REGRESSION TREES

LECTURE 2

SECTION 2

JUNE 2ND

INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

UNIVERSITY of RWANDA

MOTIVATION

MODEL INTERPRETABILITY:

Interpreting our models allows us to evaluate our model and extract new insights about our data.

M1: Linear Regression

$$y = 3.5\,X_1 - 0.1\,X_2 + 100{,}000$$

y: price in \$

X_1: m^2

X_2: distance to city center (m)

Train log-likelihd: -2056

Test log-likelihd: -2564

Is this a reasonable model?

What does the model tell us about housing prices?

M2: Polynomial Regression

$$y = 0.01\,X_1^2 - 0.02\,X_2^2 + 2.1\,X_1 X_2 + 0.1\,X_1 - 0.2\,X_2 + 10{,}000$$

y: price in \$

X_1: m^2

X_2: distance to city center (m)

Train log-likelihd: -54

Test log-likelihd: -58

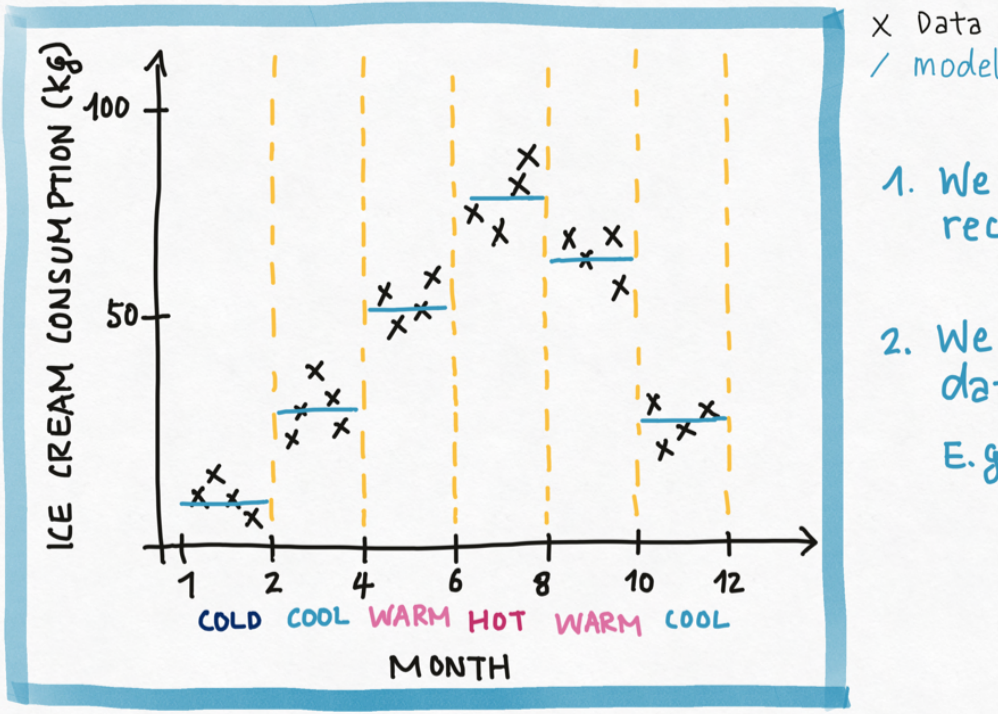
Is this a reasonable model?

What does the model tell us about housing prices?

Which model should we use for predicting house prices?

PIECEWISE-LINEAR MODELS:

We can retain the interpretability of linear models as well as the flexibility of polynomial models by using piecewise-linear models:



1. We divide the input space X into rectangular regions.

2. We fit a linear model to the data in each region.

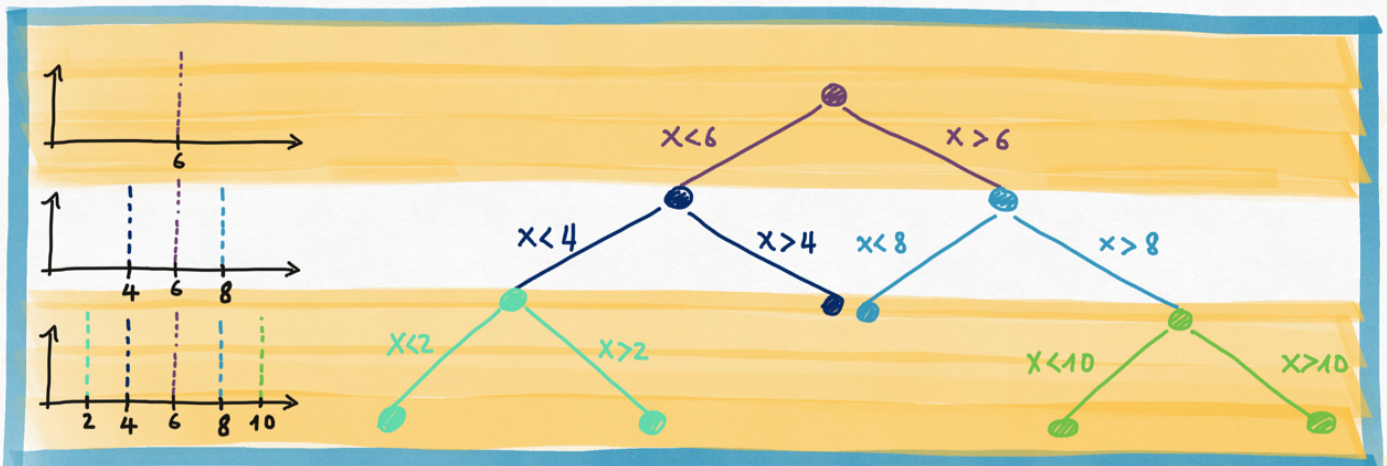
   E.g. we can take the average y-value of the data in each region

