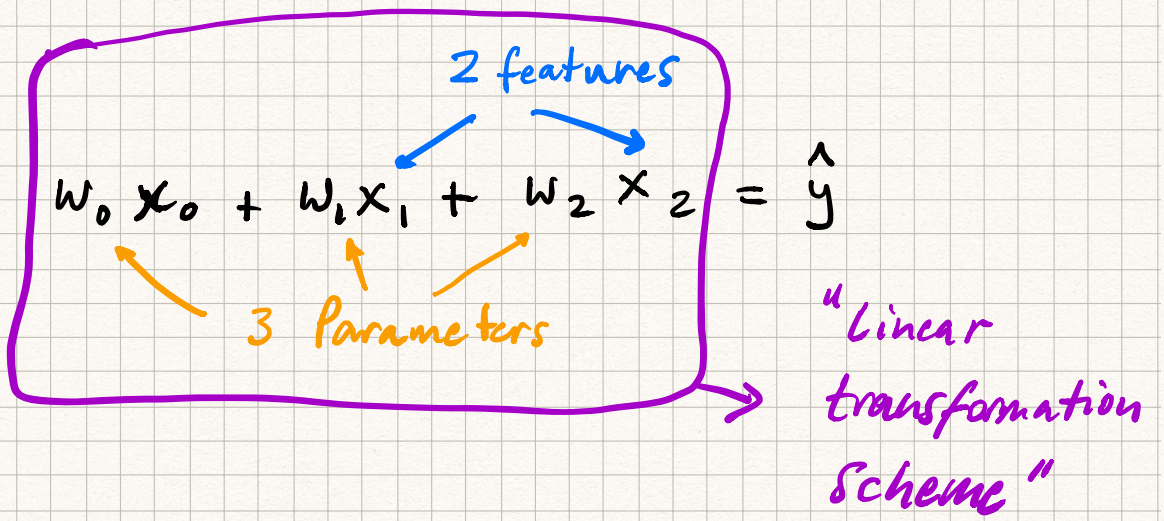
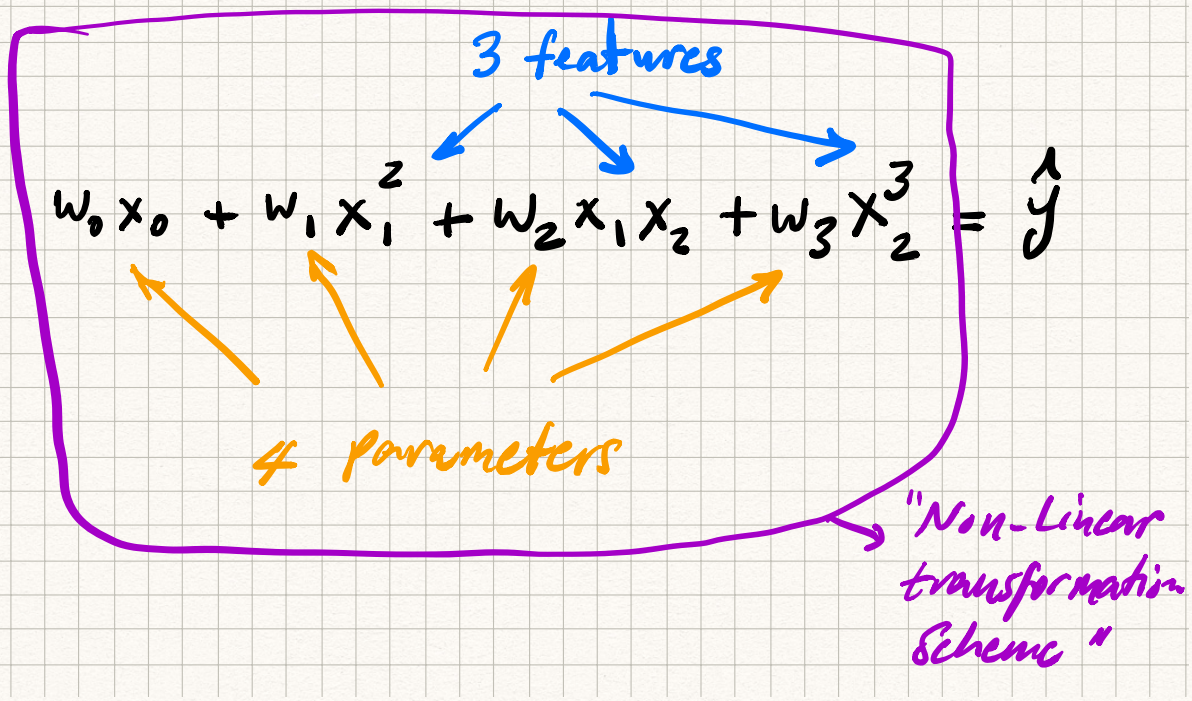
MODELS FOR MACHINE LEARNING

