Social Data Science

Soc 114
Winter 2025

Supervised Machine Learning

Learning goals for today

By the end of class, you will be able to

- use statistical learning to estimate when data are sparse
- work with models that are "wrong"

Prediction function

A **prediction function** is an input-output function:

- input a vector of predictors $\vec{x}$
- output a predicted outcome $\hat{y} = \hat{f}(\vec{x})$

	Input	**Output**
	$\vec{x} \longrightarrow$	$\hat{y}$
Example:	{Sex,Age}	Probability of Employment Given Sex and Age

cases for learning

Age	Sex	Employed
26	F	1
40	M	1
61	M	0
32	F	1

case to predict

Age	Sex	Employed
63	F	?

OLS is a prediction function

Input $\vec{x} \rightarrow$ Output $\hat{y}$

$$\hat{y} = \hat{f}(\vec{x}) = \hat{\beta}_0 + \hat{\beta}_1(\text{Sex} = \text{Male}) + \hat{\beta}_2(\text{Age})$$

- Learn $\hat{f}$ in a **learning sample** with $\{\vec{x}_i, y_i\}_{i=1}^n$
  - Computer finds $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ that predict well in the learning sample
- At a new $\vec{x}$ value, predict $\hat{f}(\vec{x})$

Logistic regression is a prediction function

Input $\vec{x} \rightarrow$ Output $\hat{y}$

$$\hat{y} = \hat{f}(\vec{x}) = \text{logit}^{-1}\left(\hat{\beta}_0 + \hat{\beta}_1(\text{Sex} = \text{Male}) + \hat{\beta}_2(\text{Age})\right)$$

- Learn $\hat{f}$ in a **learning sample** with $\{\vec{x}_i, y_i\}_{i=1}^n$
    - Computer finds $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ that predict well in the learning sample
- At a new $\vec{x}$ value, predict $\hat{f}(\vec{x})$

Matching is a prediction function

Input $\vec{x} \rightarrow$ Output $\hat{y}$

$$\hat{y} = \hat{f}(\vec{x}) = y_j$$

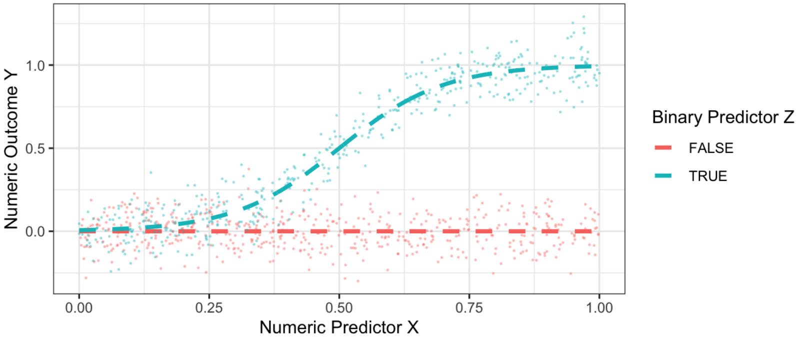
where unit j is the best match among the learning sample, which minimizes a distance from the case to predict: $d(\vec{x}, \vec{x}_j)$ is small

- Learn $\hat{f}$ in a **learning sample** with $\{\vec{x}_i, y_i\}_{i=1}^n$
  - Computer finds j with $\vec{x}_j$ most similar to $\vec{x}$
- At a new $\vec{x}$ value, predict $\hat{f}(\vec{x})$

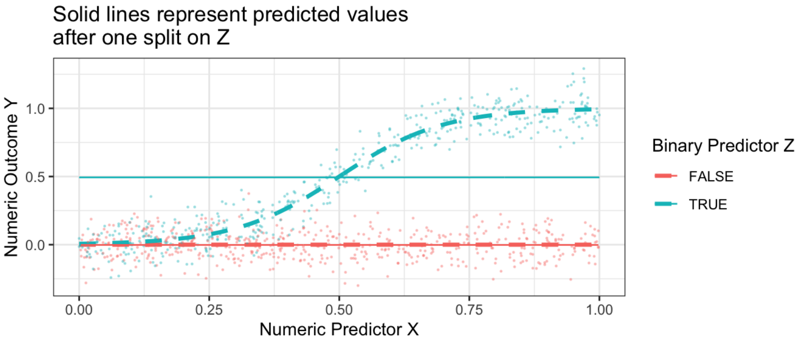
There are many prediction functions

- input a vector of predictors $\vec{x}$
- output a predicted outcome $\hat{y} = \hat{f}(\vec{x})$

