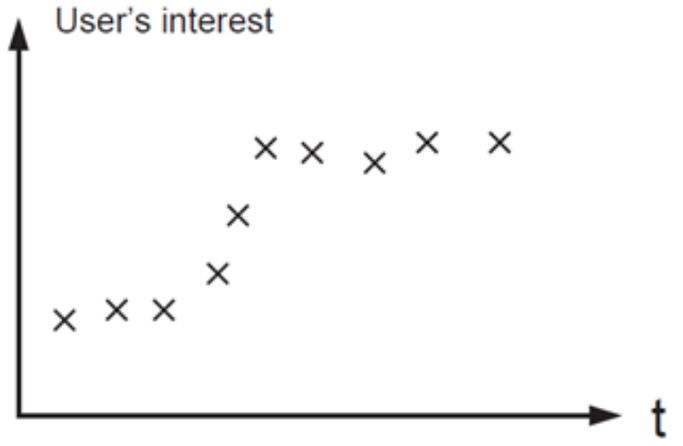
GENERAL ASSEMBLY

XGBOOST

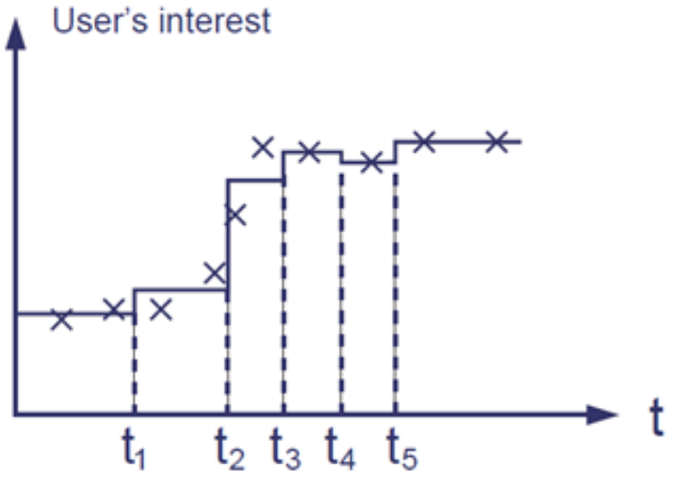
Matt Brems

Data Science Immersive, GA DC

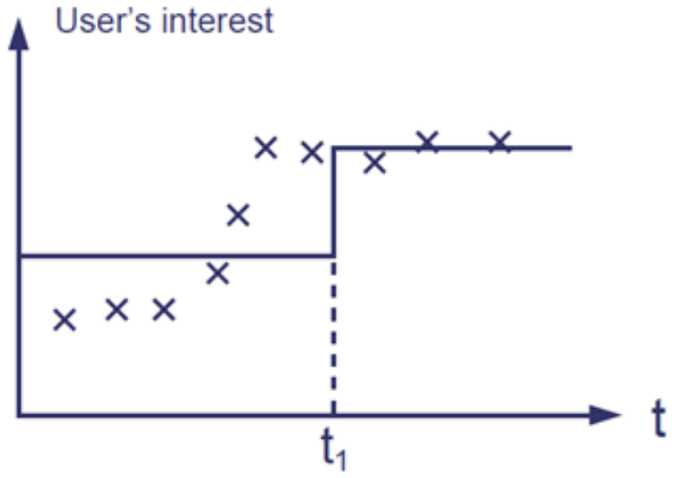
EXAMPLES



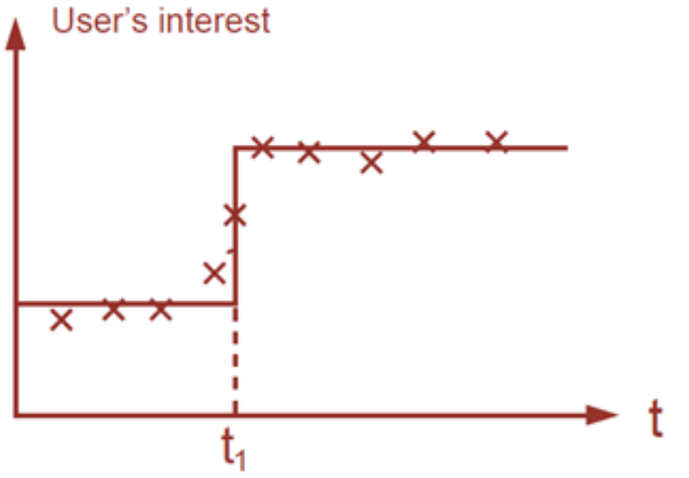
Observed user's interest on topic k against time t