Models we've Constructed

Regression, Linear



Regression, Non-Linear



Regression, Logistic, Linear

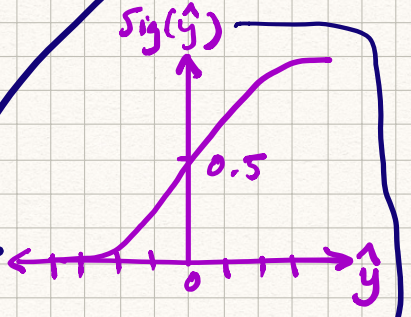
2 features

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

3 Parameters

Linear transformation scheme

Logistic transformation Scheme



If $Sig(\hat{y}) > 0.5$, predict 1

If $Sig(\hat{y}) \leq 0.5$, predict 0

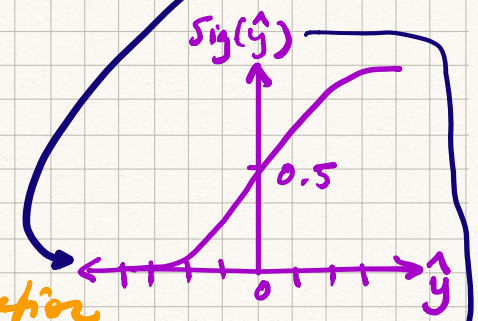
Regression, Logistic, Non-Linear

3 features

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

4 parameters

Non-Linear transformation scheme



