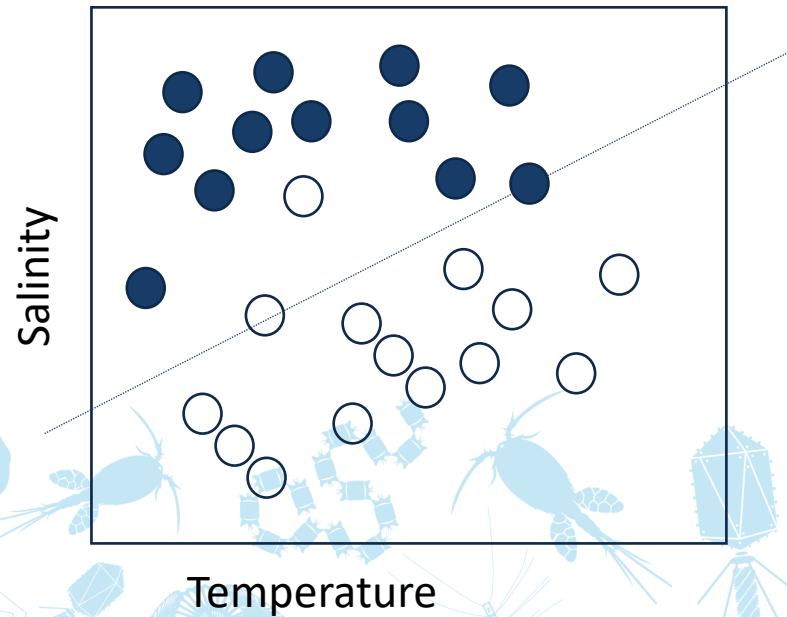
Machine Learning Algorithms

- K – nearest neighbors



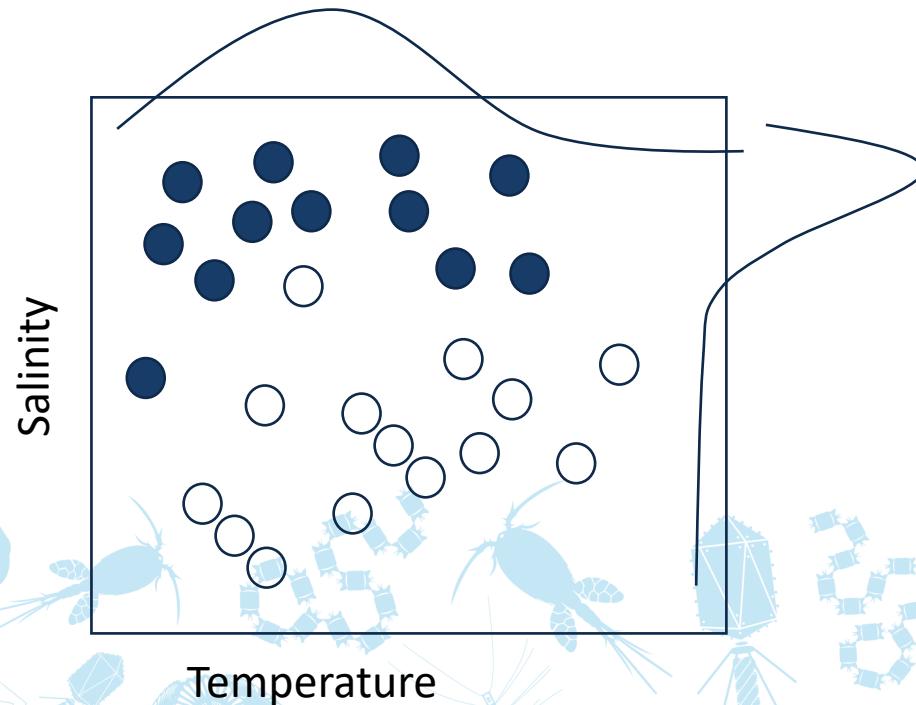
Machine Learning Algorithms

- Support Vector Machine



Machine Learning Algorithms

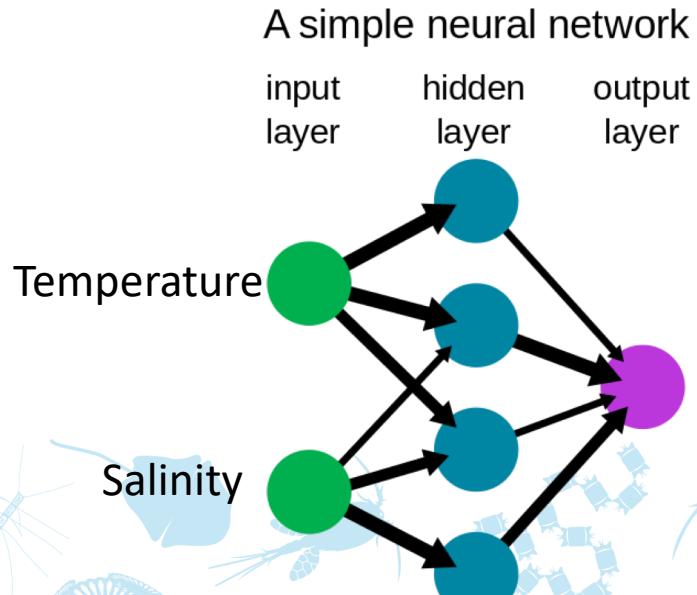
- Maximum entropy



A probability distribution
that maximizes entropy
