

Streaming Machine Learning Regression

Alessio Bernardo

Post-doc @ Politecnico di Milano

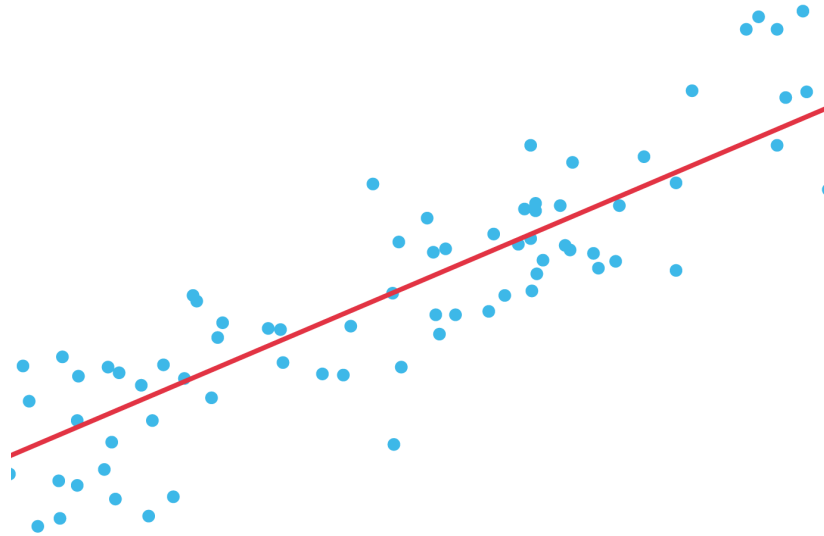
CTO & Co-founder @ Motus ml



POLITECNICO
MILANO 1863



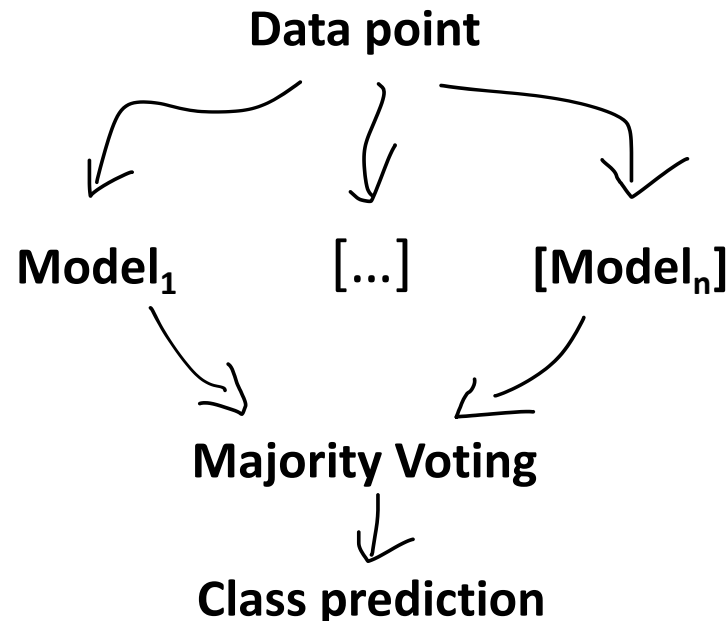
SML Regression models



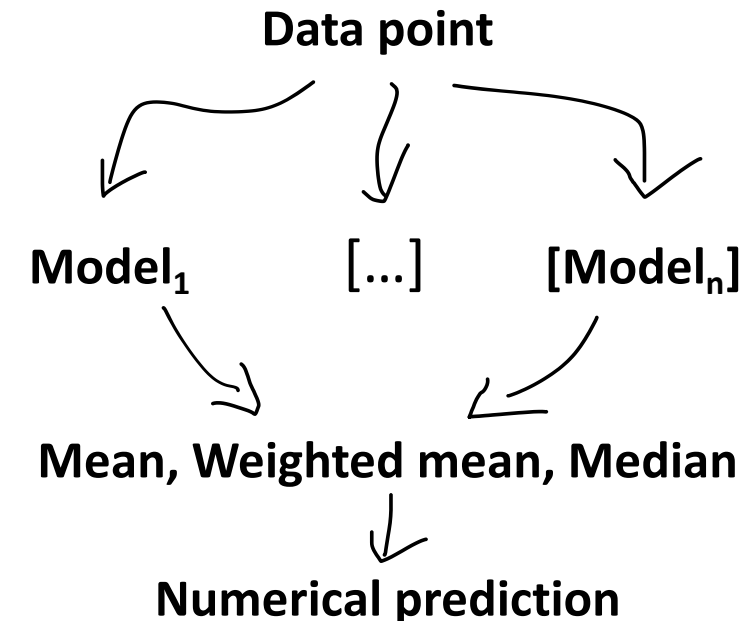
From Classification to Regression

Instead of using a **majority voting** approach, use the **mean** or **median** to aggregate results and compute the predicted target feature

Classification

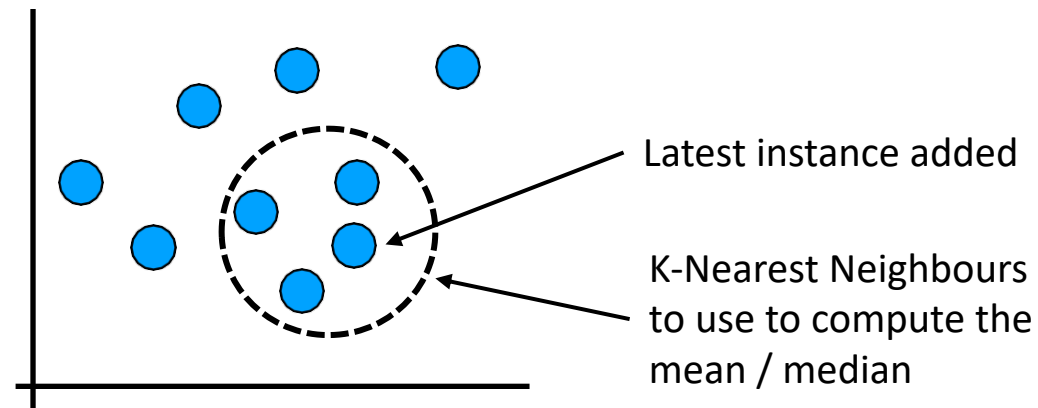


Regression



Online K-Nearest Neighbours Regressor

- Use a fixed size sliding window to save the instances
- Find the k nearest neighbours to the new sample in input
- The predicted target feature is the **mean / weighted mean / median** of the k nearest neighbours' target features



Bifet, A., Pfahringer, B., Read, J., & Holmes, G. **Efficient data stream classification via probabilistic adaptive windows**. ACM symposium on applied computing, 2013

Adaptive Model Rules (AMRules)

- The antecedent of a rule is a set of literals (conditions based on the attribute values)