Logistic transformation scheme

If $sig(\hat{y}) > 0.5$, predict 1

If $sig(\hat{y}) \leq 0.5$, predict 0

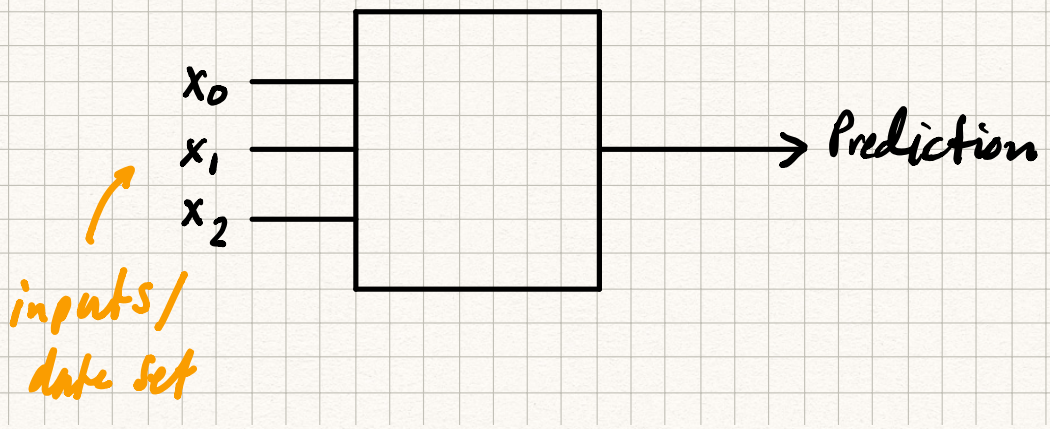
Rule to get 0 or 1

EXERCISE

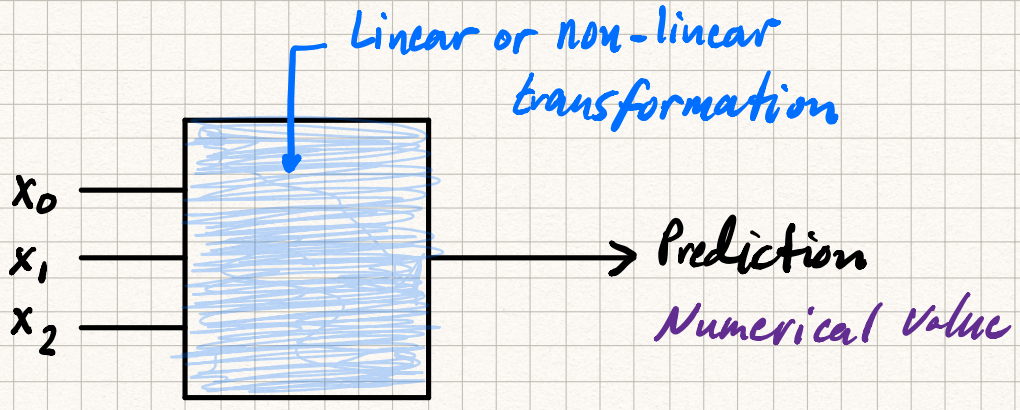
For the models above, draw a schematic data table for the original data and show how this data table is changed / transformed by each of the models above.

There are many more models available for machine learning.

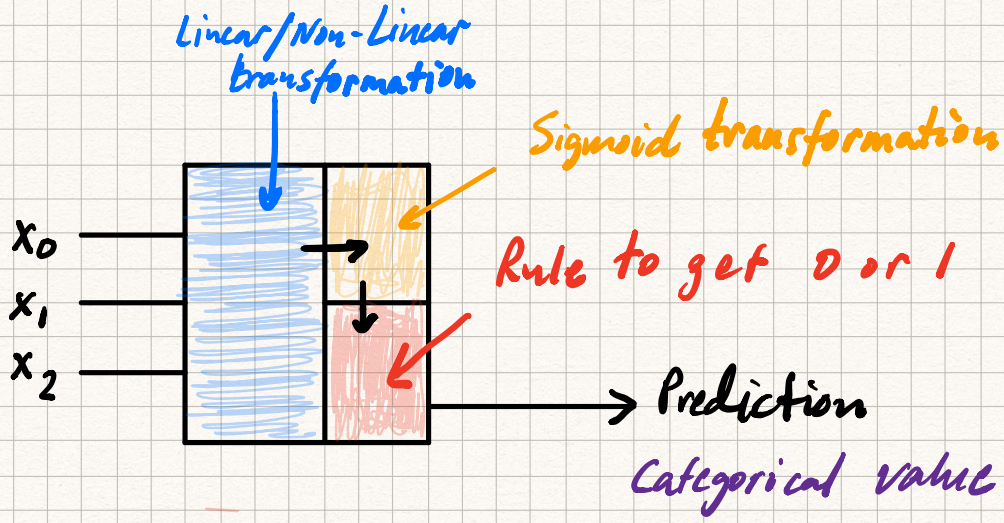
Let's use a simpler representation of a model and use this to introduce models we've not seen before.



