

# MODELS FOR MACHINE LEARNING

# Models we've Constructed

## Regression, Linear

$$w_0 x_0 + w_1 x_1 + w_2 x_2 = \hat{y}$$

2 features  
3 Parameters

"Linear transformation scheme"

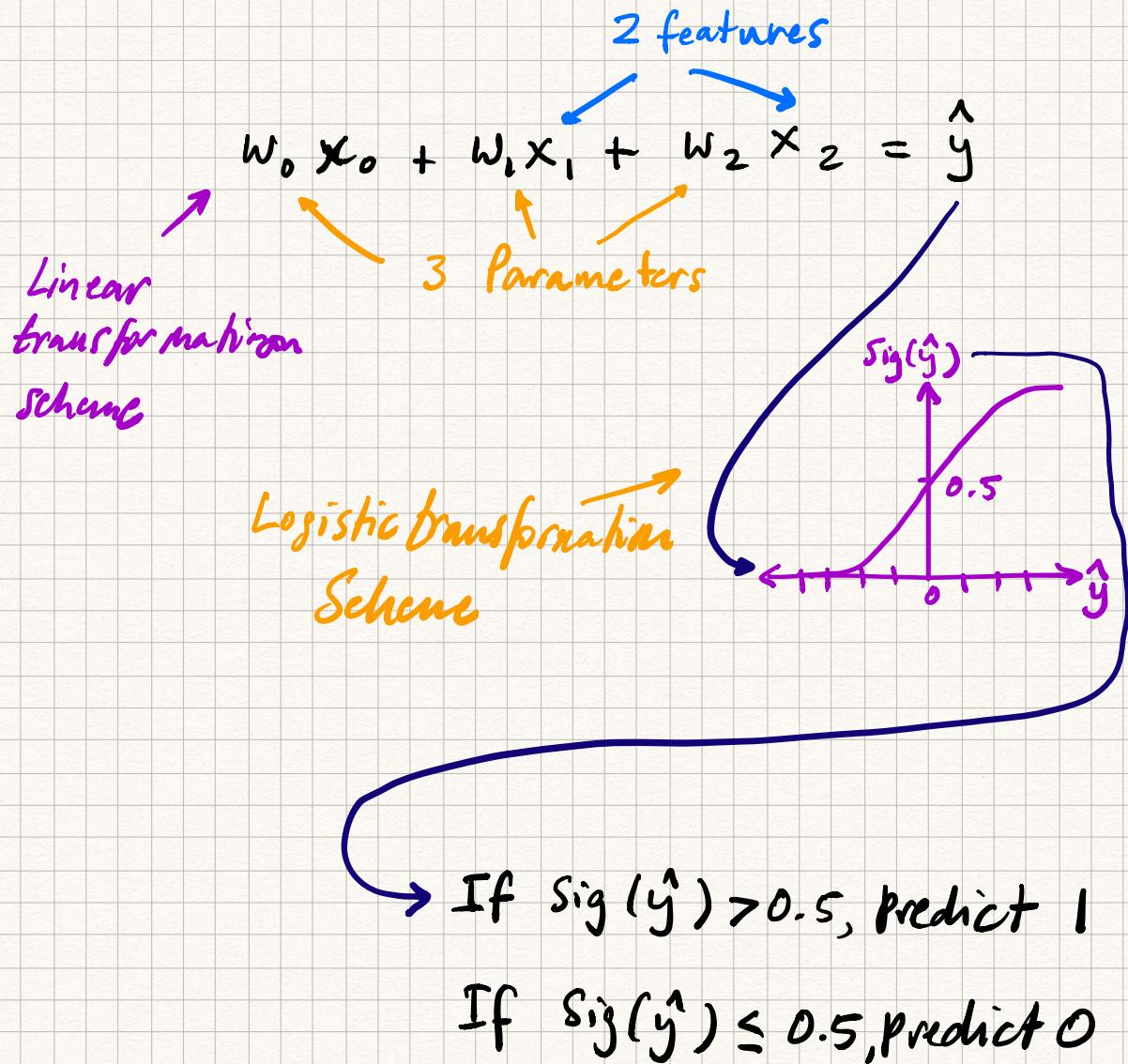
