Overview of Machine Learning: Opportunities and Challenges

John D. Lee and Linda Ng Boyle

10/28/2019

HF Considerations for ML and ML Considerations for HF

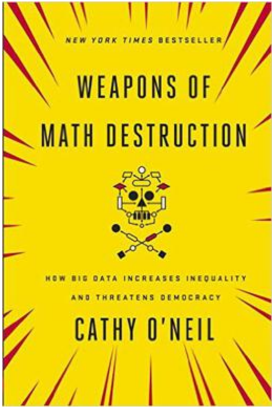
– Human-centered ML represents a looming societal need

– ML deployment sometimes neglects design and systems thinking

– ML mindset for data analysis might save us from the p-value ritual

– ML and Human-centered tradeoffs in model development (e.g., understanding and fairness)

Increasing Centrality of Machine Learning in System Design|Human implications of increasingly powerful algorithms

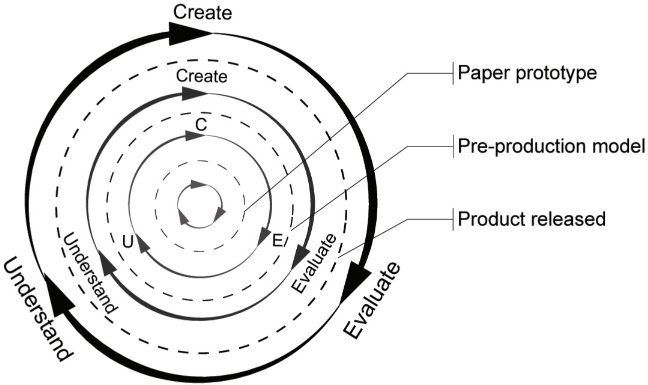


Design Thinking and Systems Thinking for ML

– Design thinking: Empathy and understanding for person-centered solutions

– Systems thinking: 5 whys? 5 whats?

– Sea-level and C-level support for success

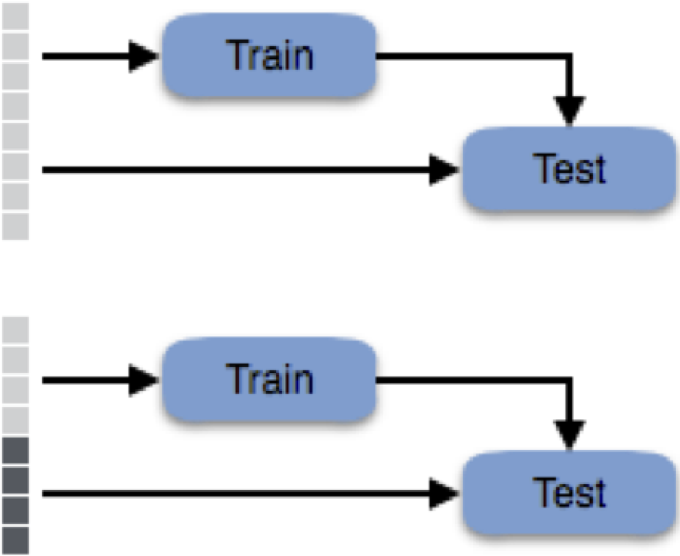


Machine Learning versus Inferential Data Analysis

– Linear regression and logistic regression as machine learning?

ML mindset	Traditional inferential mindset
Prediction	Inference
Cross validation and prediction error	Hypothesis testing and p values
Practical significance	Statistical significance

