Cuánta información debe tener?

Tiene un valor agregado? Cada cuándo debo actualizarlo?

qué ventanas de tiempo?

ventajas

algoritmos para dynamic learning

las formas de desarrollar dynamic: model-ops

* Refit: qué datos
* Transfer y finetuning
* cambiar la data

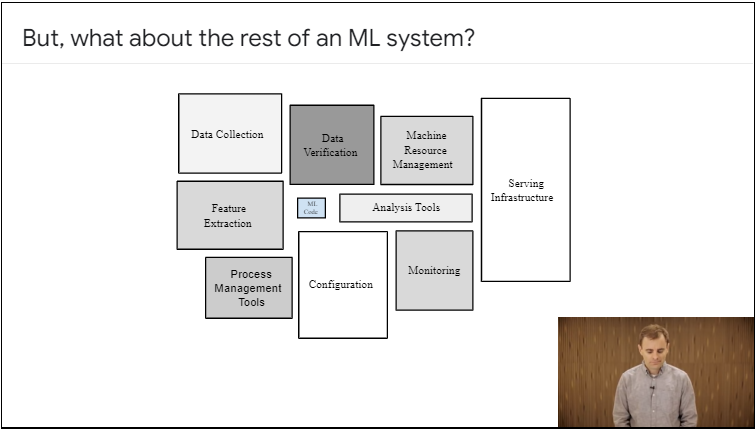
tipos de dynamic learning

xgboost -> algoritmos

**Static Learning vs Dynamic Learning**

(Google)

*Production Machine Learning Systems*



*Static Model - Trained Offline*

* Easy to build and test - use batch train & test, iterate until good
* Requires monitoring of inputs
* Easy to let this grow stale

*Dynamic Model - Trained Online*

* Continue to feed in training data over time, regularly sync out
* Use progressive validation rather than batch training & test
* Needs monitoring, model rollback & data quarantine capabilities

*Static vs Dynamic Inference*

* **Offline inference**: meaning that you make all possible predictions in a batch, using a MapReduce or something similar. You then write the predictions to an SSTable or Bigtable, and then feed these to a cache/lookup table.
  + Upside: don’t need to worry about cost of inference
  + Upside: can likely use batch quota
  + Upside: can do post-verification on predictions on data before pushing.
  + Downside: can only predict things we know about -- bad for long tail.
  + Downside: update latency likely measured in hours or days.
* **Online inference**: meaning that you predict on demand, using a server.
  + Upside: can predict any new item as it comes in -- great for long tail.
  + Downside: compute intensive, latency sensitive -- may limit model complexity.
  + Downside: monitoring needs are more intensive.

**Tipos de Dynamic Learning**

* Refit: qué datos
* Transfer y finetuning
* cambiar la data

**Algoritmos apropiados para Dynamic Learning**

*Gradient Boosting Machines (GBM)*

(Joaquín Amat Rodrigo)

Un modelo Gradient Boosting está formado por un conjunto de árboles de decisión individuales, entrenados de forma secuencial, de forma que cada nuevo árbol trata de mejorar los errores de los árboles anteriores. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo.

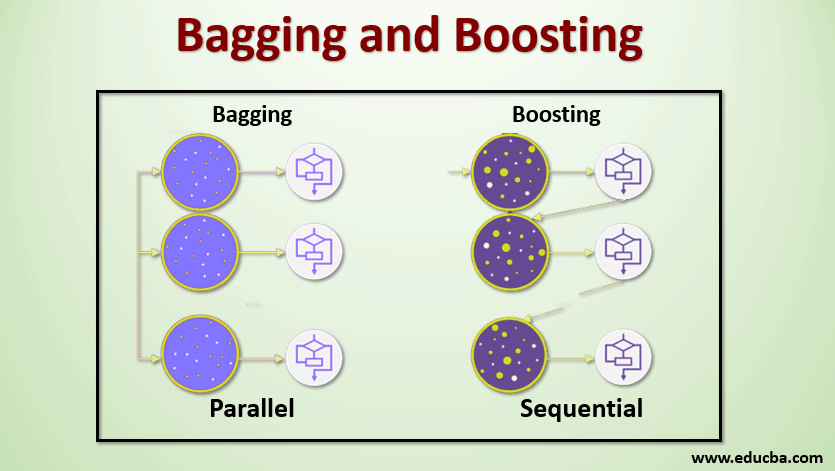
Muchos métodos predictivos generan modelos globales en los que una única ecuación se aplica a todo el espacio muestral. Cuando el caso de uso implica múltiples predictores, que interaccionan entre ellos de forma compleja y no lineal, es muy difícil encontrar un único modelo global que sea capaz de reflejar la relación entre las variables. Los métodos estadísticos y de machine learning basados en árboles engloban a un conjunto de técnicas supervisadas no paramétricas que consiguen segmentar el espacio de los predictores en regiones simples, dentro de las cuales es más sencillo manejar las interacciones. Es esta característica la que les proporciona gran parte de su potencial.