## Regression, Non-Linear

$$w_0 x_0 + w_1 x_1^2 + w_2 x_1 x_2 + w_3 x_2^3 = \hat{y}$$

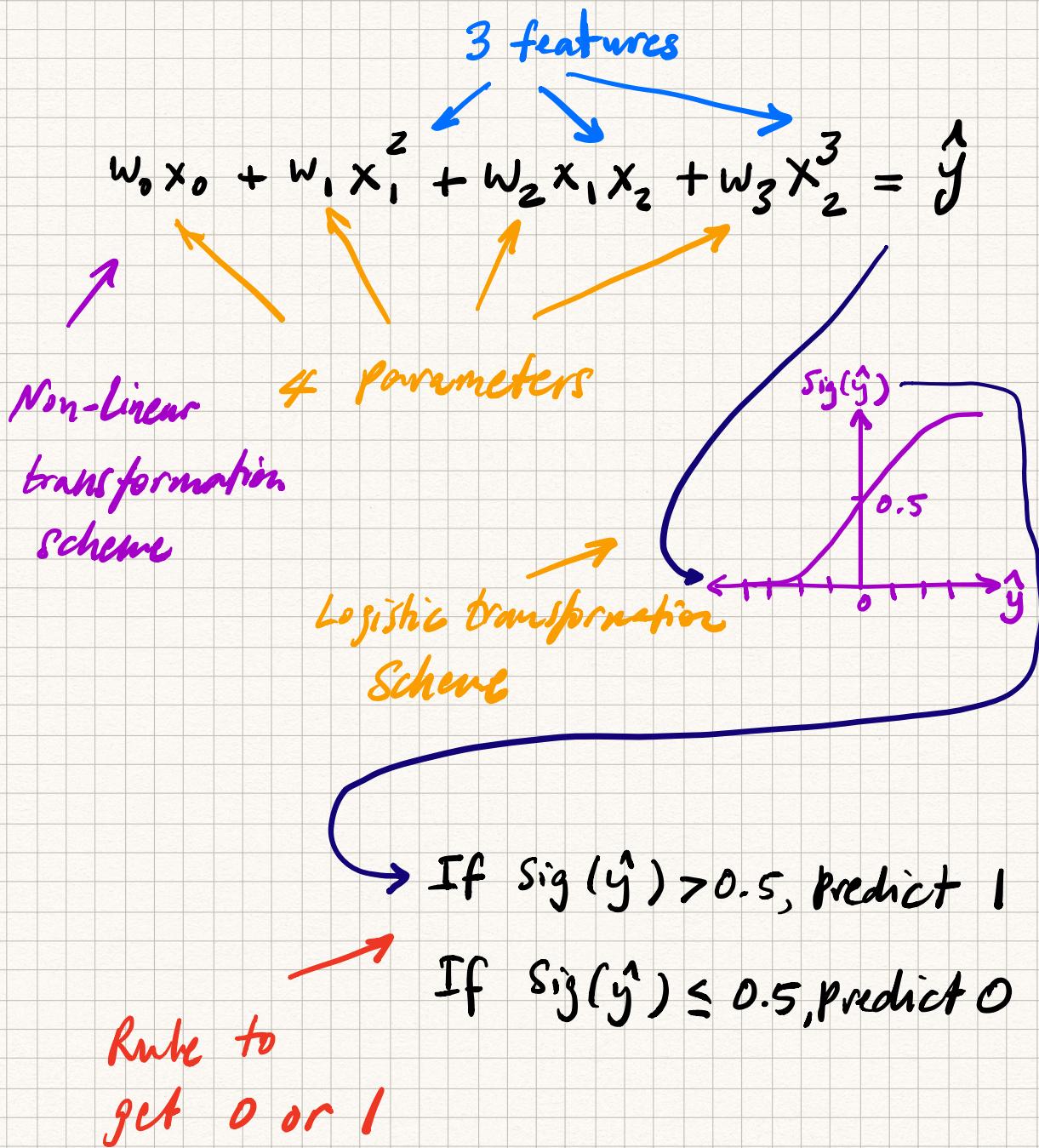
3 features  
4 parameters

"Non-Linear transformation scheme"

