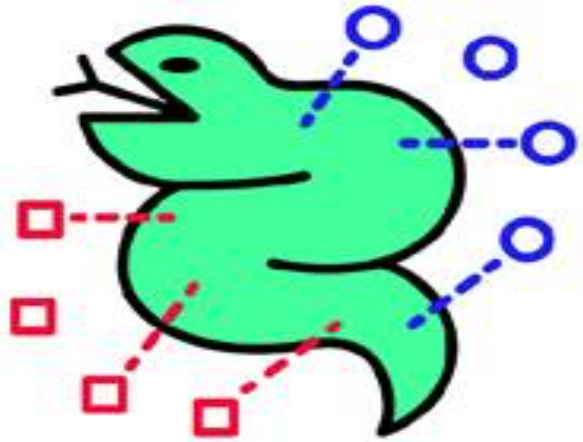
Decision Trees



Alexander Jung

Assistant Professor

Department of Computer Science

Aalto University

What are three main components of machine learning?

1. Data

Data

- set of “data points” (atomic unit of information)
- data point has features and labels
- features are properties that can measured easily
- labels =higher-level facts or quantities of interest

Data Point = ”Some Movie”

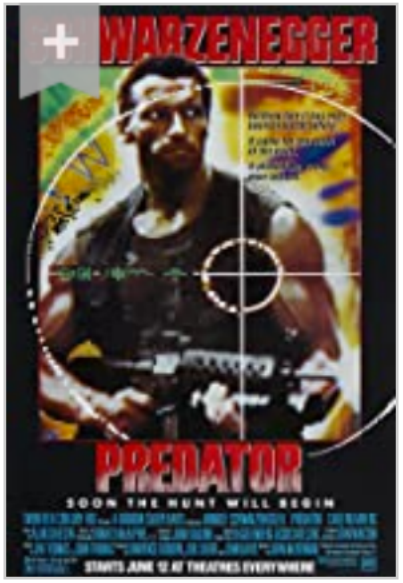
features:

- x1 = movie duration in minutes
- x2 = screen time of Arnold Schwarzenegger

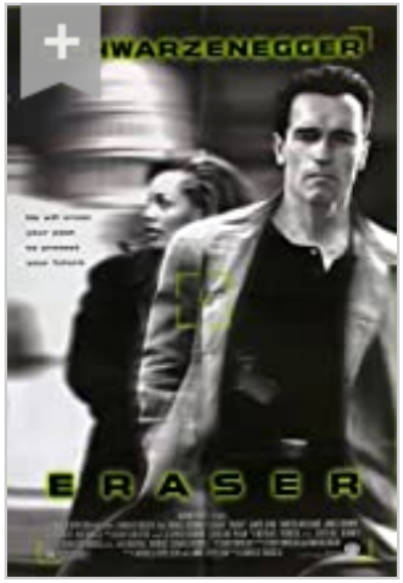
label: the movie rating (scale 0 … 10)

Data = Bunch of Data Points

The 6th Day
Adam Gibson
(2000)



Predator
Dutch
(1987)

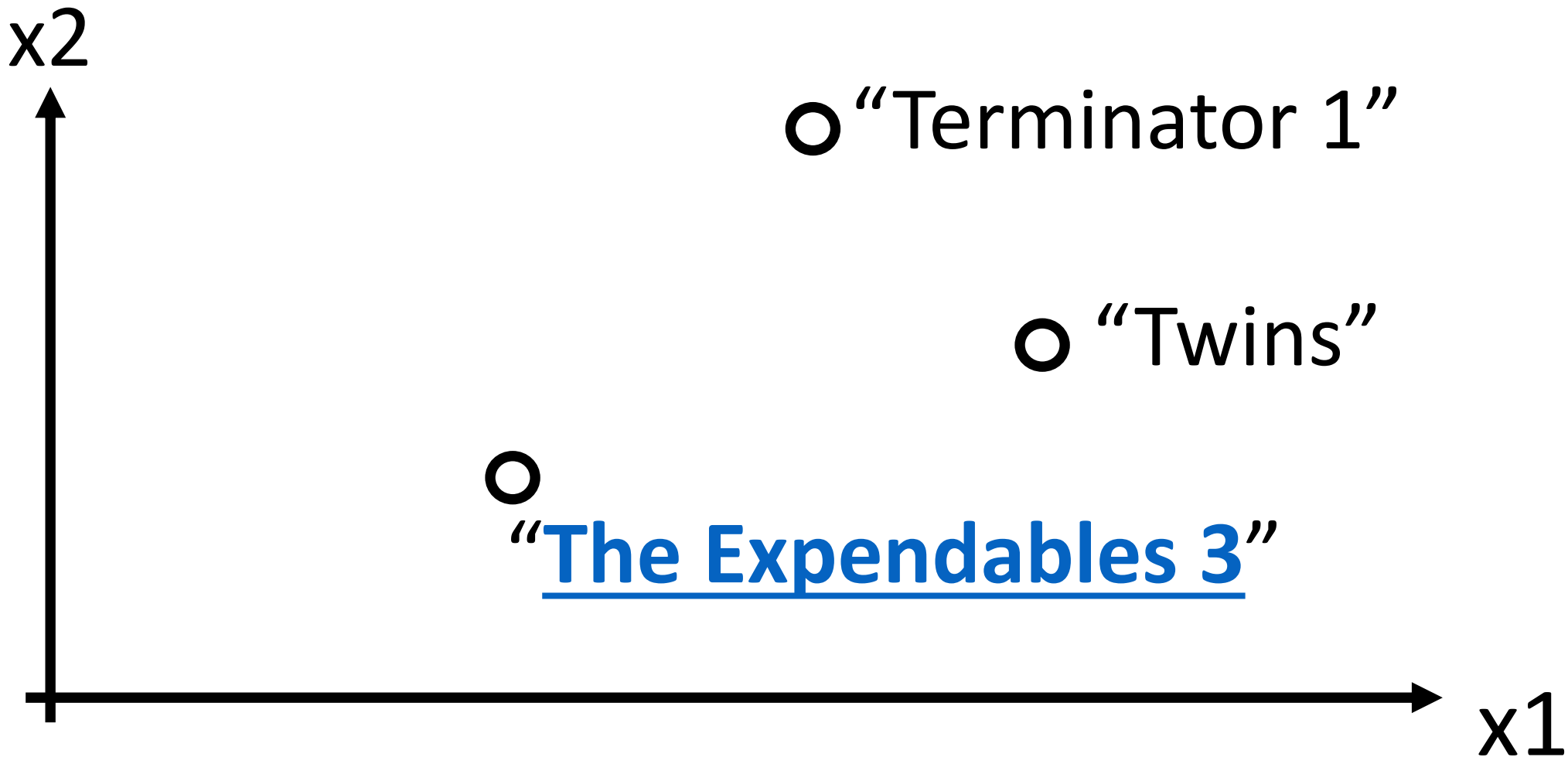


Eraser
U.S. Marshal John 'The E…
(1996)