REGRESSION TREES

REGRESSION TREES:

In order to fit a piecewise-linear function, we need to divide the domain space sequentially.

The easiest way to keep track of these divisions is through a tree:



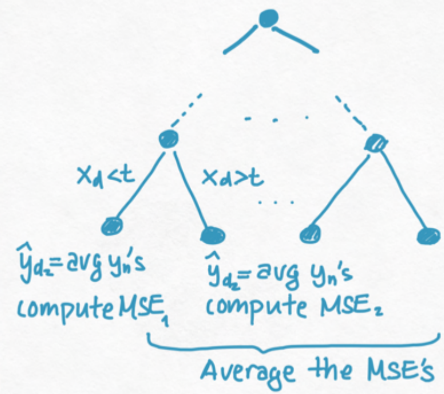
Every tree corresponds to division and every division can be described by a tree.

TRAINING REGRESSION TREES:

Given a dataset $D = \{(x^{(1)}, y^{(1)}) \dots (x^{(N)}, y^{(N)})\}$, where $x^{(n)} \in \mathbb{R}^D$ how do we decide on the sequence of cuts to divide our domain space?

Regression tree algorithm:

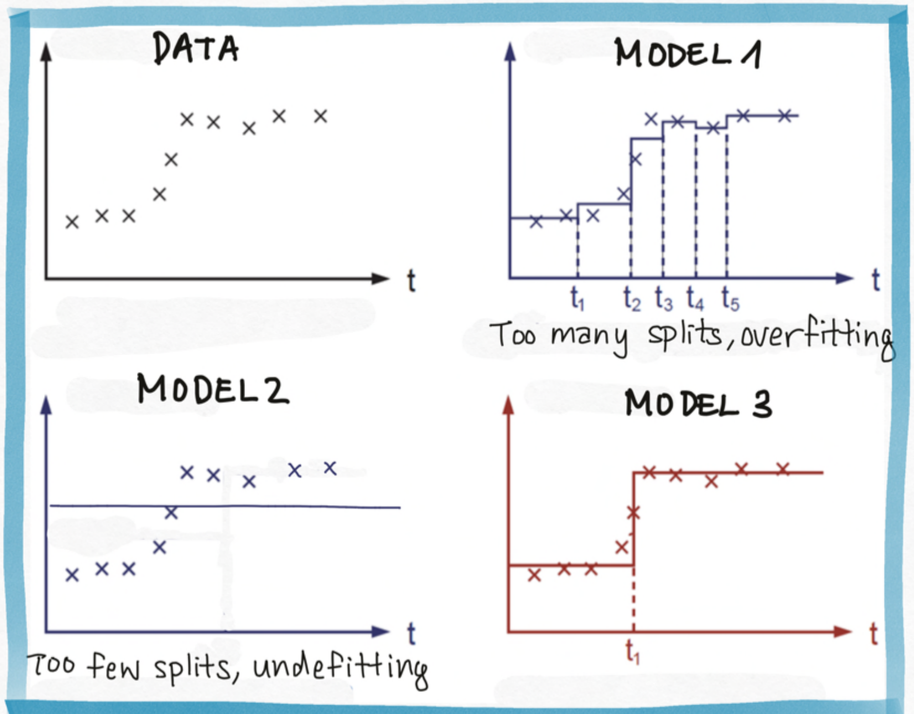
0. Start with a node containing all the data.
1. If stopping condition is not met:
   A. Choose an input dimension d and threshold t and divide the data in the node into two sets $x_d < t$ and $x_d > t$, these become new nodes.
   B. Choose d, t such that the MSE of the piecewise-linear model is minimal.
2. For each new node, repeat step 1.



The stopping condition is usually a maximum depth or a minimum MSE.

UNDERFITTING & OVERFITTING:

We need to decide how many cuts to make, i.e. the depth of the tree.



When the tree is deep, we create many regions, this can cause overfitting.

When the tree is shallow, we create very few regions, this can cause underfitting.

Just like with polynomial models there is a trade-off:

flexibility of model v.s. robustness to noise