# Regression, Logistic, Linear



# Regression, Logistic, Non-Linear

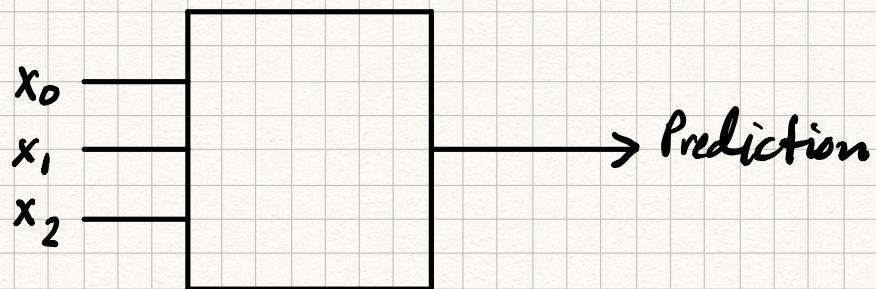


## EXERCISE

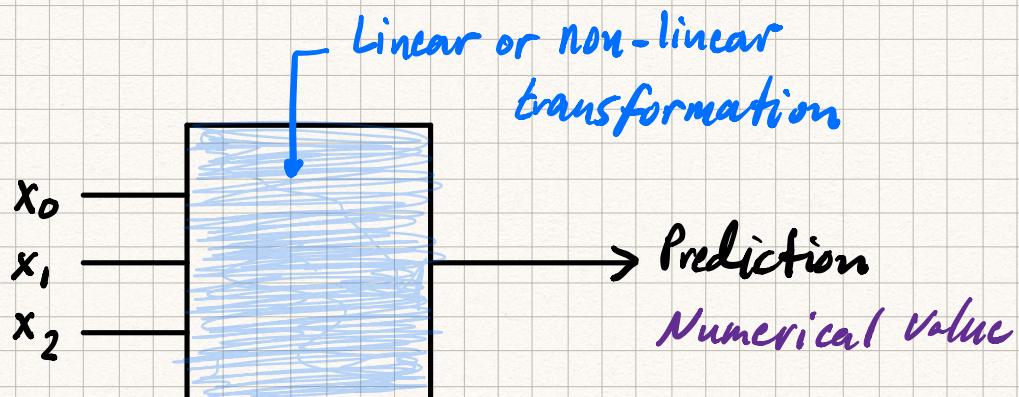
For the models above,  
draw a schematic data table  
for the original data and  
show how this data table  
is changed / transformed by  
each of the models above.

There are many more  
models available for  
Machine learning.

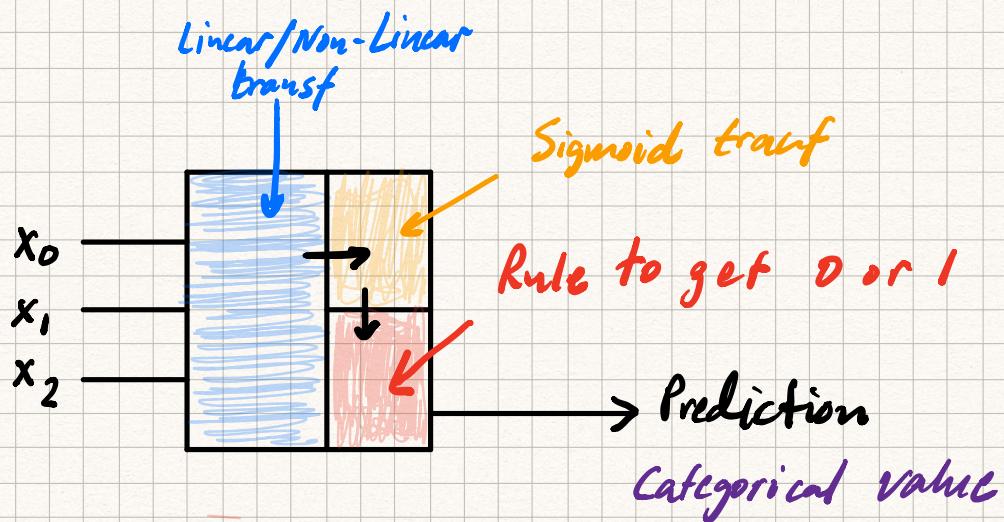
Let's use a simpler  
representation of a model  
and use this to introduce  
models we've not seen before.