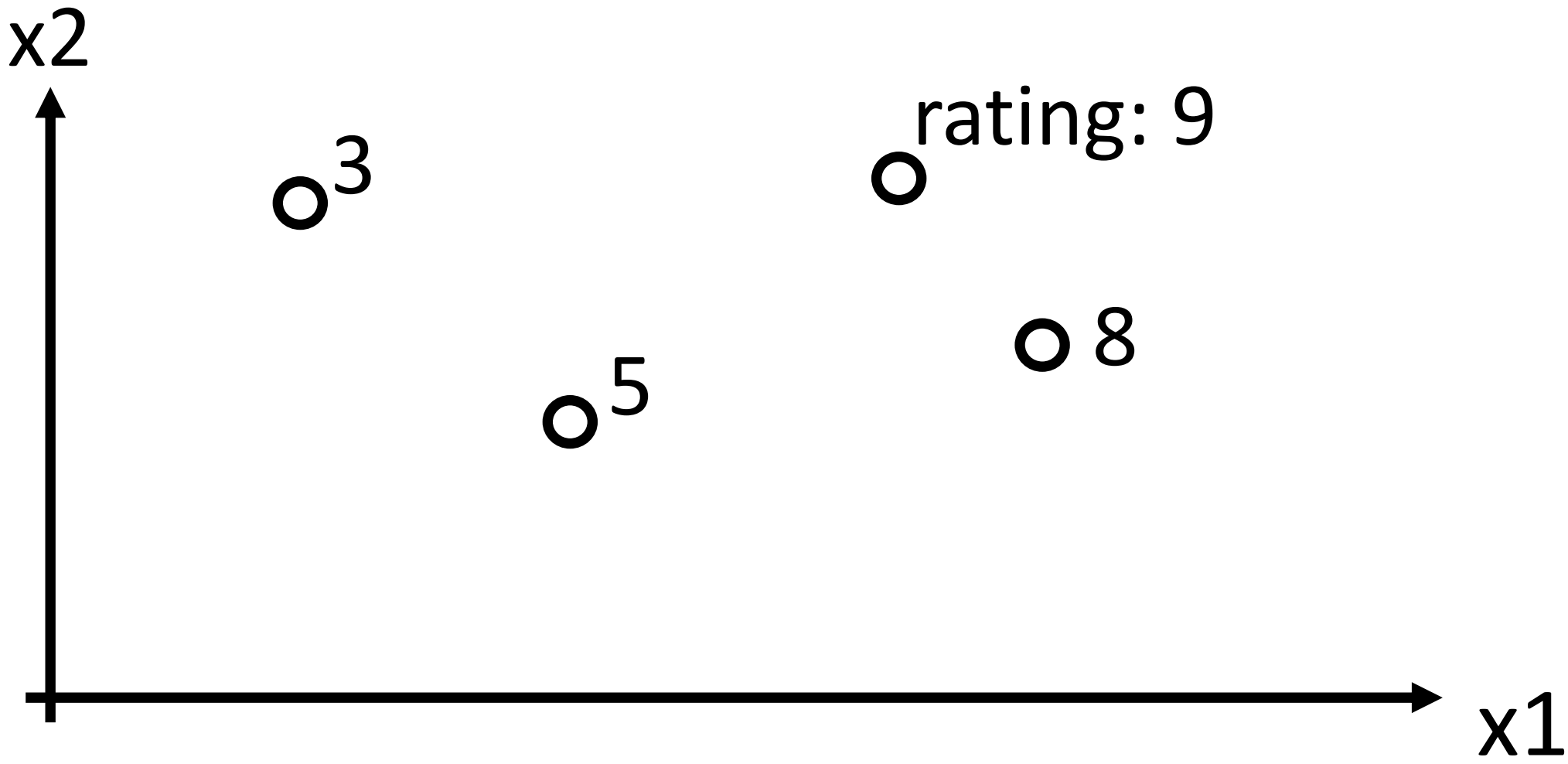
Last Action Hero
Jack Slater
(1993)