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



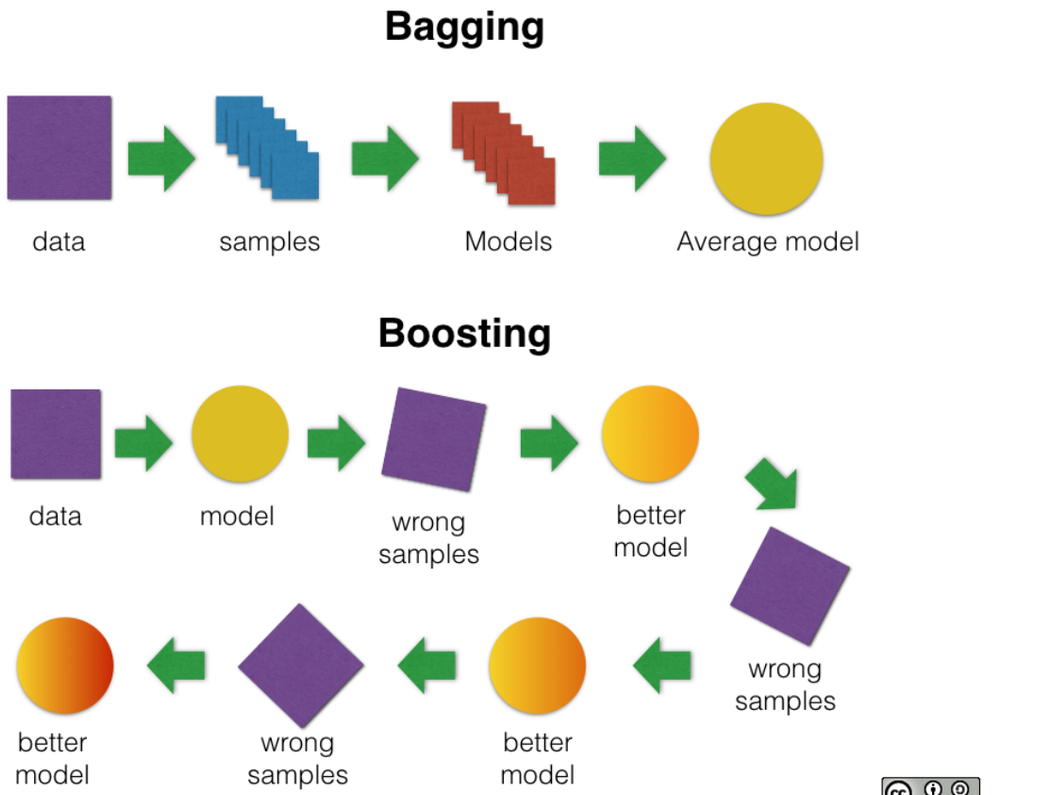
☒ Wrong split point, $L(f)$ is high



☑ Good balance of $\Omega(f)$ and $L(f)$

OPENING

- Recall that **boosting** is a sequential additive modeling technique that fits a model, reweights the observations based on whether they were classified properly or improperly, then fits an additional model, and so on until the algorithm stops.



OPENING

Adaboost

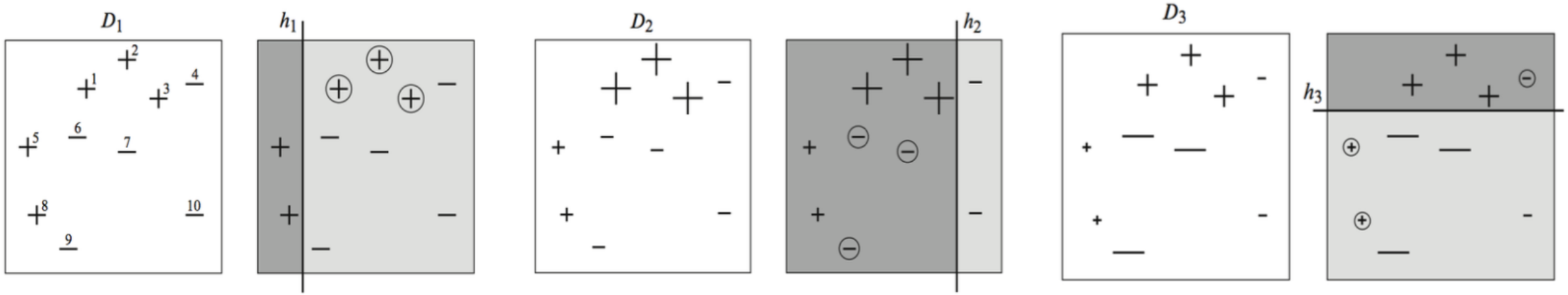


Figure: AdaBoost. Source: Figure 1.1 of [Schapire and Freund, 2012]