## Regression, Linear/Non-Linear

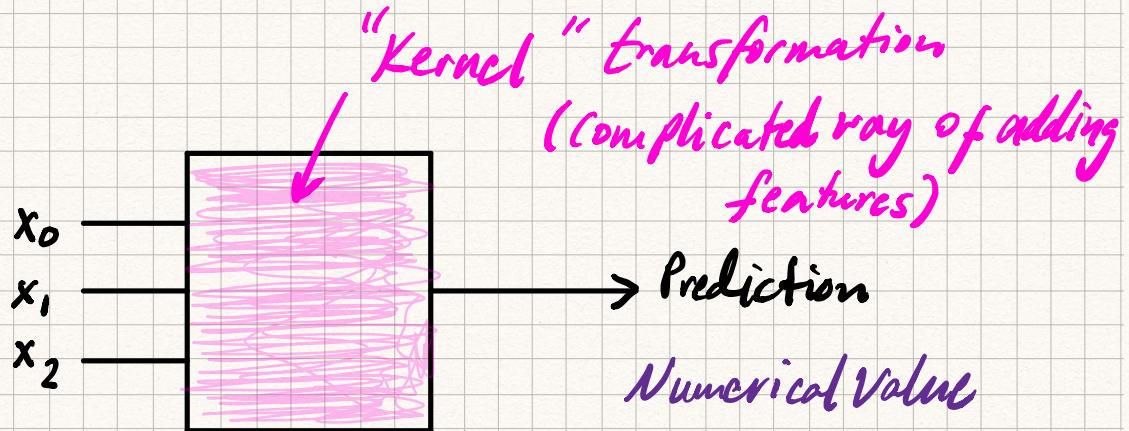


## Regression, Logistic, Linear/Non-Linear



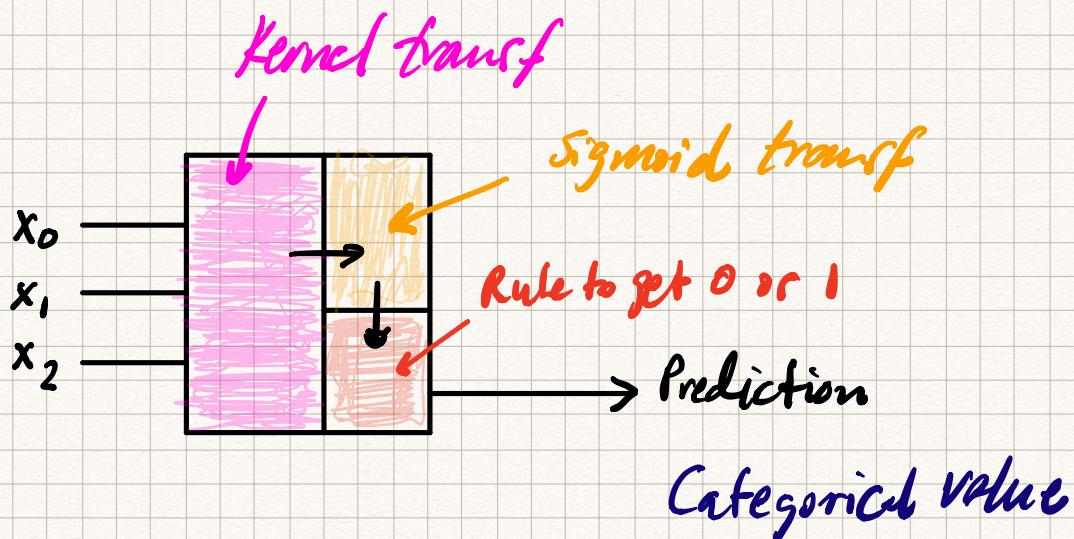
## Regression, Support Vector Machine, Linear/ Non-Linear

