

# NON-LINEAR REGRESSION TECHNIQUES

## LASSO REGRESSION

→ sklearn linear\_model LassoCV

GOOD AT HANDLING COLINEAR VARIABLES  
"SHRINKS" COEFFICIENTS FROM CORRELATED VARIABLES.

## k-NN REGRESSOR

SIMILAR POINTS IN OUR FEATURE SPACE WILL HAVE SIMILAR OUTCOMES.

SIMPLEST NON-LINEAR MODEL.

GOOD BASELINE ERROR FOR OTHER ML MODELS

## DECISION TREES

HIGHLY FLEXIBLE NON-LINEAR MODEL. GREAT AT ESTIMATING INTERACTIONS BETWEEN FEATURES.

PRONE TO OVERFITTING.

DIFICULT ADJUSTMENT OF HYPERPARAMETERS

## ENSEMBLE METHODS

CURRENT STANDARD MODELLING TECHNIQUE VIA BOOSTING ALGORITHMS.

ENSEMBLE STRATEGIES: BAGGING VS BOOSTING

USING catboost AND LIBRARY

## WHAT ARE HYPERPARAMETERS??

PARAMETERS IN A MODEL ARE THE QUANTITIES OUR MODEL ESTIMATES:

\* THE SLOPE & INTERCEPT OF A REGRESSION

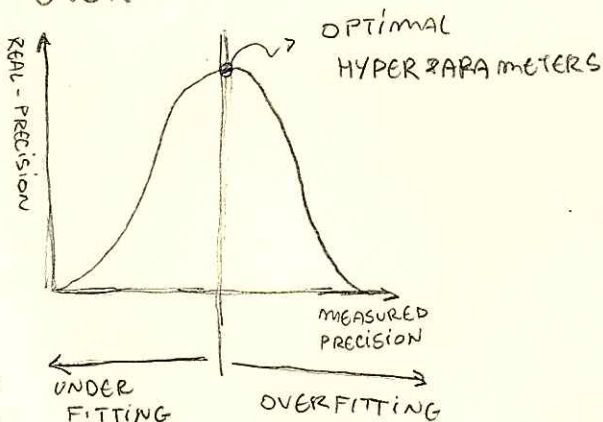
\* NODE VARIABLES & CUT VALUES IN DECISION TREES

HYPERPARAMETERS ARE OPTIMIZATION PARAMETERS FOR EACH ML ALGORITHM

\* K-neighbors in kNN

\* TREE DEPTH & LEAF SIZES IN DECISION TREES

HYPERPARAMETERS REPRESENT THE TRADE-OFF BETWEEN PRECISION & OVERFITTING



REAL PRECISION  $\approx$  TEST ERROR

MEASURED PRECISION  $\approx$  TRAIN ERROR

WE MUST USE OPTIMIZATION TECHNIQUES TO FIND THE BEST HYPERPARAMETERS FOR EACH PROBLEM/MODEL!