Holdout validation and prediction error

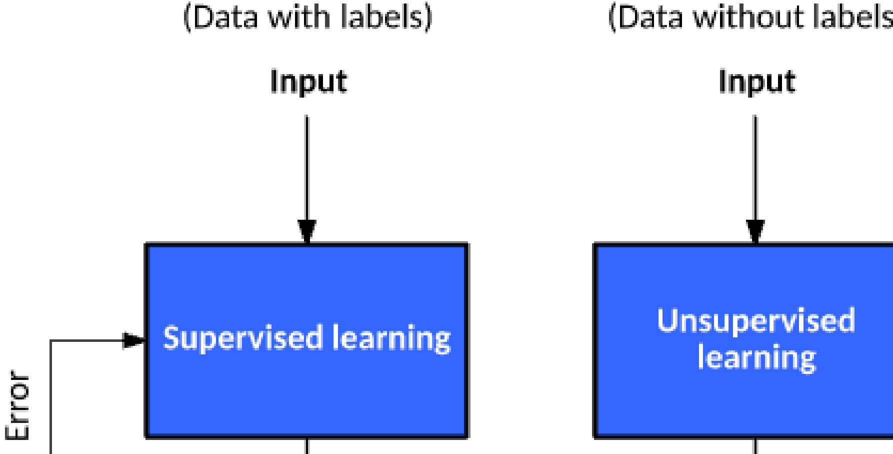


Types of Machine Learning

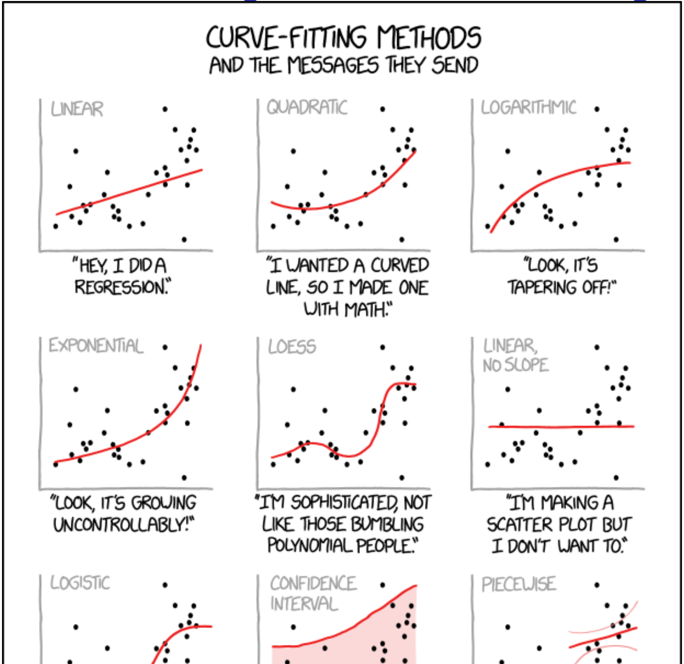
– Supervised learning: Predicting a known output

– Unsupervised learning: Identifying unknown patterns or clusters

– Reinforcement learning: Learning through interactions with a system



Supervised Learning: More than curve fitting?



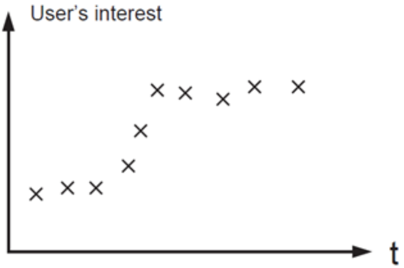
Essential Tradeoffs in ML Design|Variance-bias tradeoff of a model θ

$\Omega(\theta)$ = Model complexity, ideally small variance and good generalization

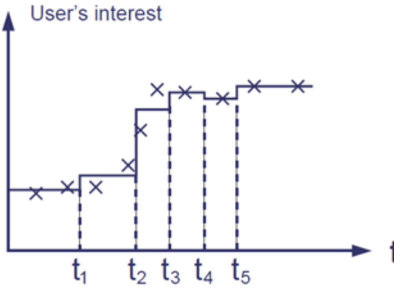
$L(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ = Model error, ideally small bias and precise predictions

Objective function$(\theta) = \Omega(\theta) + L(\theta)$

https://xgboost.readthedocs.io/en/latest/index.html



Observed user's interest on topic k against time t



☒ Too many splits, $\Omega(f)$ is high

Essential Tradeoffs in ML Design| Human-centered tradeoffs

– Different errors: Cost of a miss may differ than a cost of a false alarm

$L(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ Positive errors equally problematic as negative errors??

– Trust depends on more than AUC (Area Under the Receiver-Operator Curve): Hard hits don't compensate for easy misses

– People might value understandable models more than precise models

$\Omega(\theta)$ Model complexity does not typically reflect perceived complexity

Essential Tradeoffs in ML Design| Hyperparameters

Hyperparameters (as in λ below) adjust algorithms to promote:

– A better model fit

– A more robust model

– A more understandable model

Standard regression: $RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$

Ridge regression:
$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. New York: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

HF Considerations for ML and ML Considerations for HF

– Human-centered ML represents a looming societal need

– Design and systems thinking is sometimes neglected with ML deployment

– ML mindset for data analysis might save us from the p-value ritual

– ML and Human-centered tradeoffs in model development