---

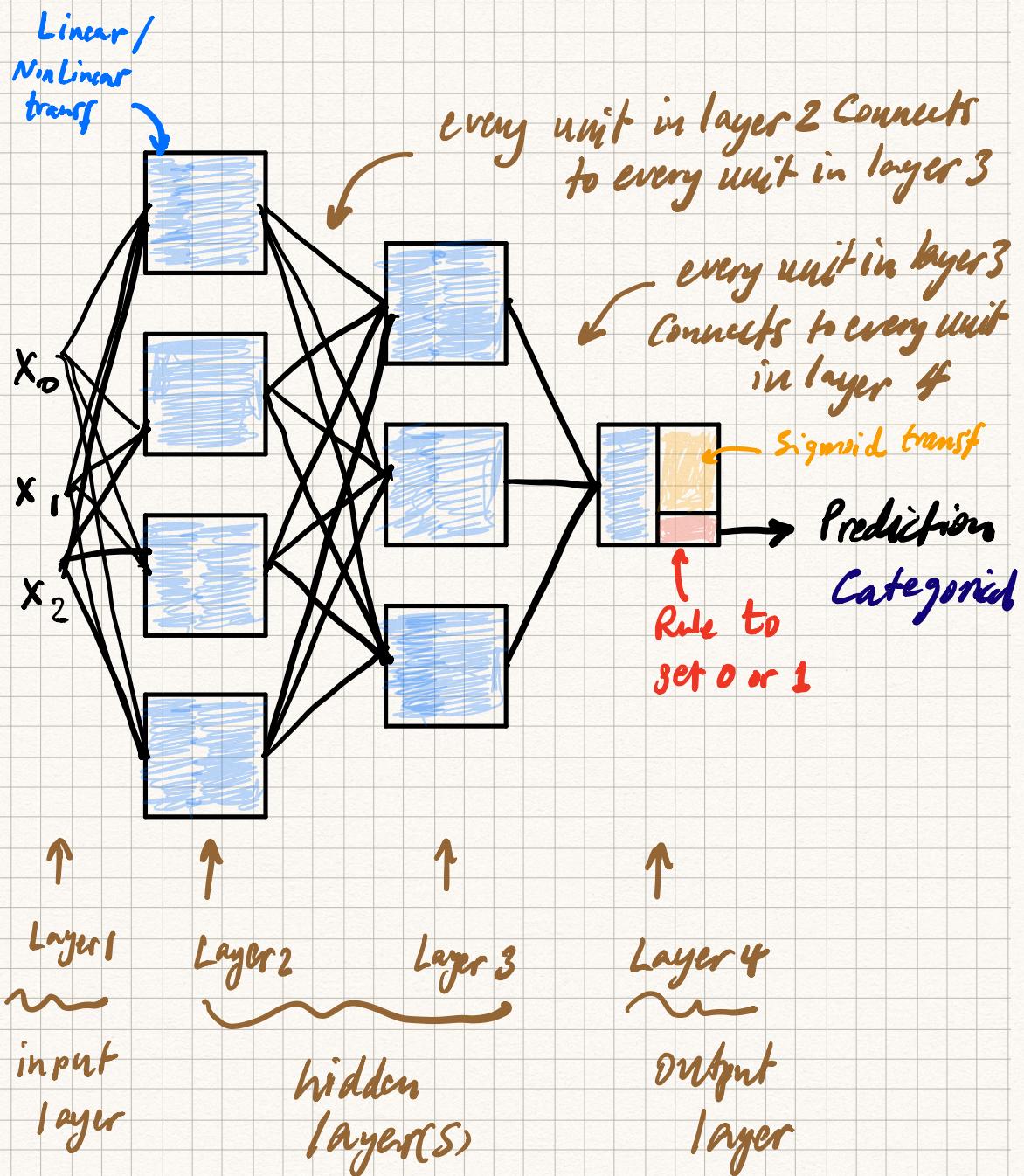


## Regression, Logistic, SVM, Linear/Non-Linear

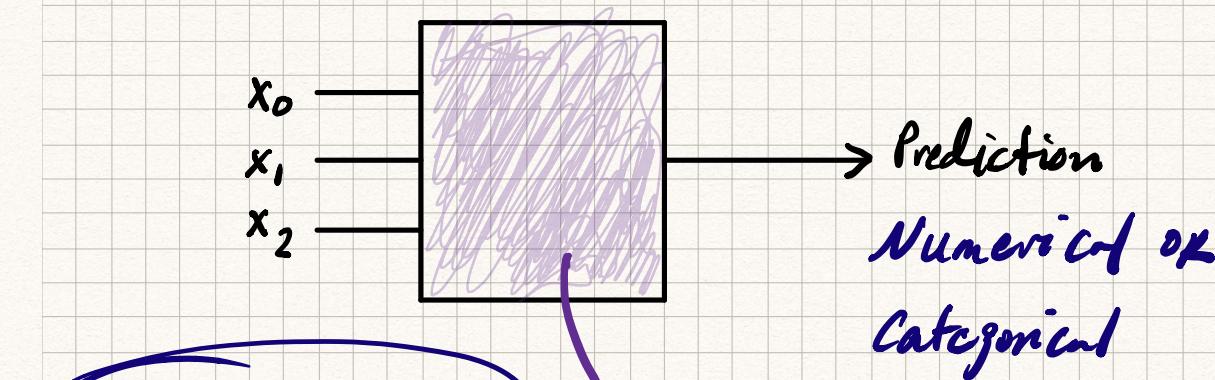
---



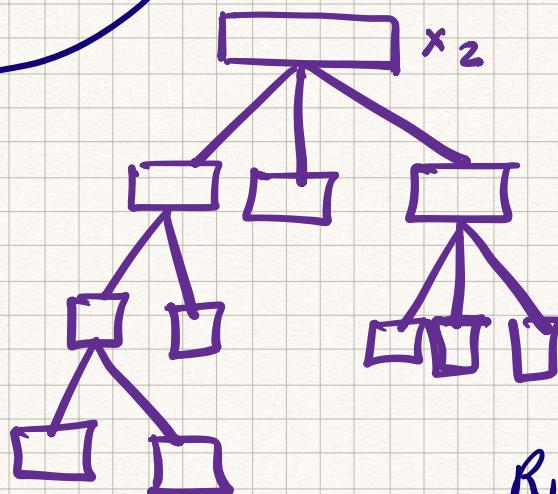
# Regression, Logistic, Neural Network, Linear/ Non-Linear