* Ventaja: Son métodos robustos
* Ventaja: Independientes de la distribución de los datos
* Ventaja: Pueden usarse para clasificación o regresión
* Desventaja: No son tan interpretables
* Desventaja: Cuando se usan para regresión, se pierde parte de la información, al agrupar los intervalos de valores en los distintos nodos
* Desventaja: Cuando se usan para regresión, no se pueden usar para predecir fuera del intervalo del conjunto de entrenamiento

Tres de los algoritmos de boosting más empleados son AdaBoost, Gradient Boosting y Stochastic Gradient Boosting. Gradient Boosting es una generalización del algoritmo AdaBoost que permite emplear cualquier función de coste, siempre que esta sea diferenciable. Tiempo después de la publicación del algoritmo de Gradient Boosting, se le incorporó una de las propiedades de bagging, el muestreo aleatorio de observaciones de entrenamiento. En concreto, en cada iteración del algoritmo, el weak learner se ajusta empleando únicamente una fracción del set de entrenamiento, extraída de forma aleatoria y sin reemplazo (no con bootstrapping). Al resultado de esta modificación se le conoce como Stochastic Gradient Boosting y aporta dos ventajas: consigue mejorar la capacidad predictiva y permite estimar el out-of-bag-error.

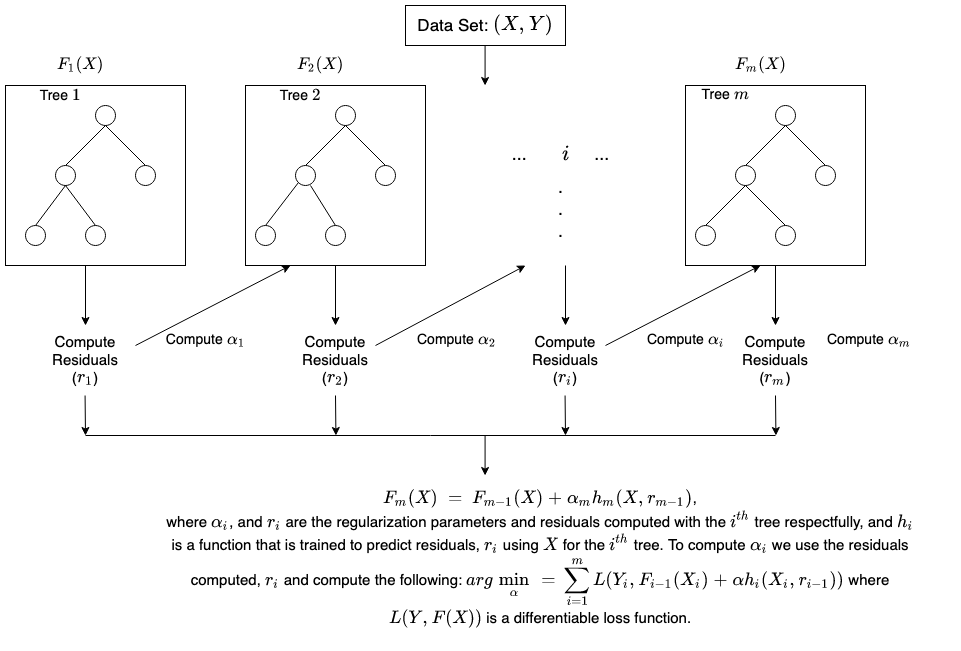
*XGBoost*

(Machine Learners)