IMDb

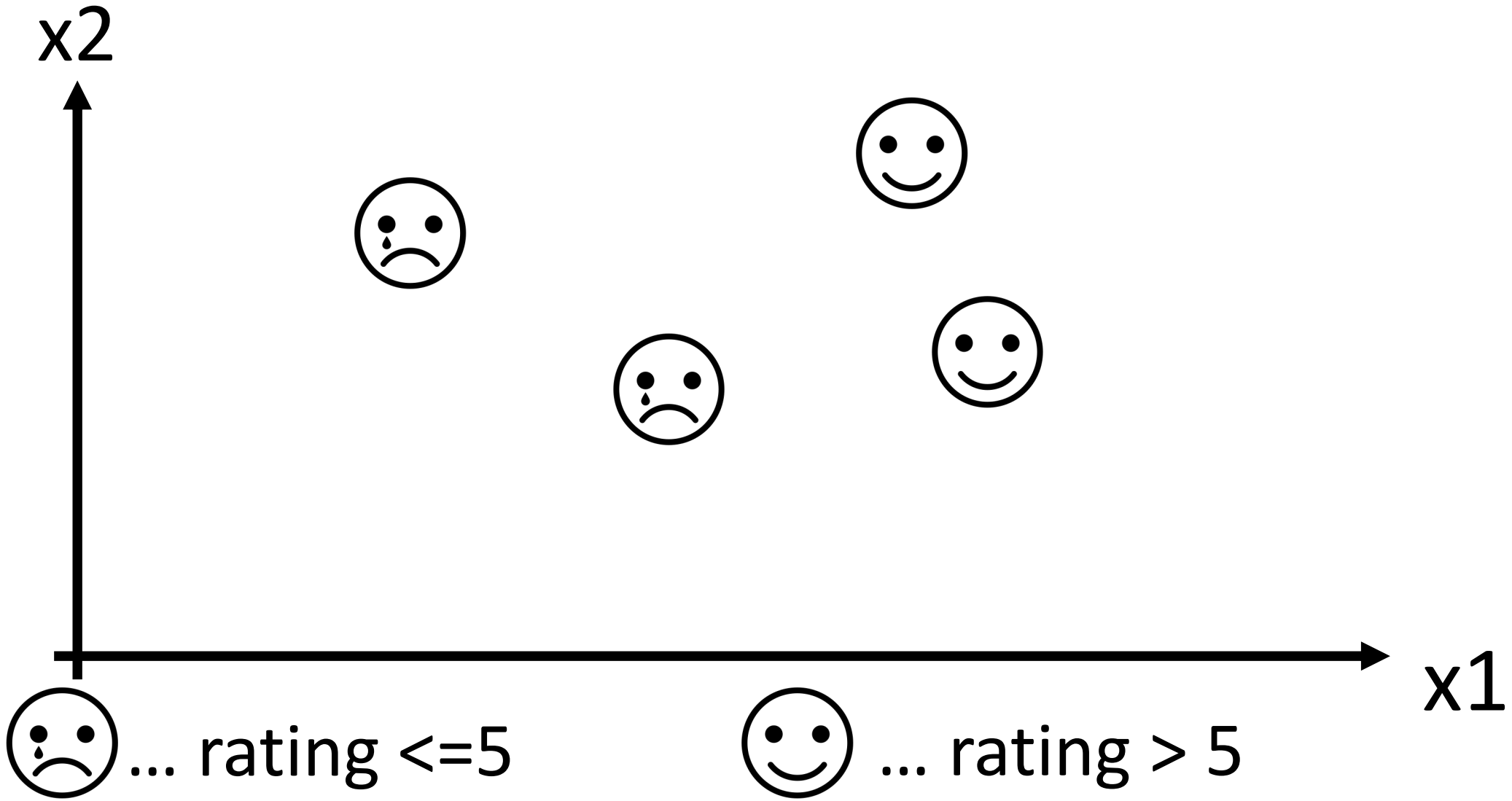
Scatter Plot



Scatter Plot

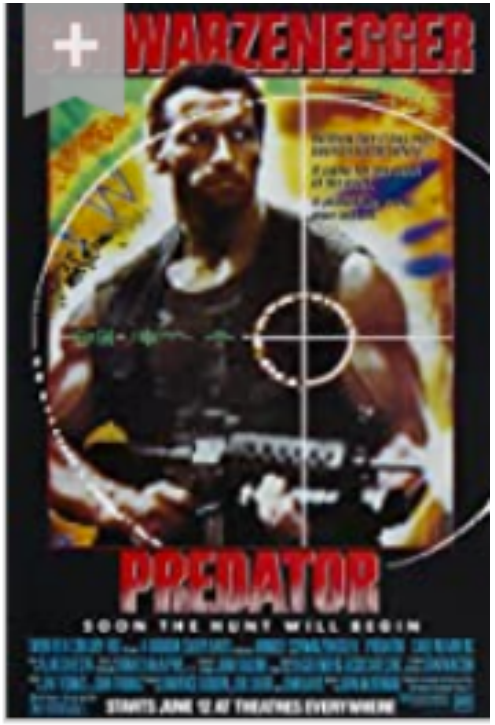


Different Choice of Label



2. Hypothesis Space

How Many Predictors Are There?



predictor

predicted rating

h(x1,x2)