## Regression/Logistic regression, Decision Tree, Linear



No  $w_0, w_1, w_2$  parameters



## Random Forest

Bunch of trees, each using  
a random subset of the attributes

Rules for:  
• what to split on  
• when to stop

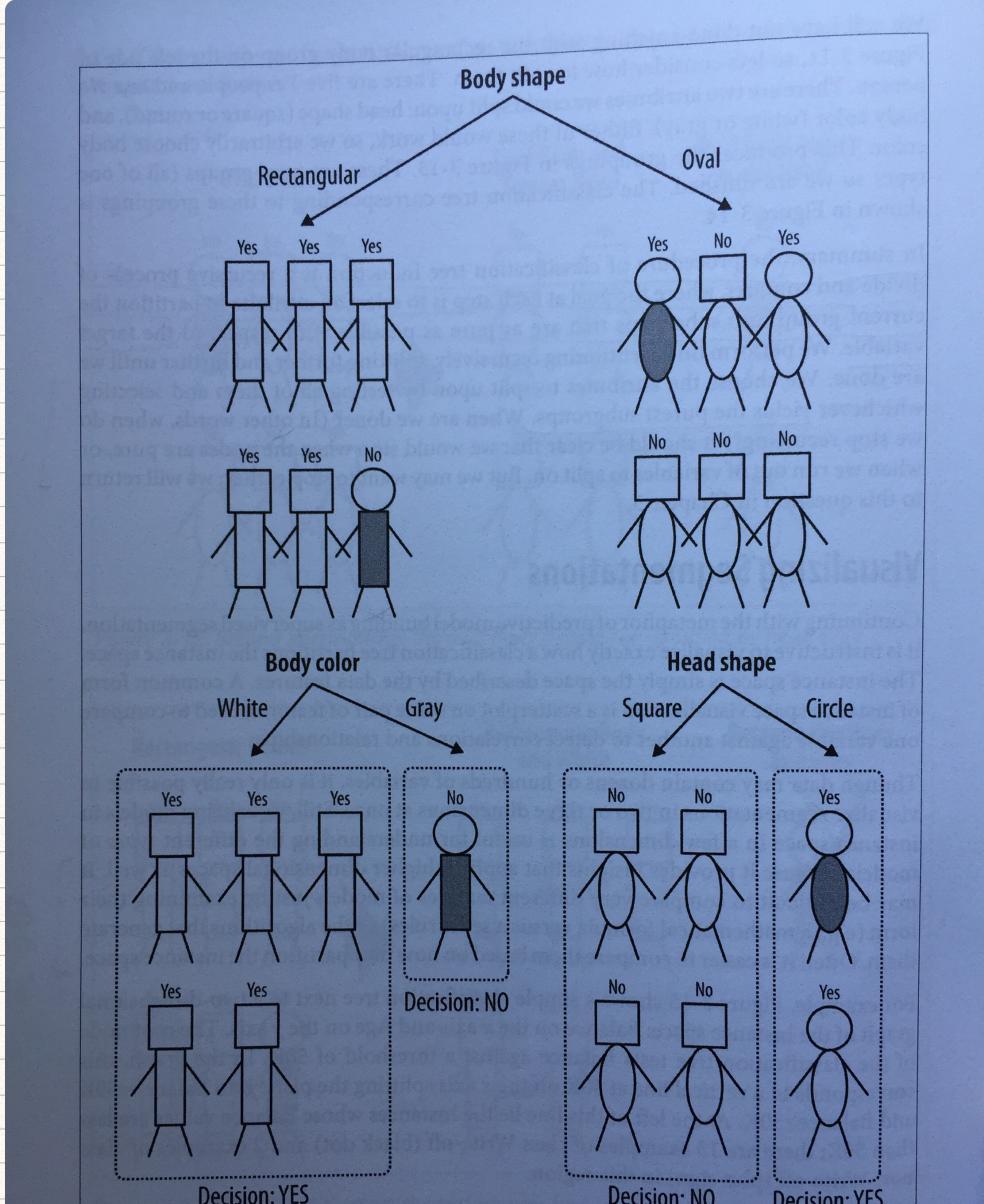
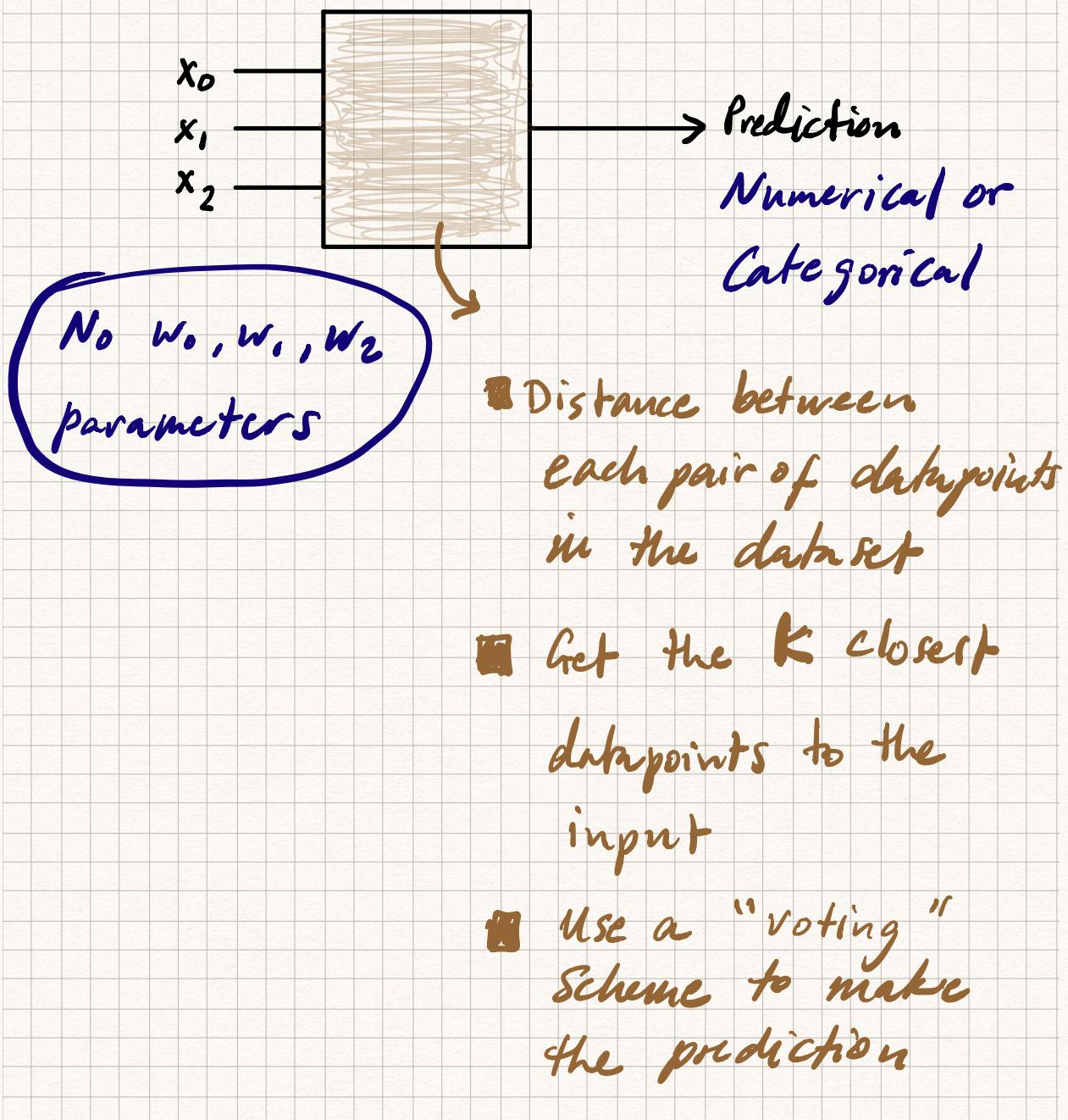