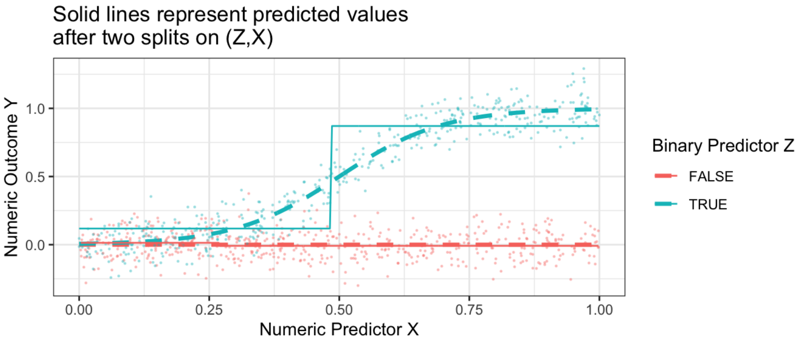
Trees as a prediction function



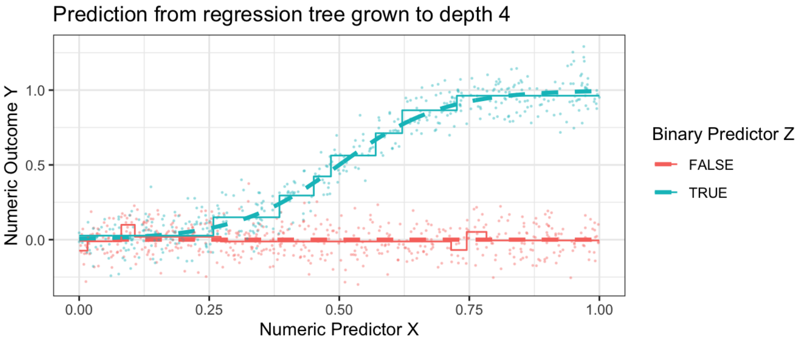
Trees as a prediction function



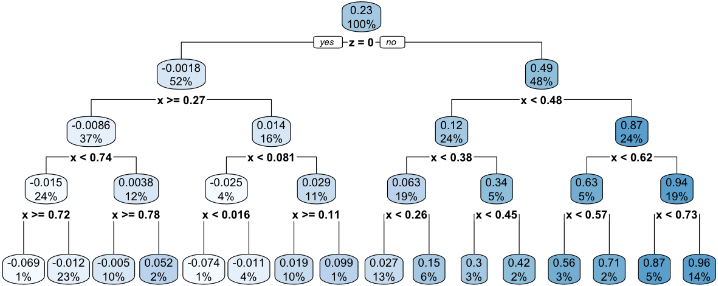
Trees as a prediction function



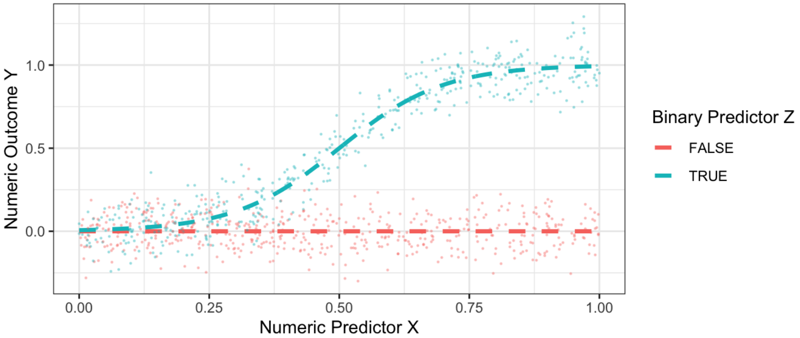
Trees as a prediction function



Trees as a prediction function

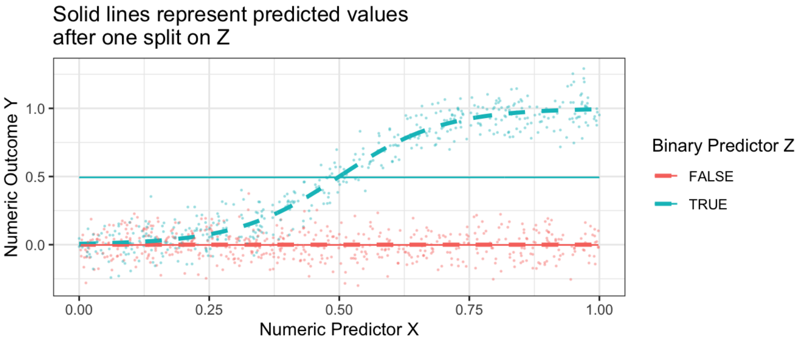


Trees as a prediction function: How that worked

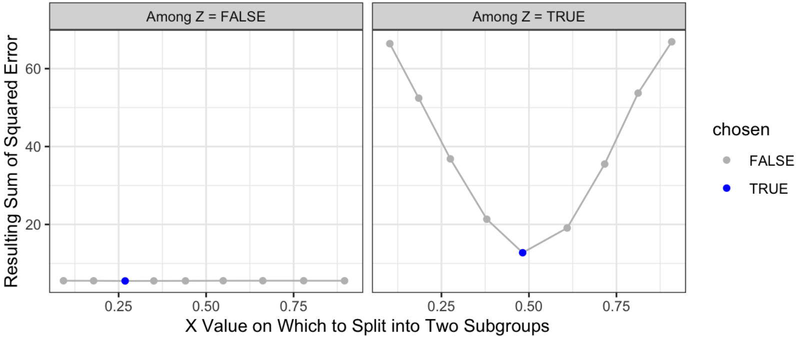


Trees as a prediction function: How that worked

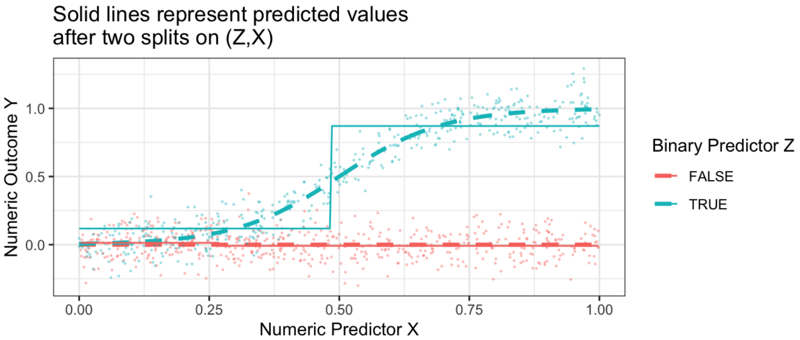
