

# CATBOOST

Working, Comparative performance  
and Intuition

# Overview

- *The term CatBoost is an acronym that stands for "Category" and "**Boosting.**"*
- *"Boosting" in CatBoost refers to the **gradient boosting.***
- *CatBoost is an algorithm for **gradient boosting on decision trees.***
- *So in nutshell we can say that CatBoost uses categories of features (categorical or numerical) with an implementation very similar to gradient boosting technique.*

# Gradient Boosting

f1	f2	f3	Salary	Prediction	R1	R2	R3 ..
			50	75	-25	-23	
			70	75	-5	-3	
			80	75	5	3	
			100	75	25	20	

1. Initialize model by average.
2. Calculate the Residual (loss). Here loss = actual – pred.
3. Construct a sequential decision tree to minimize the loss.
4. Final prediction value,  $F(x) = h_0(x) + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots$   
In our case,  $F(x) = 75 + 0.1(-25) + 0.1(-23) = 70.2$

- *CatBoost grows Symmetric trees in sequential manner until selected loss function is no longer minimized.*
- *CatBoost makes several permutation of the given data and each model is trained on a different permutation in greedy way. (logn selections)*
- *CatBoost estimates the training time at first iteration itself.*

## Symmetric trees