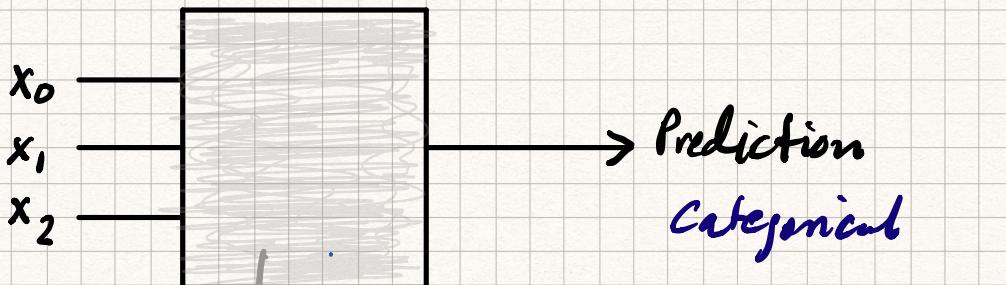
Figure 3-14. The classification tree resulting from the splits done in Figure 3-11 to Figure 3-13.

# Regression / Logistic Regression, K-NN, Linear



## Regression, Logistic, Naive Bayes, Linear

---



No parameters  
 $w_0, w_1, w_2$

What is  $\hat{y}$  if  $x_0 = 1$   
 $x_1 = 2.2$   
 $x_2 = -4.8$  ?

- Use Bayes' Rule to make the prediction

(we'll learn about Bayes' Rule later when we build an email spam filter)

# SUMMARY

	Linear	Non-lin	Numerical	Categorical
Regression	✓	✓	✓	✗
Logistic Regression	✓	✓	✗	✓
SVM	✓	✓	✓	✓
Neural Network	✗	✓	✗	✓
Decision Tree	✓	✗	✓	✓
Random Forest	✓	✗	✓	✓
Ada Boost	✓	✗	✓	✓
K-NN	✓	✗	✓	✓
Naive Bayes	✓	✗	✗	✓

Q: How do you choose which model  
to use ?

A: Try multiple models and  
then choose the one that  
performs the best.