$\hat{y}$ =8.3

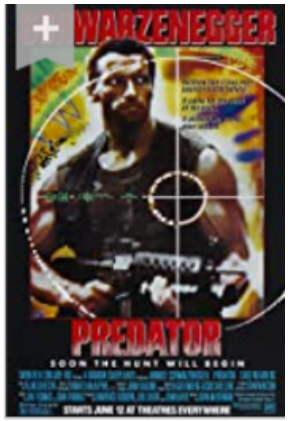
x1 = 123.348 ,

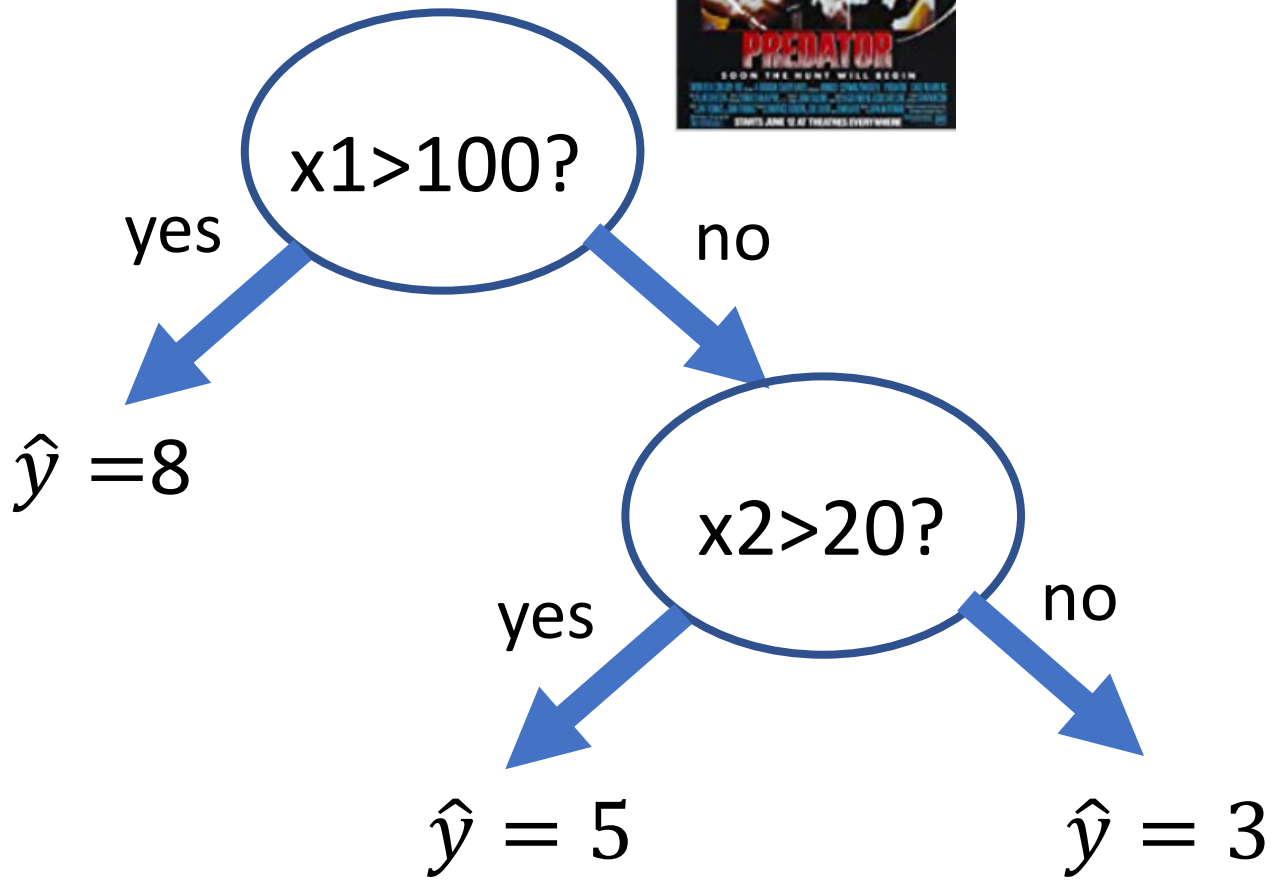
x2 = 40.456

Hypothesis Space “Decision Tree”

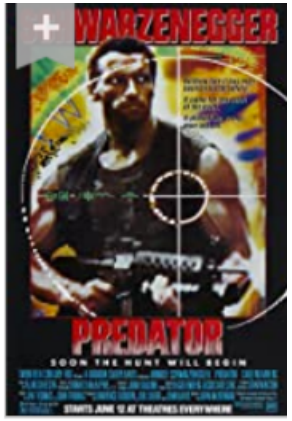
- infinitely many functions from x1,x2 to $\hat{y}$
- restrict to subset of maps (hypothesis space)
- subset of maps given by decision trees

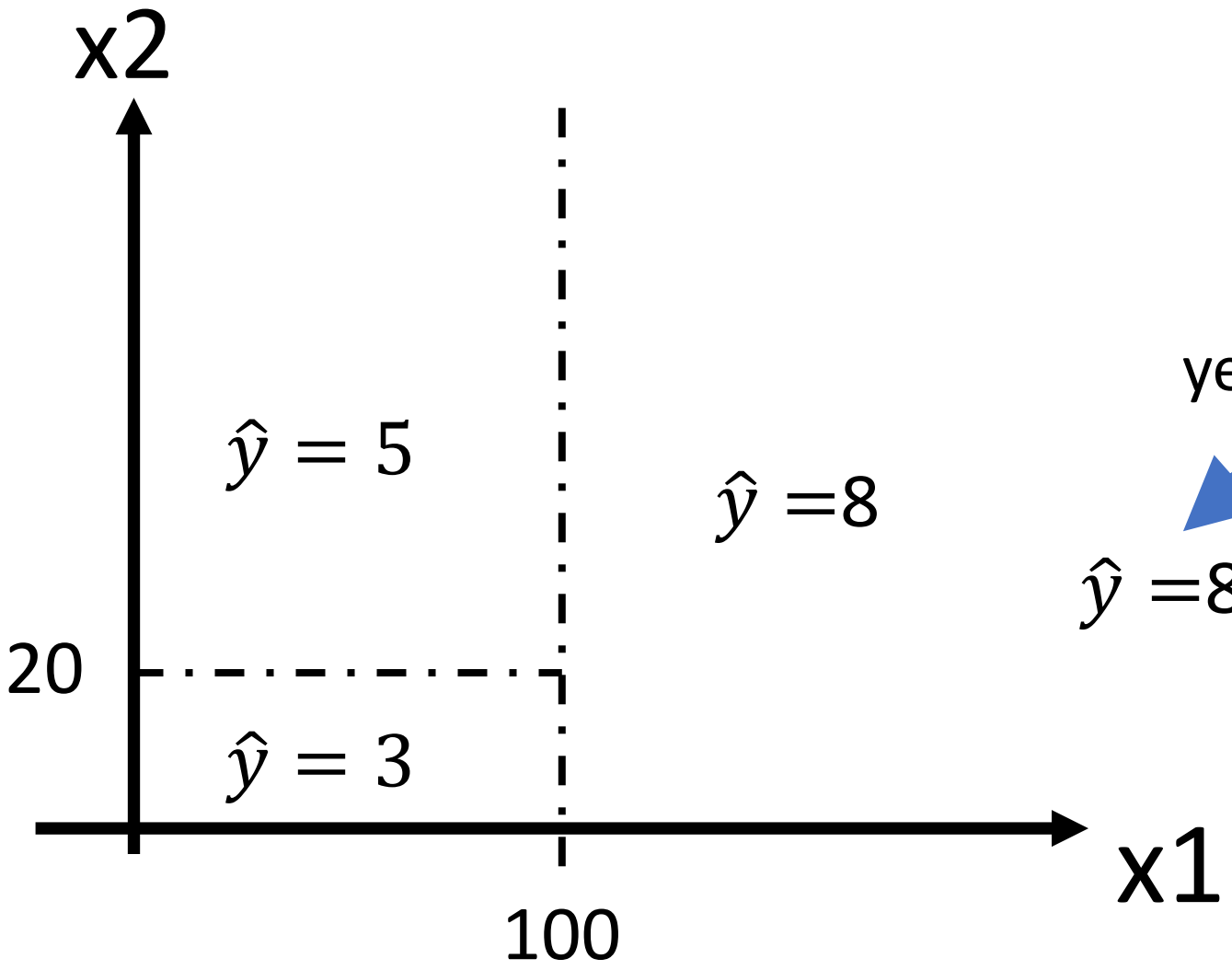
A Decision Tree (DT)