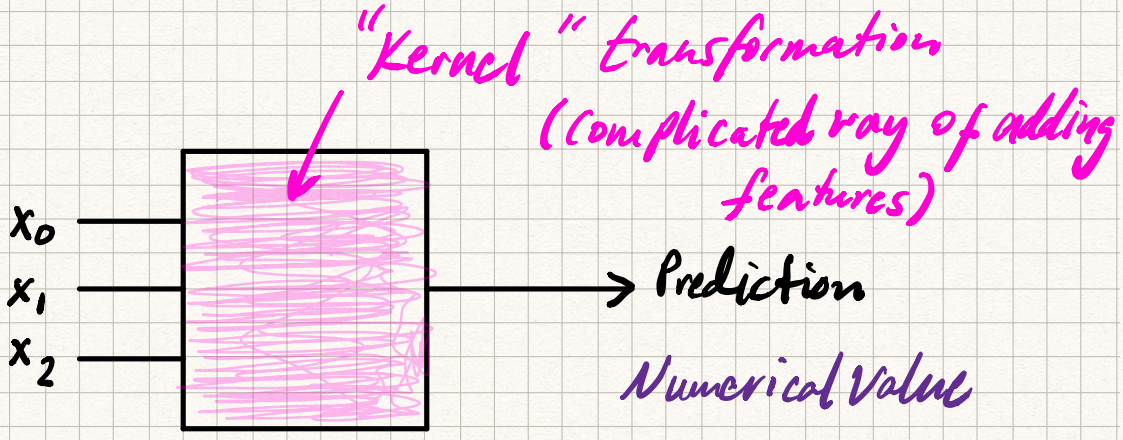
Regression, Linear/Non-Linear



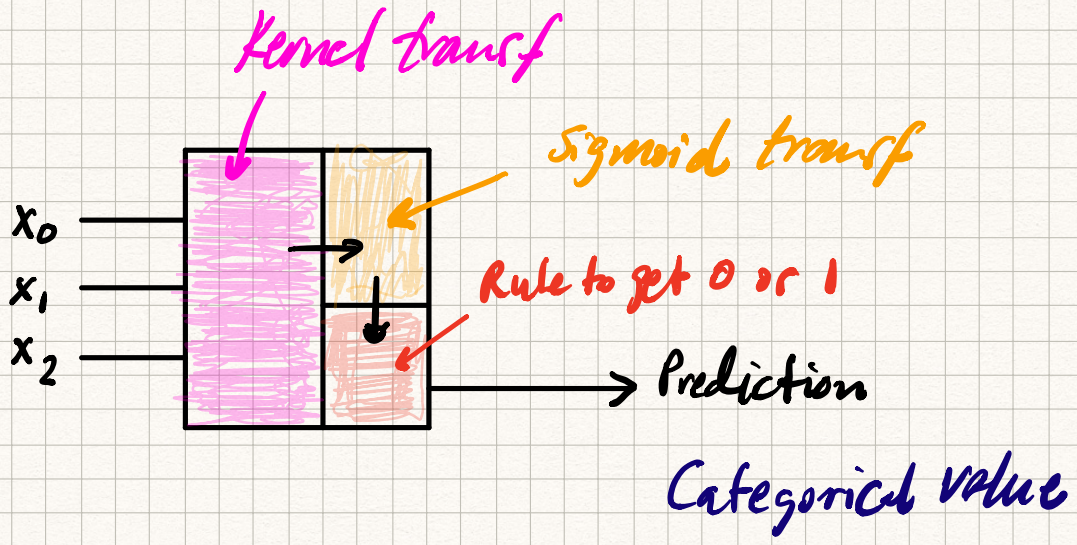