Trees as a prediction function: How that worked



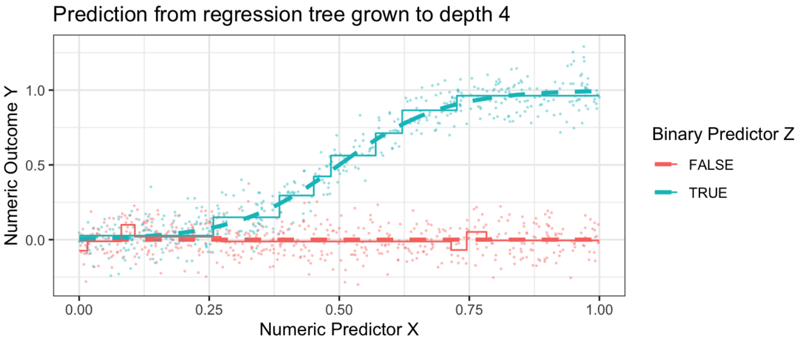
Trees as a prediction function: How that worked



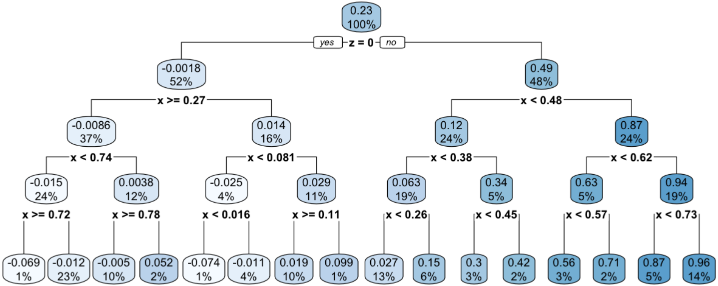
Trees as a prediction function: How that worked



Trees as a prediction function: How that worked



Trees as a prediction function: How that worked



A tree can be interpretable

TODO: real data example

Learning goals for today

By the end of class, you will be able to

- use statistical learning to estimate when data are sparse
- work with models that are “wrong”