Lecture 19 announcements

- Midterm grading is in progress; please be patient
- My office hours tomorrow will be 11 AM - 12 PM only
- HW 10 has been posted

Lecture 19 outline

- Adaptive basis function models
- Classification and Regression Trees (CART)

NOTE: CORRECTION ON PAGE 5.

ADAPTIVE BASIS FUNCTION MODEL (ABM)

[MURPHY 16.1, LAST 2 PARAGRAPHS].

$$\hat{f}(\underline{x}) = w_0 + \sum_{m=1}^{M} w_m \phi_m(\underline{x})$$

IN WHICH $\phi_m(\underline{x})$ IS LEARNED FROM THE DATA.

IF THE $\phi_m(\underline{x})$ ARE PARAMETRIC, THEN:

$$\phi_m(\underline{x}) = \phi(\underline{x}; \underline{v}_m)$$

PARAMETERS OF ϕ_m, TO BE LEARNED FROM THE DATA.

CART

[MURPHY 16.2.0–16.2.4, INCLUSIVE]

(ALSO CALLED "DECISION TREES")

MODEL:

$$\hat{f}(\underline{x}) = \sum_{m=1}^{M} w_m \mathbb{I}(\underline{x} \in R_m)$$

$$= \sum_{m=1}^{M} w_m \phi(\underline{x}; \underline{v}_m)$$

IN WHICH R_m = $m^{\underline{th}}$ REGION,

ν_m = PARAMETERS OF ϕ_m (LEARNED FROM DATA).

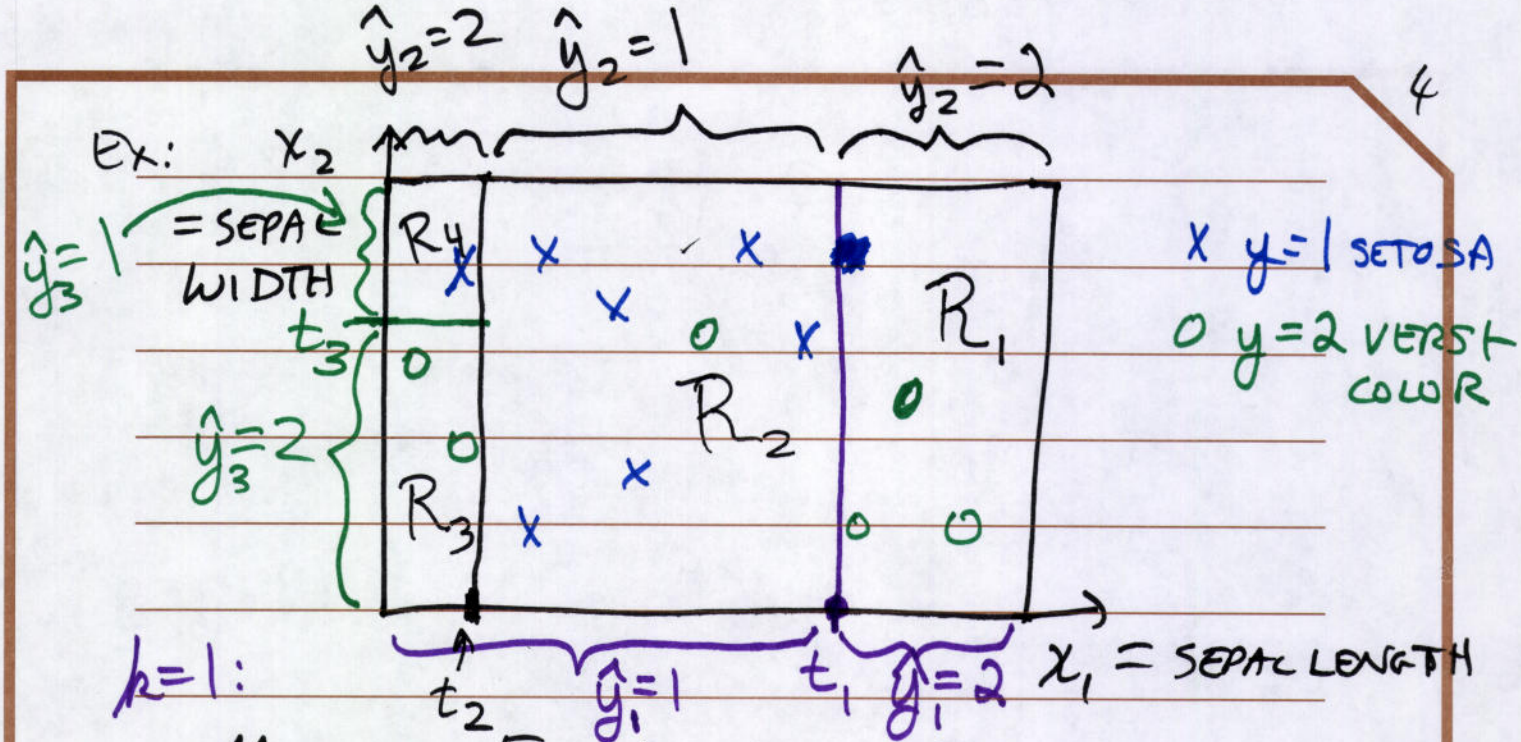
$\Rightarrow \hat{f}(\underline{x})$ IS A PIECEWISE CONSTANT APPROX. TO $f(\underline{x})$.