* Es un algoritmo paralelizable.
* Para datasets grandes (más de 1000 observaciones, esto se debe a que el boosting mejora con iteraciones) y que el número de variables sea menor que el número de observaciones.
* Para problemas con una mezcla de variables categóricas y numéricas, o sólo numéricas (al contrario que la regresión lineal, el KNN o el K-Means, este algoritmo sí que acepta variables categóricas tal cual están).
* Hay una gran variedad de problemas supervisados en los que otros algoritmos de Deep Learning son capaces de darnos mucho mejor rendimiento que el XGBoost. Entre ellos, los más destacados son:
  + Reconocimiento de Imágenes
  + Problemas de Visión por Computador
  + NLP

(Amazon)



(TDS)

Now that you understand decision trees and gradient boosting, understanding XGBoost becomes easy: it is a gradient boosting algorithm that uses decision trees as its “weak” predictors. Beyond that, its implementation was specifically engineered for optimal performance and speed.

-> Retrain (por construcción) <https://github.com/dmlc/xgboost/issues/3055>

-> Append new trees (?) <https://datascience.stackexchange.com/questions/8120/can-i-fine-tune-the-xgboost-model-instead-of-re-training-it> <https://stackoverflow.com/questions/38079853/how-can-i-implement-incremental-training-for-xgboost>

*ANN*

A

**Incremental Learning**

<https://arthurdouillard.com/post/incremental-learning/>

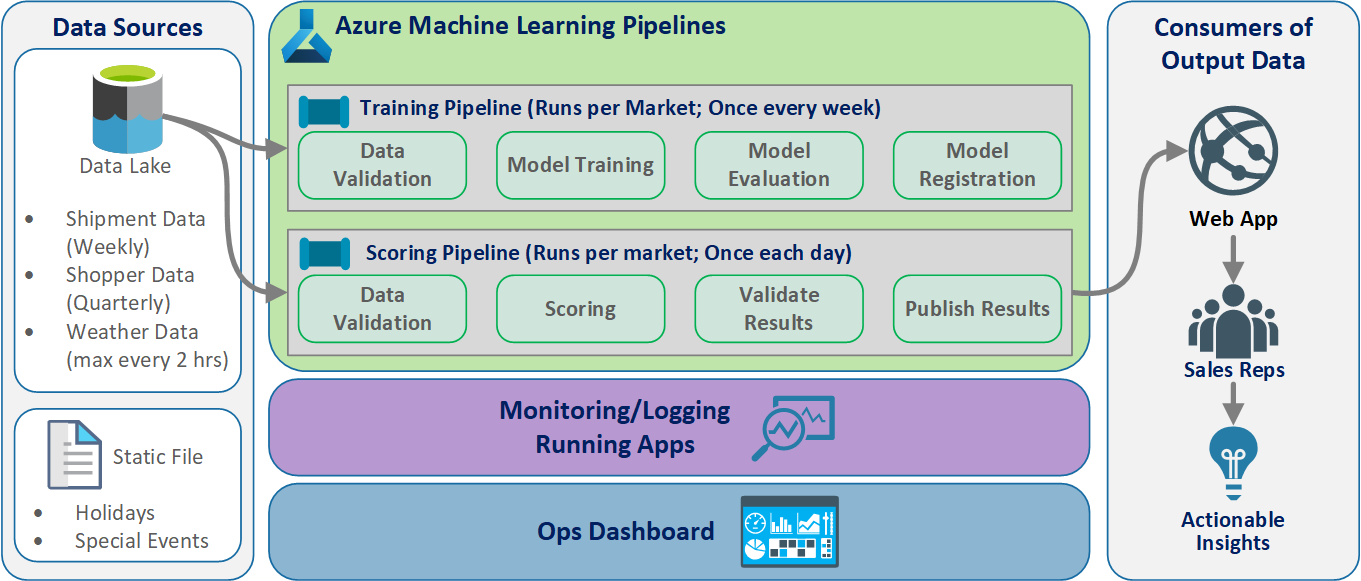
(Papers with Code)