Regression, Logistic, Linear/Non-Linear



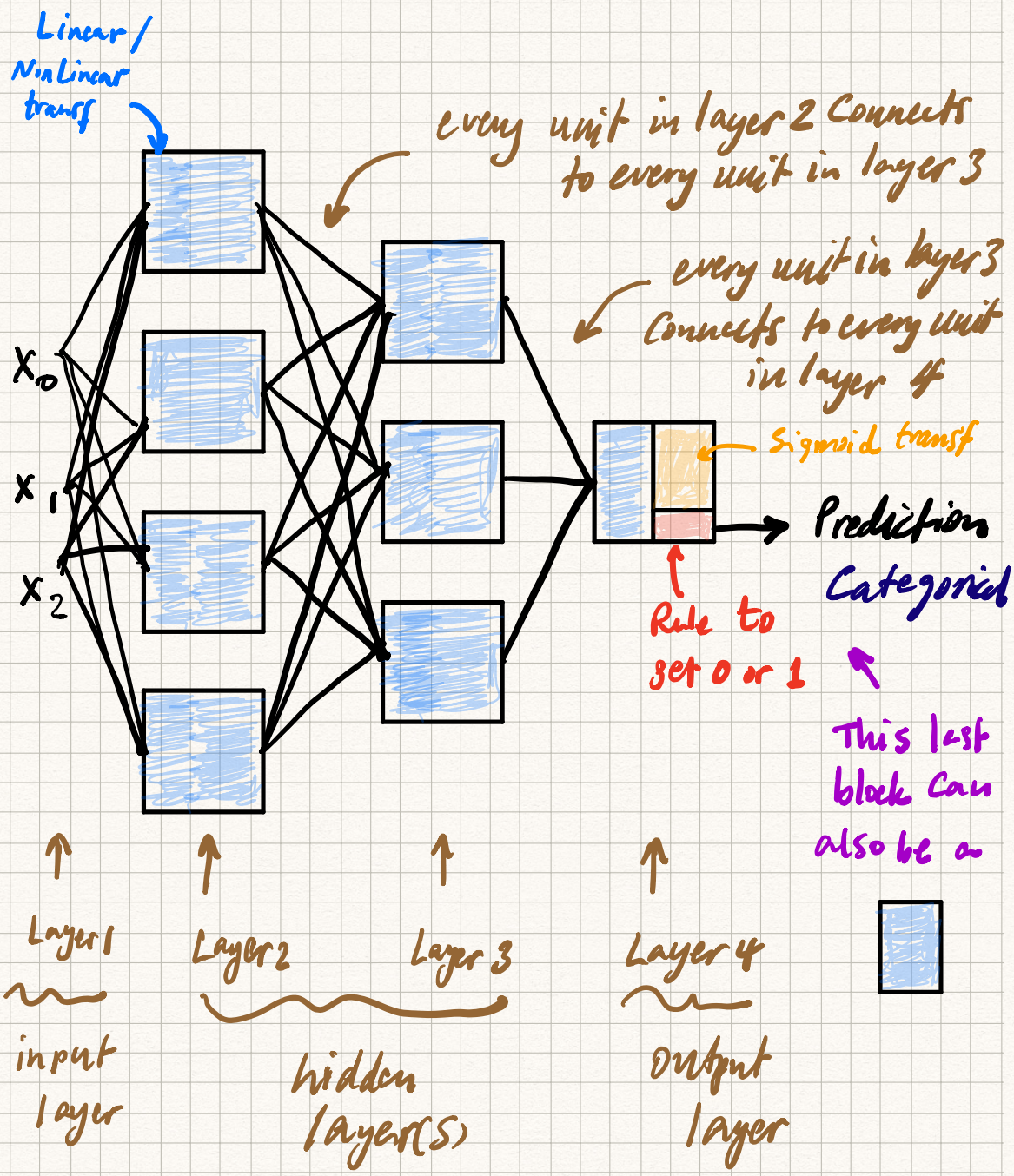
Regression, Support Vector Machine, Linear/Non-Linear