CART: FORMS A TREE, AND A SET OF REGIONS R_m IN FEATURE SPACE.

TREE & REGIONS R_m COME FROM

RECURSIVE SPLITTING OF A REGION INTO 2, WITH SPLIT PERFORMED BY THERSHOLDING ONE COORDINATE VARIABLE (ONE FEATURE).

Ex:



$k=1$:

MINIMIZE E_{in} AT EACH ITERATION.

START: R_0 : ALL FEATURE SPACE $\in \hat{y}=1$

$\Rightarrow E_{in} = 6/13$.

$k=1$: CHOOSE x_1 OR x_2 TO THRESHOLD

PICK THRESHOLD VALUE t_1

CHOOSE ~~CLAS~~ REGION LABELS.

$\Rightarrow E_{in} =$ ~~[illegible]~~ $3/13$

$k=2$: CHOOSE R_i TO SPLIT

CHOOSE x_1 OR x_2 TO THRESHOLD

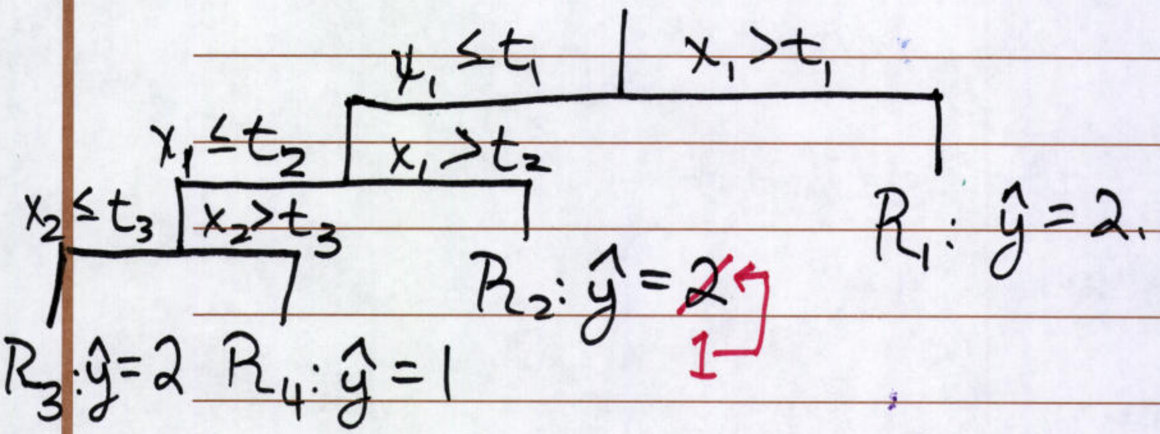
" THRESH. VALUE t_2.

CHOOSE REGION LABELS (2 NEW REGIONS)

$\Rightarrow E_{in} = 2/13$

$k=3$: SEE ABOVE. $\Rightarrow E_{in} = 1/13$.

TREE FOR ABOVE:



$\exists$ A VARIETY OF HALTING CONDITIONS, SUCH AS:

- MAX. DEPTH OF TREE.
- MIN. REDUCTION IN COST FCN. TO SPLIT A REGION.
- MIN. # OF DATA PTS. IN A FINAL REGION [ref: Murphy].