OPENING

Adaboost

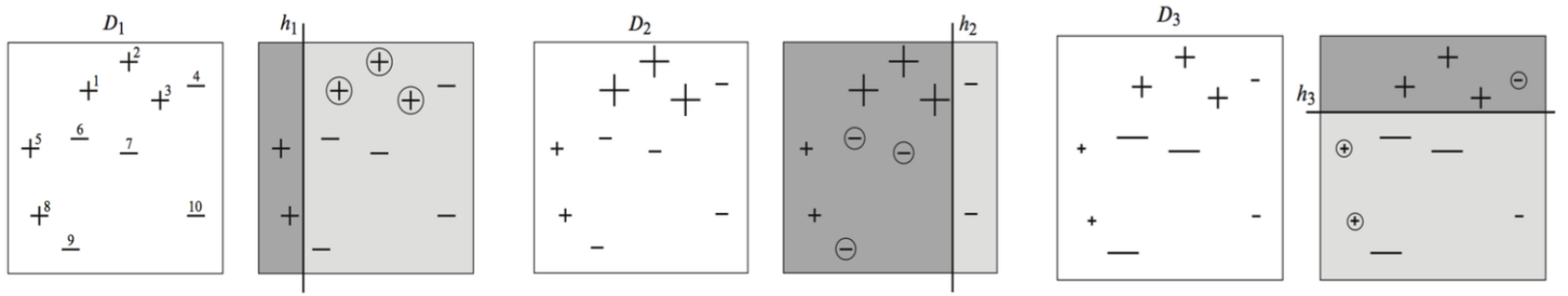


Figure: AdaBoost. Source: Figure 1.1 of [Schapire and Freund, 2012]

- Fit an additive model (ensemble) $\sum_t \rho_t h_t(x)$ in a forward stage-wise manner.
- In each stage, introduce a weak learner to compensate the shortcomings of existing weak learners.
- In Adaboost, "shortcomings" are identified by high-weight data points.

OPENING

Adaboost

$$H(x) = \sum_t \rho_t h_t(x)$$

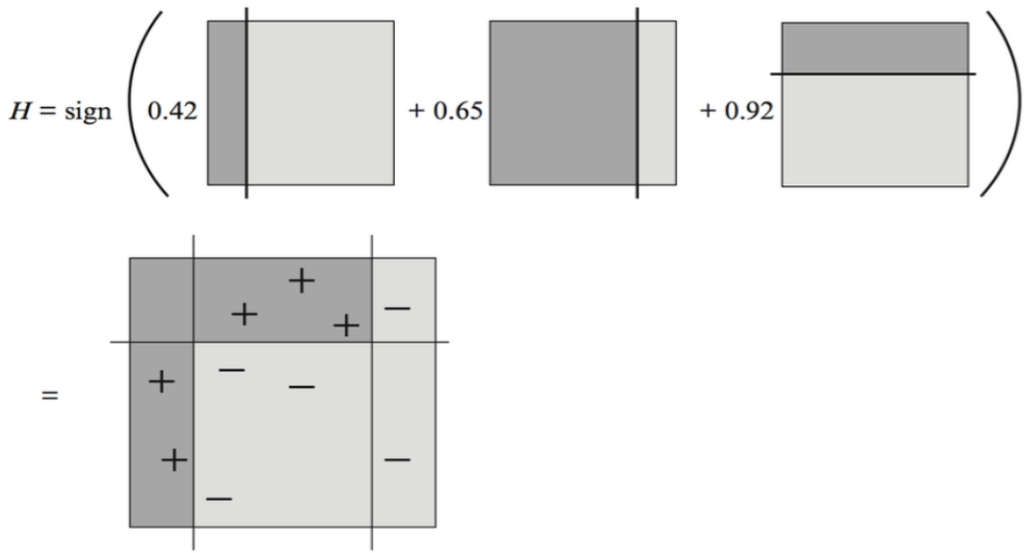


Figure: AdaBoost. Source: Figure 1.2 of [Schapire and Freund, 2012]

OPENING

- This method, AdaBoost, was generalized to include the gradient descent method. The highlights:
  - AdaBoost is an iterative, additive model.
  - AdaBoost iterates by reweighting observations based on misclassification.
  - These misclassifications are “residuals” and, treating these residuals as gradients, we can formulate this as a gradient descent problem with a particular loss function.
  - As a gradient descent problem, we can replace the loss function associated with AdaBoost with other loss functions to make this more general.
  - This generalization is known as “gradient boosting.”

XGBOOST

- XGBoost, or “eXtreme Gradient Boosting,” is similar to gradient boosting.
  - Random Forests are our boosted models.
  - XGBoost Model = RF1 + RF2 + …
- XGBoost has some very good computational benefits.
- XGBoost is capable of handling missing values.