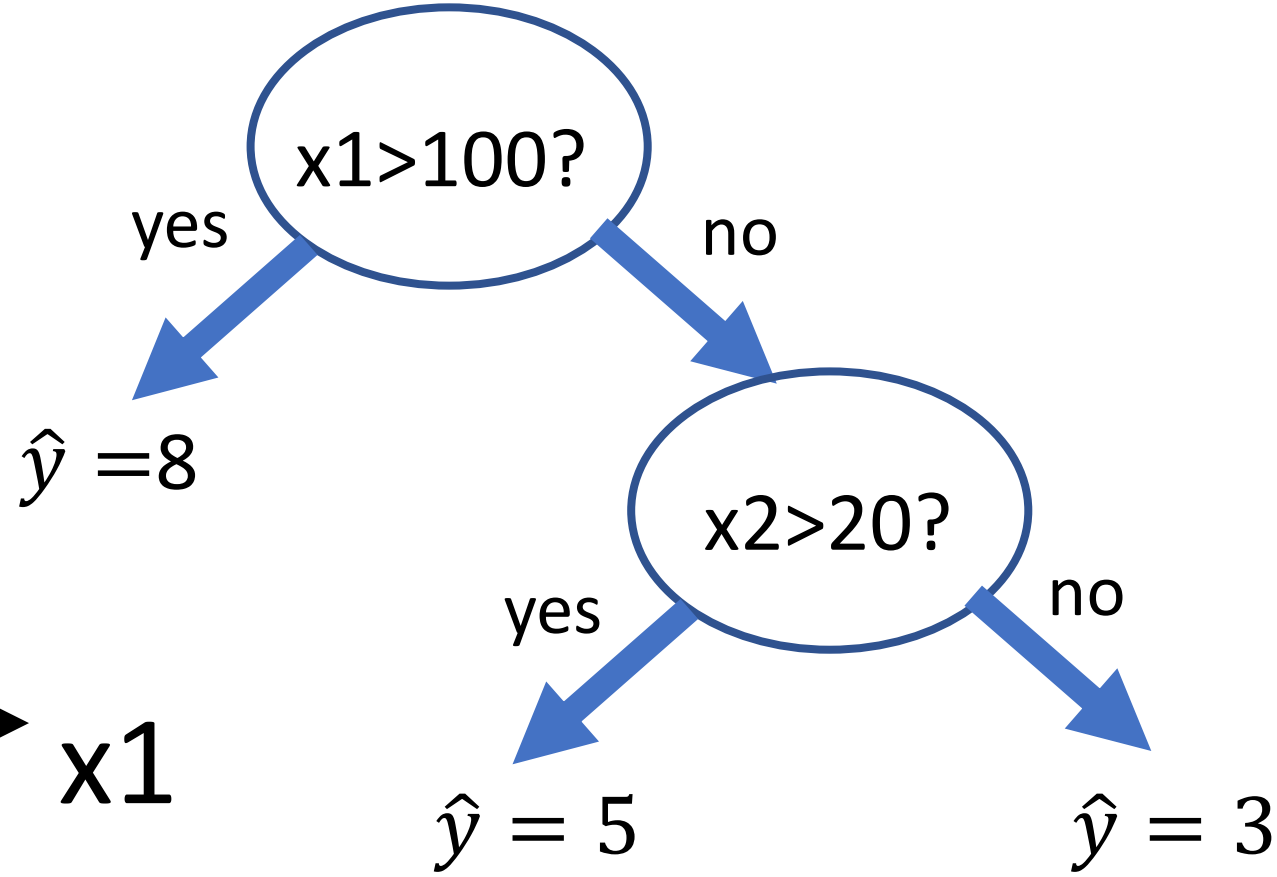




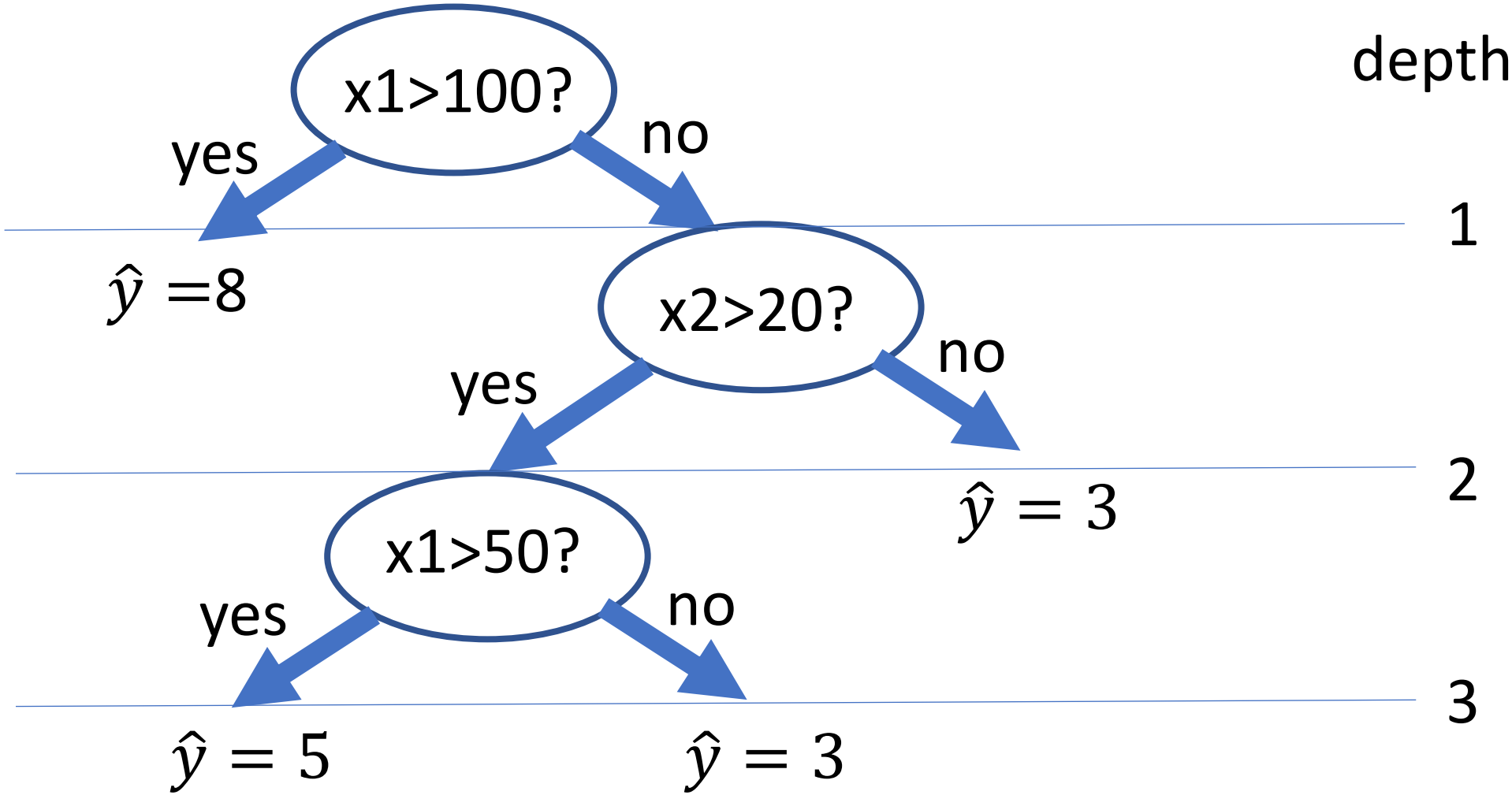
DT in Feature Space







Tree Depth


x1>100?
yes
no
$\hat{y}$ =8
x2>20?
yes
no
$\hat{y}$ = 3
x1>50?
yes
no
$\hat{y}$ = 5
$\hat{y}$ = 3
depth
1
2
3

Hypothesis Space of DTs

space of predictor maps given by DT that involve threshold tests for x1 and x2

sklearn.tree.DecisionTreeRegressor

class sklearn.tree.**DecisionTreeRegressor**(*, *criterion='mse'*, *splitter='best'*, *max_depth=None*, *min_samples_split=2*, *min_samples_leaf=1*, *min_weight_fraction_leaf=0.0*, *max_features=None*, *random_state=None*, *max_leaf_nodes=None*, *min_impurity_decrease=0.0*, *min_impurity_split=None*, *presort='deprecated'*, *ccp_alpha=0.0*) [source]

A decision tree regressor.

numeric labels

sklearn.tree.DecisionTreeClassifier

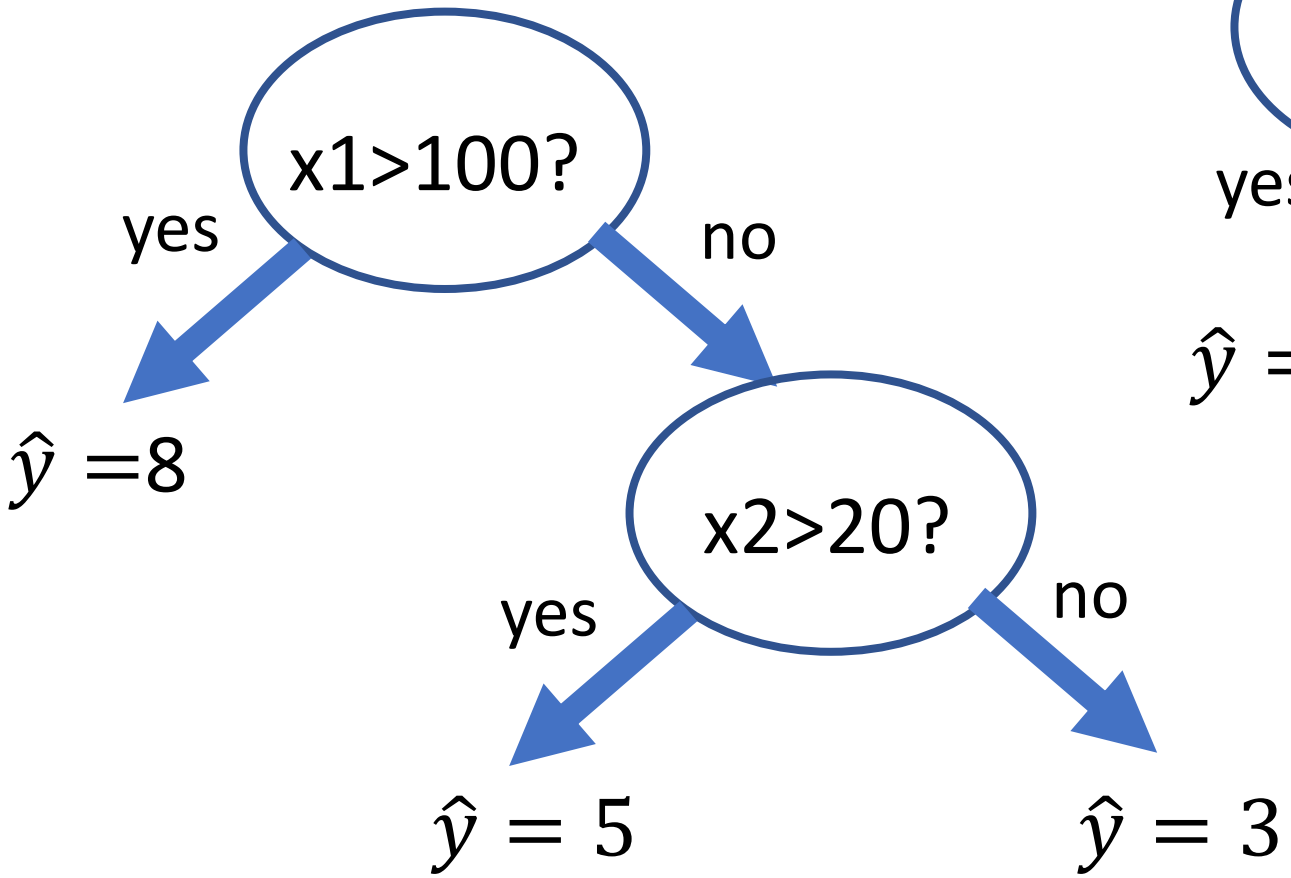
class sklearn.tree.**DecisionTreeClassifier**(*, *criterion='gini'*, *splitter='best'*, *max_depth=None*, *min_samples_split=2*, *min_samples_leaf=1*, *min_weight_fraction_leaf=0.0*, *max_features=None*, *random_state=None*, *max_leaf_nodes=None*, *min_impurity_decrease=0.0*, *min_impurity_split=None*, *class_weight=None*, *presort='deprecated'*, *ccp_alpha=0.0*) [source]