NOTE: OPTIMAL FINAL TREE BALANCES COMPLEXITY OF TREE WITH N AND COMPLEXITY OF TARGET FCN.

MORE RIGOROUSLY:

AT EACH ITERATION (EA. NODE OF TREE), WE DIVIDE ONE REGION R_m INTO TWO, BY THRESHOLDING ONE FEATURE x_j; THUS:

AT $k^{\underline{th}}$ ITERATION:

$$\min_{m,\, j,\, t_k,\, w_{m_1},\, w_{m_2}} \left\{ f_{obj}^{(k)} \left(w_{m_1}, w_{m_2}, \mathcal{D}; j, t_k, m \right) \right\}$$

in which:

$$f_{obj}^{(k)} = \text{cost}_k \left\{ (\underline{x}_i, y_i) \in \mathcal{D} \right\}_{\text{after split of } R_m}$$

FOR COST FCNS. THAT ARE ADDITIVE BY REGION, THAT IS:

$$\text{cost} \left\{ (\underline{x}_i, y_i) \in \mathcal{D} \right\} = \alpha \sum_{m=1}^{M} \text{cost} \left\{ (\underline{x}_i, y_i) \in R_m \right\}$$

WE CAN INSTEAD USE THE INCREMENTAL CHANGE IN COST:

$$\hat{f}_{obj}^{(k)} = \Big[\text{cost} \left\{ (\underline{x}_i, y_i) \in R_{m_1} \right\} + \text{cost} \left\{ (\underline{x}_i, y_i) \in R_{m_2} \right\} - \text{cost} \left\{ (\underline{x}_i, y_i) \in R_m \right\} \Big]$$

IN WHICH:

$$R_{m_1} : R_m \cap \{x_j \leq t_k\}$$

$$R_{m_2} : R_m \cap \{x_j > t_k\}$$

FOR REGRESSION, COST FCN. IS TYPICALLY:

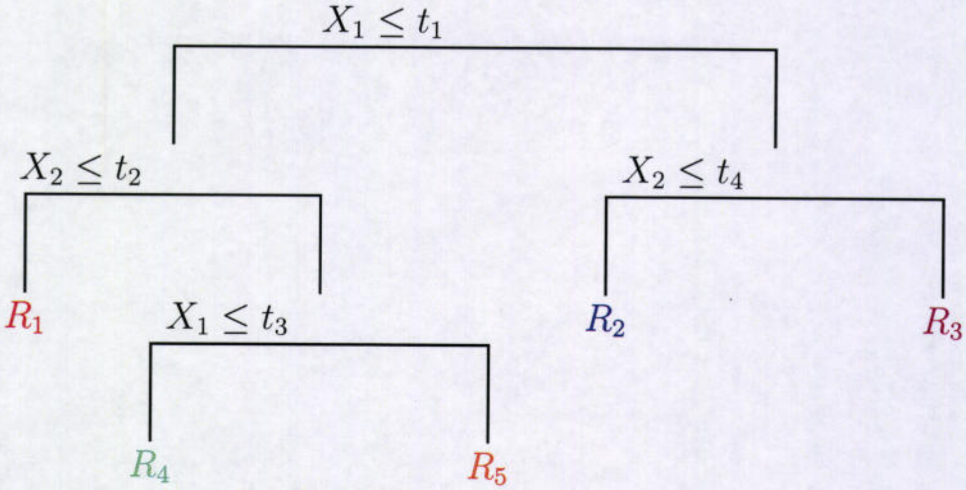
$$\text{cost}\{(\underline{x}_i, y_i) \in R_{m'}\} = \sum_{\underline{x}_i \in R_{m'}} (y_i - w_{m'})^2$$

FOR GIVEN $R_{m'}$, THIS IS MINIMIZED BY:

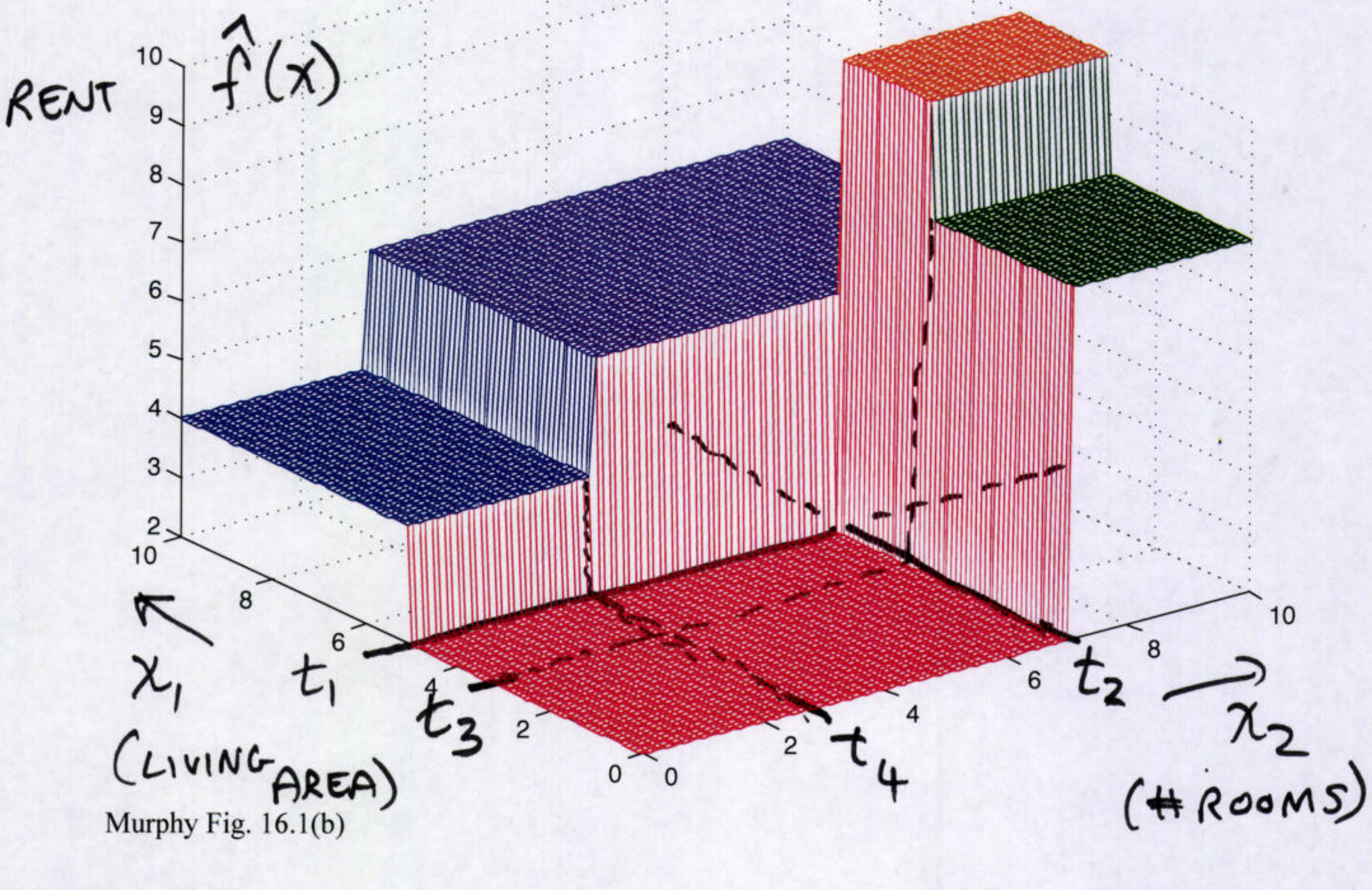
$$w_{m'} = w^*_{m'} = \bar{y}_{R_{m'}} \triangleq \frac{1}{N_{R_{m'}}} \sum_{\underline{x}_i \in R_{m'}} y_i$$

[MURPHY FIG. 16.1]

NOTE: TO SAVE ON COMPUTATION, CART TYPICALLY CYCLES THROUGH ALL REGIONS R_m, $m = 1, 2, \cdots$, SPLITTING EACH INTO 2 IF THE HALTING CONDITION ISN'T MET, INSTEAD OF FINDING THE BEST REGION TO SPLIT AT EACH ITERATION.



Murphy Fig. 16.1(a)



Murphy Fig. 16.1(b)