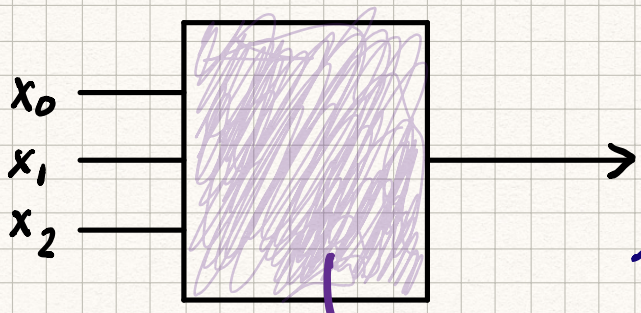
Regression, Logistic, SVM, Linear/Non-Linear



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

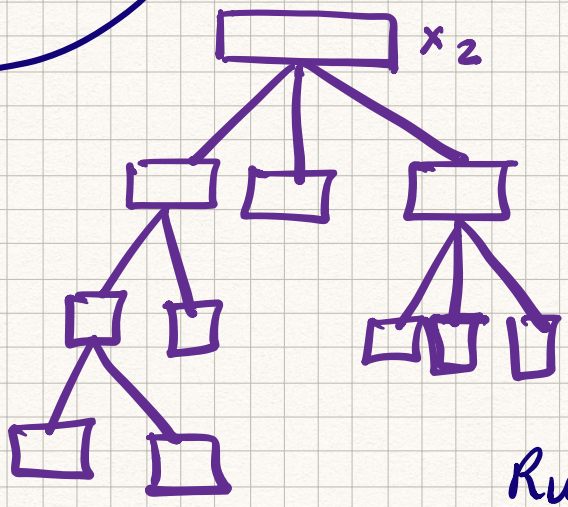


Regression / Logistic Regression, Decision Tree, Linear



Numerical or Categorical

No w_0, w_1, w_2 parameters



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

Random Forest

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

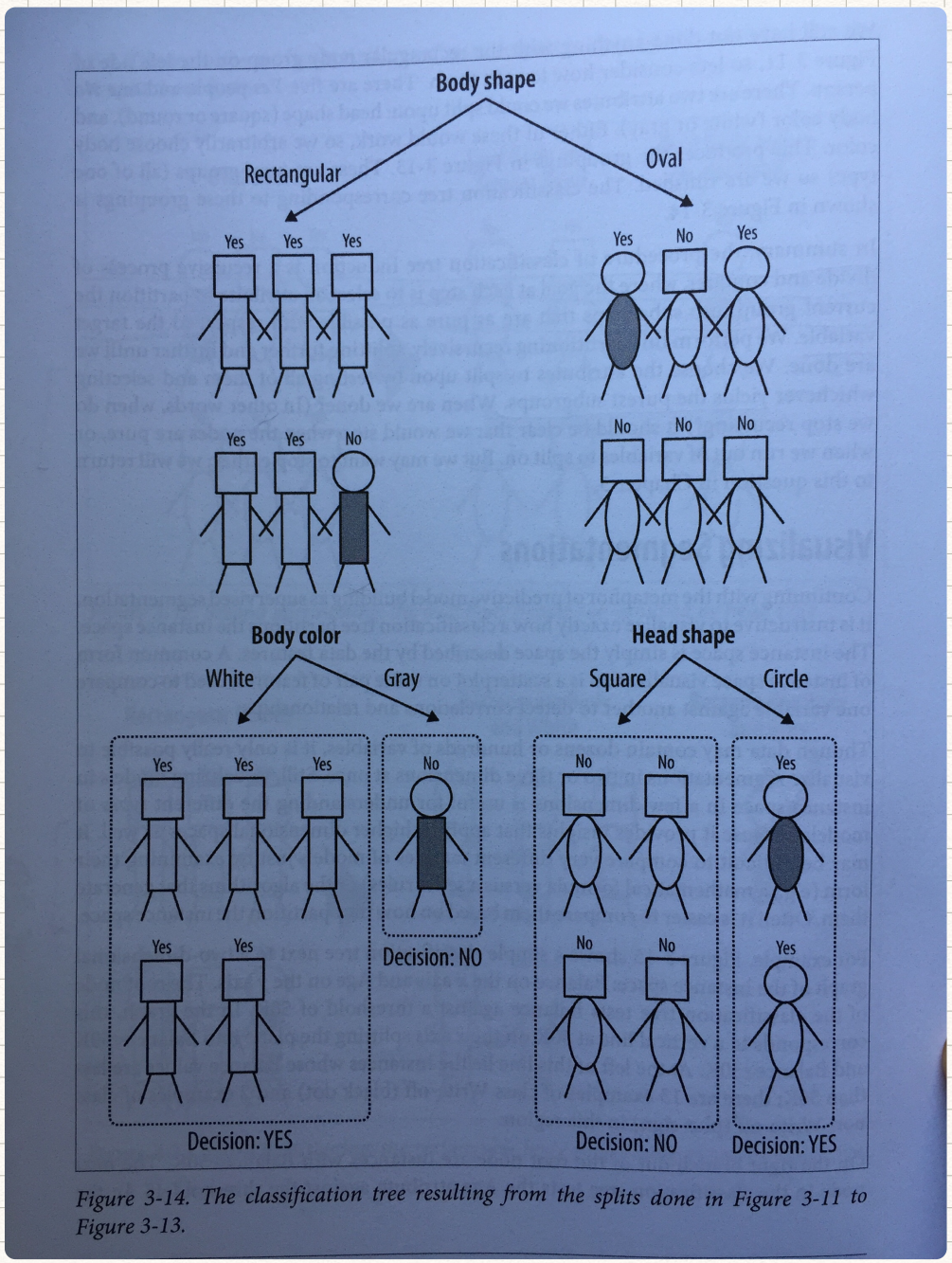
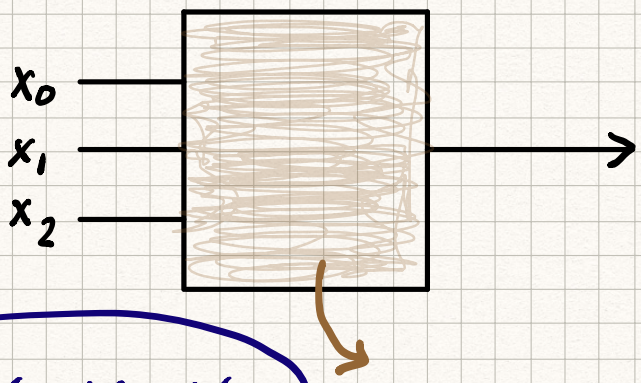


Figure 3-14. The classification tree resulting from the splits done in Figure 3-11 to Figure 3-13.