... minimizes assumptions
... maximizes uncertainty

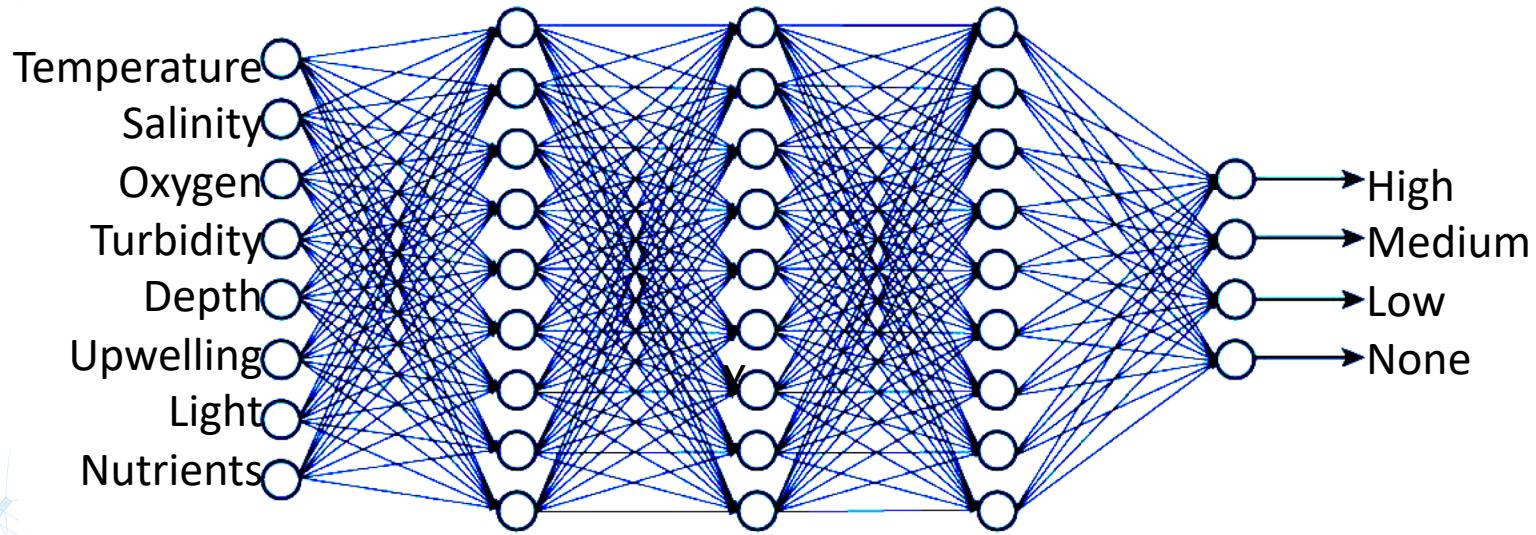
Machine Learning Algorithms

- Neural Network



Machine Learning Algorithms

- Deep Neural Network



Machine Learning Algorithms

- Not magic
- You'll never get further than the information in your data
 - (without theory or some other model)