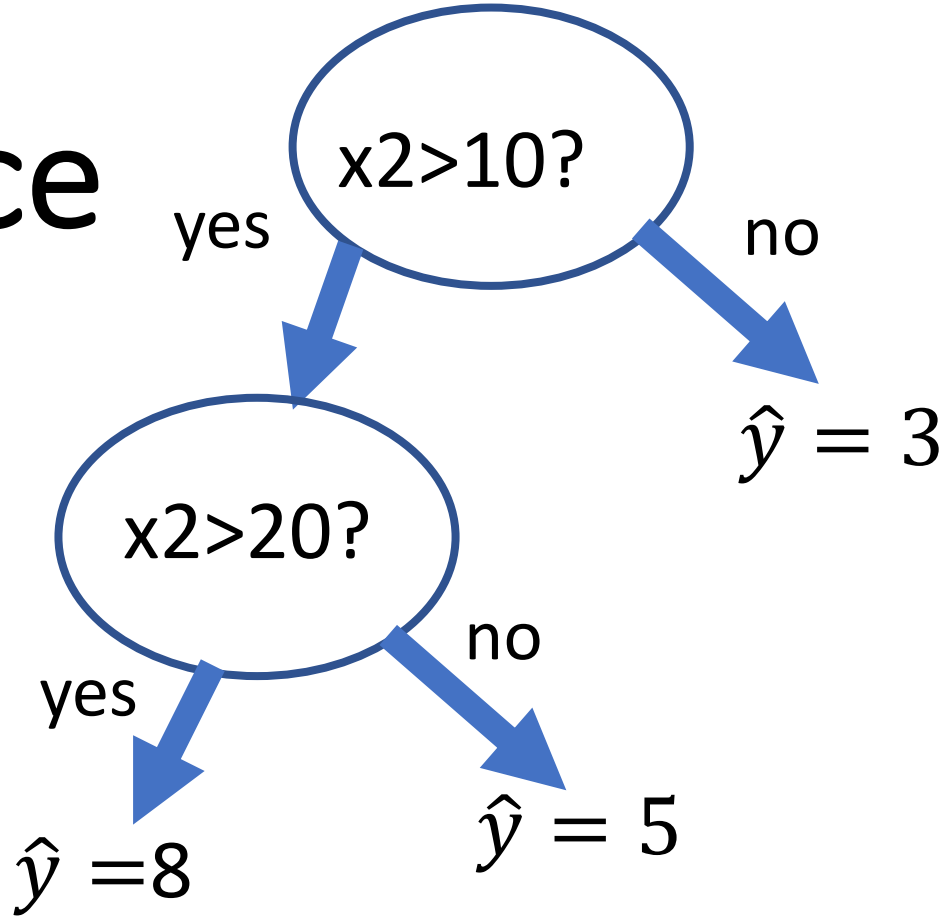
A decision tree classifier.

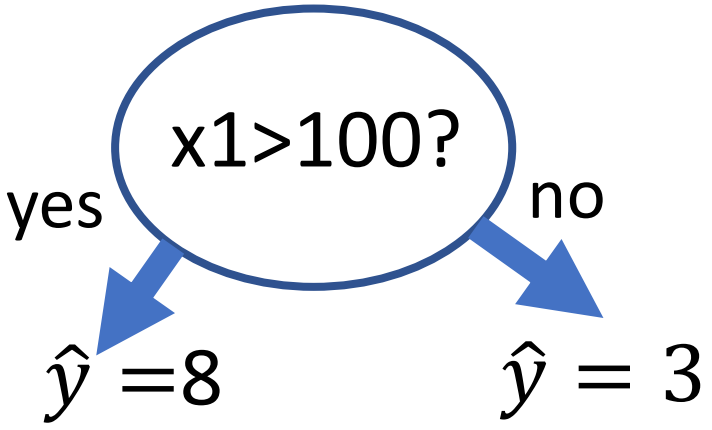
Read more in the User Guide.