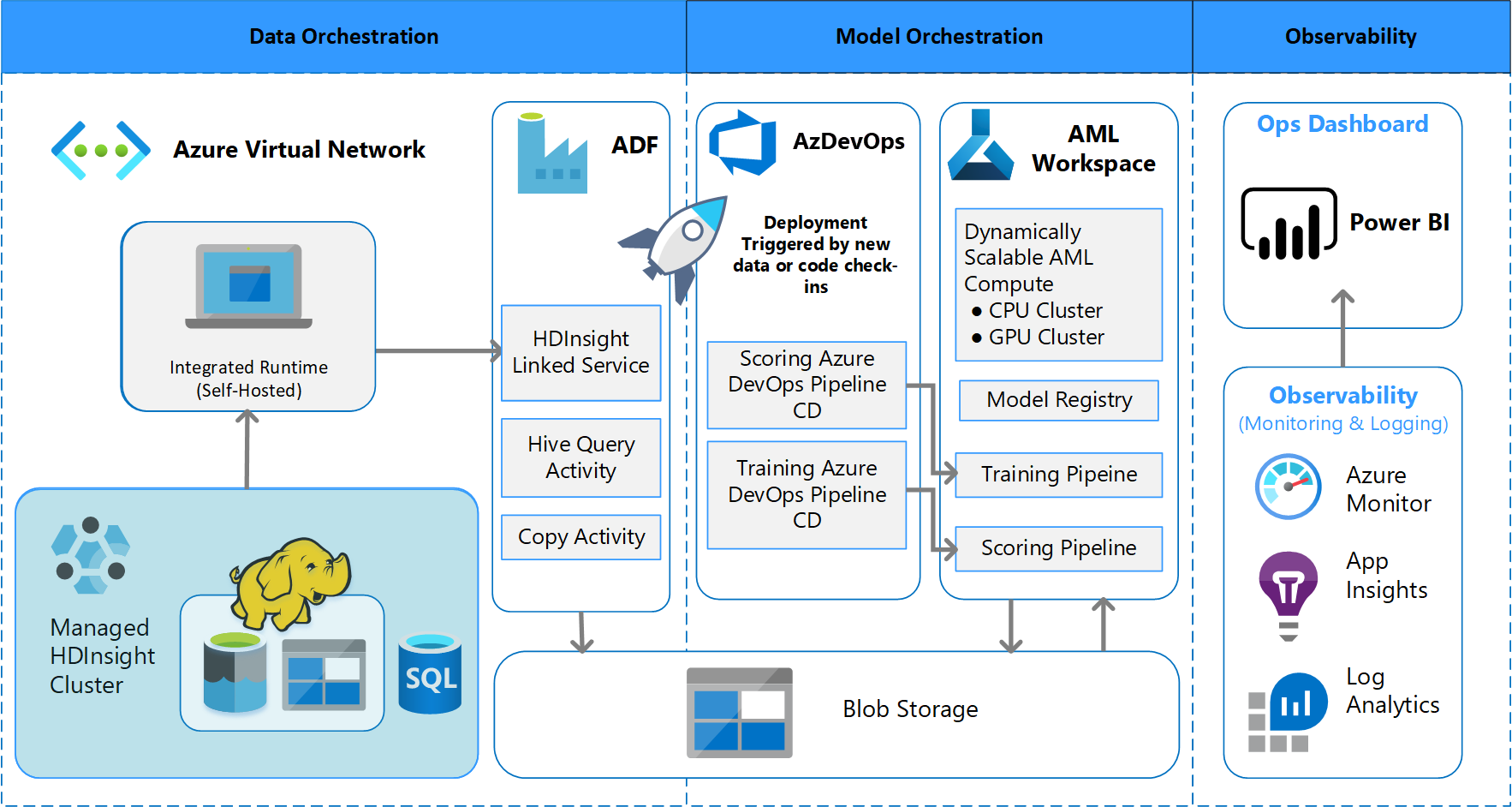
Continual Learning is a concept to learn a model for a large number of tasks sequentially without forgetting knowledge obtained from the preceding tasks, where the data in the old tasks are not available any more during training new ones.