- The consequent of a rule is a function that minimizes the mean square error of the target attribute computed from the set of samples covered by rule
- Each rule uses a Page-Hinkley test to detect changes and react to changes by pruning the rule set
- Each rule is also equipped with outliers' detection mechanisms to avoid model adaption using anomalous examples
- Multiple rules form a set of rules, similarly to a tree. Hoeffding Bound is used to grow the set

Duarte, J., Gama, J. & Bifet, A. **Adaptive Model Rules From High-Speed Data Streams**. ACM TKDD, 2016

ARF Regressor

- **Base Learners:** Hoeffding Tree Regressors
- **Diversity:** Leveraging Bagging + Local Random Subspaces
- **Combination:**
 - Flat architecture
 - Weighted mean or median as voting strategy
- **Adaptation:** Adaptive window + warning period (train background learners)

Gomes, H. M., et al. **Adaptive random forests for data stream regression**. ESANN, 2018



Exercise 5: Stream Regression [optional]





Credits

- Albert Bifet DATA STREAM MINING 2020-2021 course at Telecom Paris
- Alessio Bernardo & Emanuele Della Valle

Streaming Machine Learning Regression

Alessio Bernardo

Post-doc @ Politecnico di Milano

CTO & Co-founder @ Motus ml



POLITECNICO
MILANO 1863