discrete valued label

Hypothesis Space







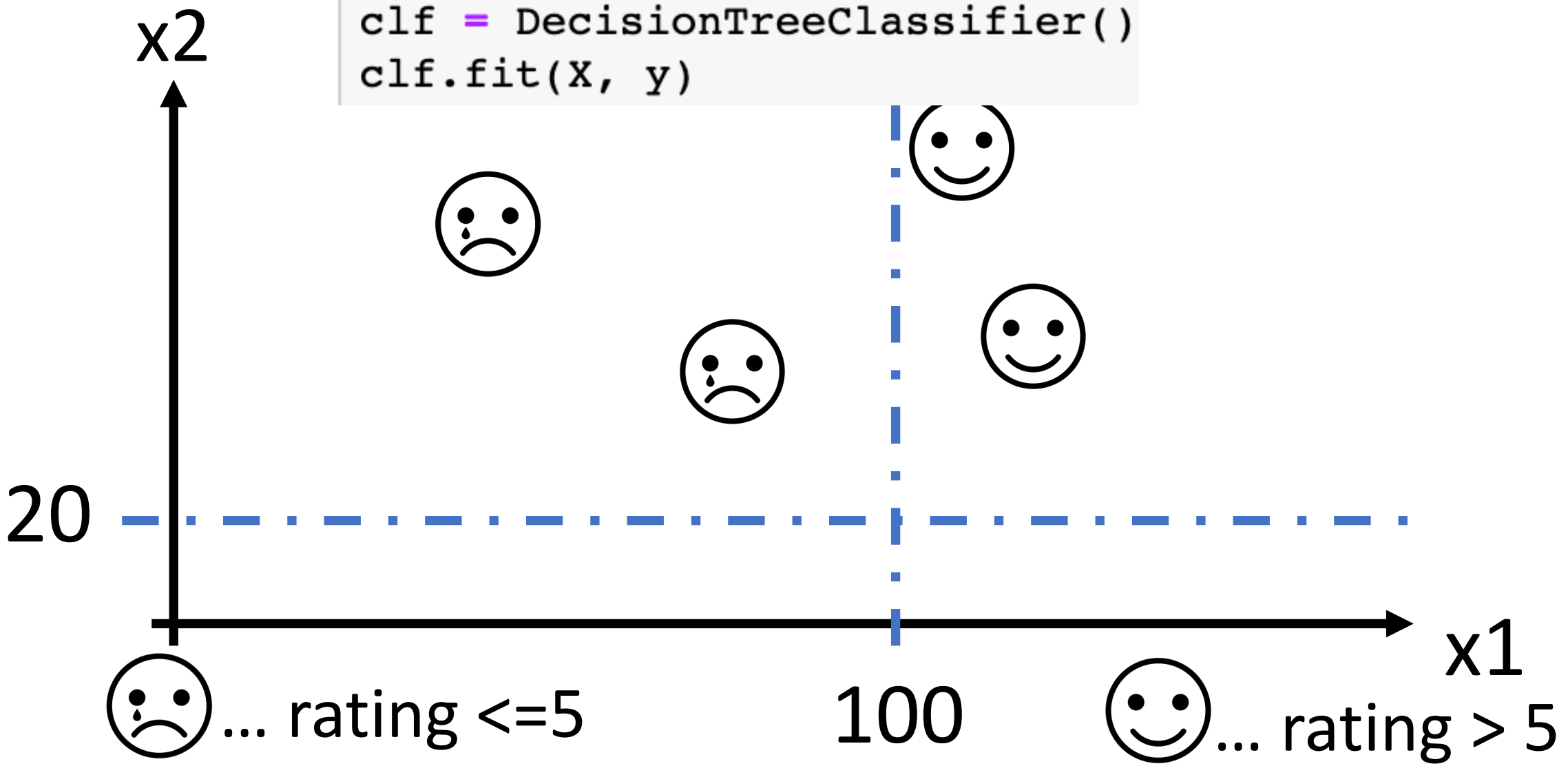
3. Loss Function

Learning a Good DT

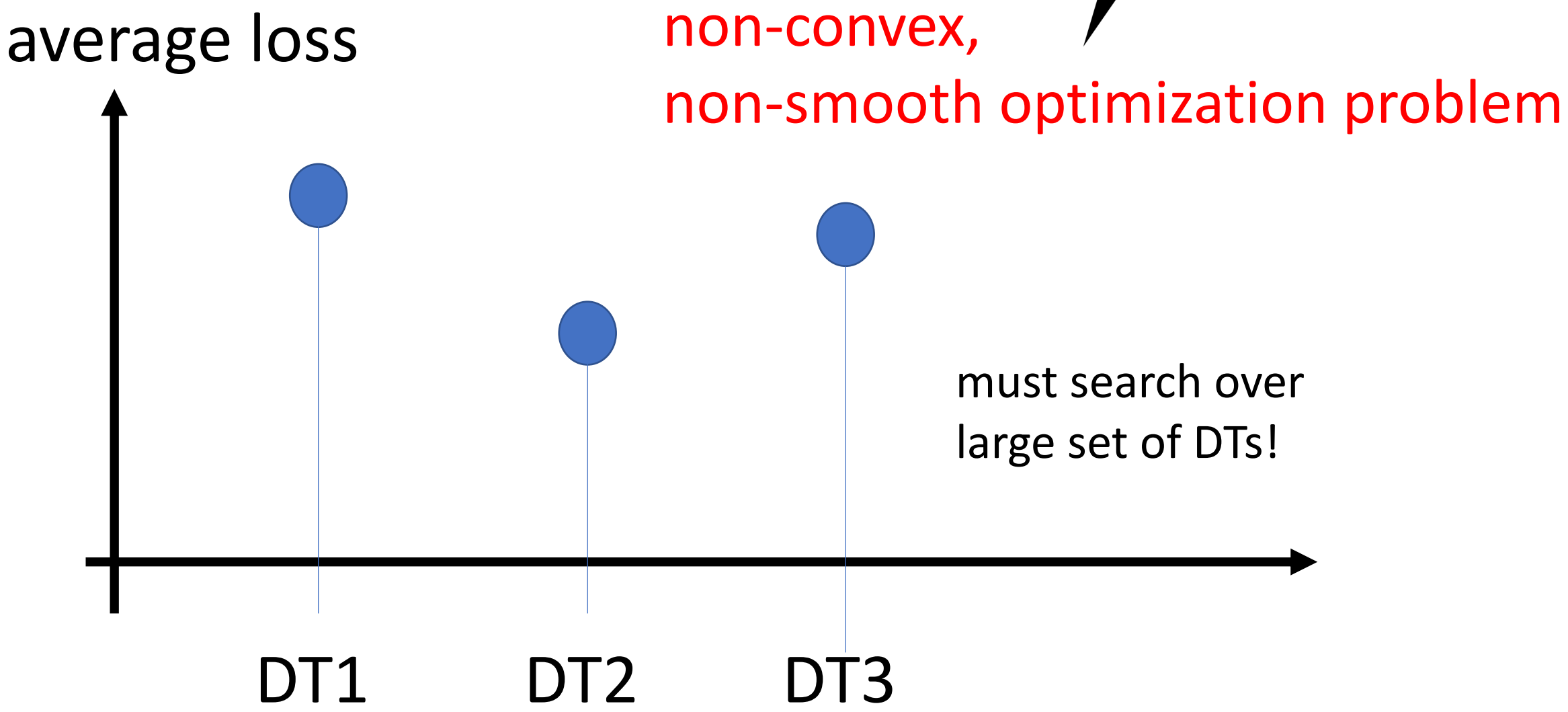
- decision tree defines a hypothesis space
- set of maps that are represented by DT
- quality of map measured by (average) loss
- can use any loss function

Some Labeled (Training) Data

```
clf = DecisionTreeClassifier()
clf.fit(X, y)
```

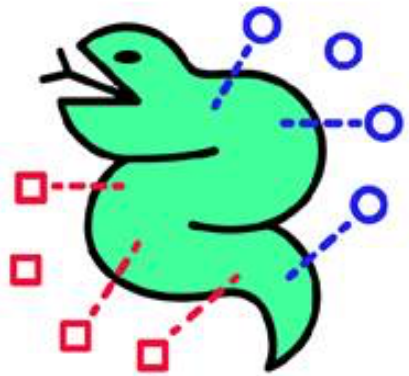


Loss Minimization



So What ?

- DT is a flow chart of predictor map
- DT define a hypothesis space
- DT can be combined with different loss functions
- DT can be used for regression or classification



Machine Learning
With Python

FITech
NETWORK UNIVERSITY

Thank You !