Data Science for Business p.68

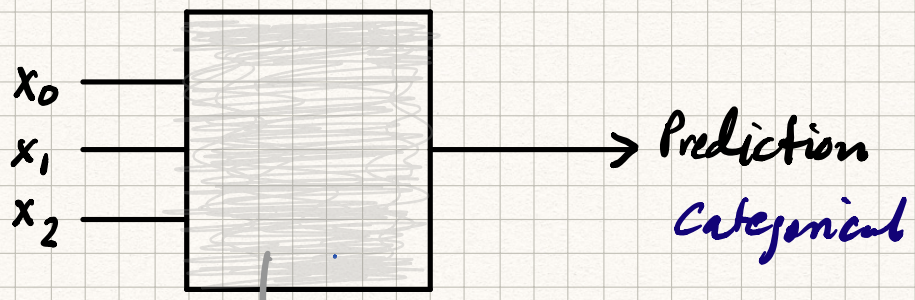
Regression / Logistic Regression, K-NN, Linear



No w_0, w_1, w_2 parameters

- Distance between each pair of datapoints in the dataset
- Get the K closest datapoints to the input
- Use a "voting" scheme to make the prediction

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	Categorial
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.

Q: Why do some models do better than others on the same data set?

A: This is a deep question about the nature of the universe!