<https://hal.archives-ouvertes.fr/hal-01418129/document>

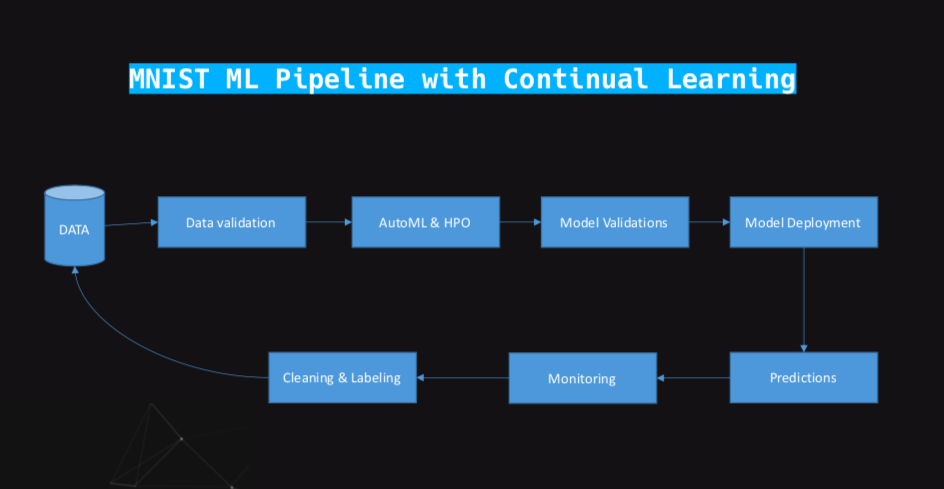
**MLOps (Frameworks)**

(Azure)





(TDS)



**Fuentes**

1. Google. Static vs Dynamic Learning. <https://developers.google.com/machine-learning/crash-course/static-vs-dynamic-training/video-lecture?hl=es_419>
2. Techno Sapien. Dynamic Learning. <https://techno-sapien.com/blog/dynamic-learning>
3. A Simple Dynamic Learning Rate Tuning Algorithm For Automated Training of DNNs. <https://arxiv.org/abs/1910.11605>
4. Joaquín Amat Rodrigo. Gradient Boosting con Python. <https://www.cienciadedatos.net/documentos/py09_gradient_boosting_python.html>
5. Machine Learners. XGBoost con Python. <https://www.themachinelearners.com/xgboost-python/>
6. Amazon. How XGBoost works. <https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html>
7. TDS. XGBoost Theory and Practice. <https://towardsdatascience.com/xgboost-theory-and-practice-fb8912930ad6>
8. TDS. How to apply continual learning to your ML models. <https://towardsdatascience.com/how-to-apply-continual-learning-to-your-machine-learning-models-4754adcd7f7f>
9. Papers with code. Continual Learning. <https://paperswithcode.com/task/continual-learning>
10. PG Madhavan. Static and Dynamical ML. <https://www.datasciencecentral.com/static-dynamical-machine-learning-what-is-the-difference/>
11. PG Madhavan. Generalized Dynamical ML. <http://www.jininnovation.com/Gen_DynamicML_paper.pdf>
12. Amazon. Introducing Dynamic Training for DL with Amazon EC2. <https://aws.amazon.com/es/blogs/machine-learning/introducing-dynamic-training-for-deep-learning-with-amazon-ec2/>
13. Coursera. ML Pipelines on Google Cloud. <https://es.coursera.org/lecture/ml-pipelines-google-cloud/continuous-training-ahas6>
14. Azure. MLOps Technical paper. <https://docs.microsoft.com/es-es/azure/architecture/example-scenario/mlops/mlops-technical-paper>
15. Stats Stack Exchange. Is it legitimate to refit my best model with my test data as a final step in the model building process. <https://stats.stackexchange.com/questions/174026/is-it-legitimate-to-refit-my-best-model-with-my-test-data-as-a-final-step-in-the>
16. Stats Stack Exchange. Training on the full dataset after cross-validation? <https://stats.stackexchange.com/questions/11602/training-on-the-full-dataset-after-cross-validation>