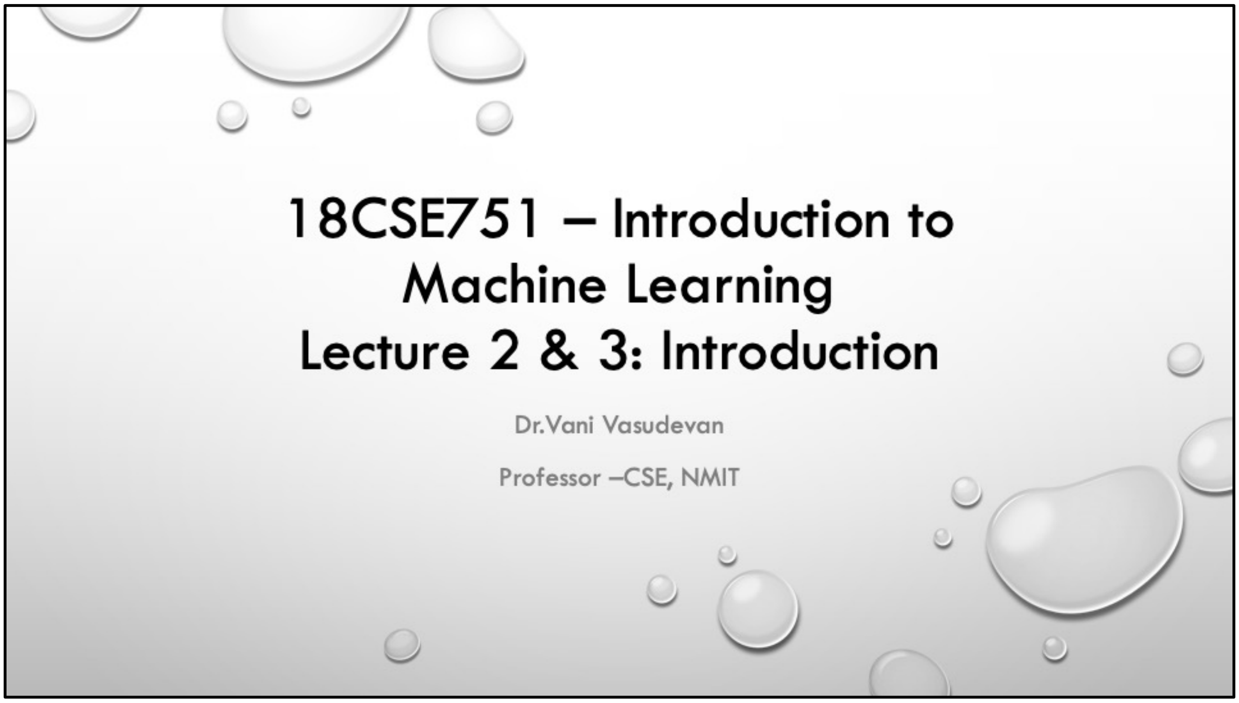
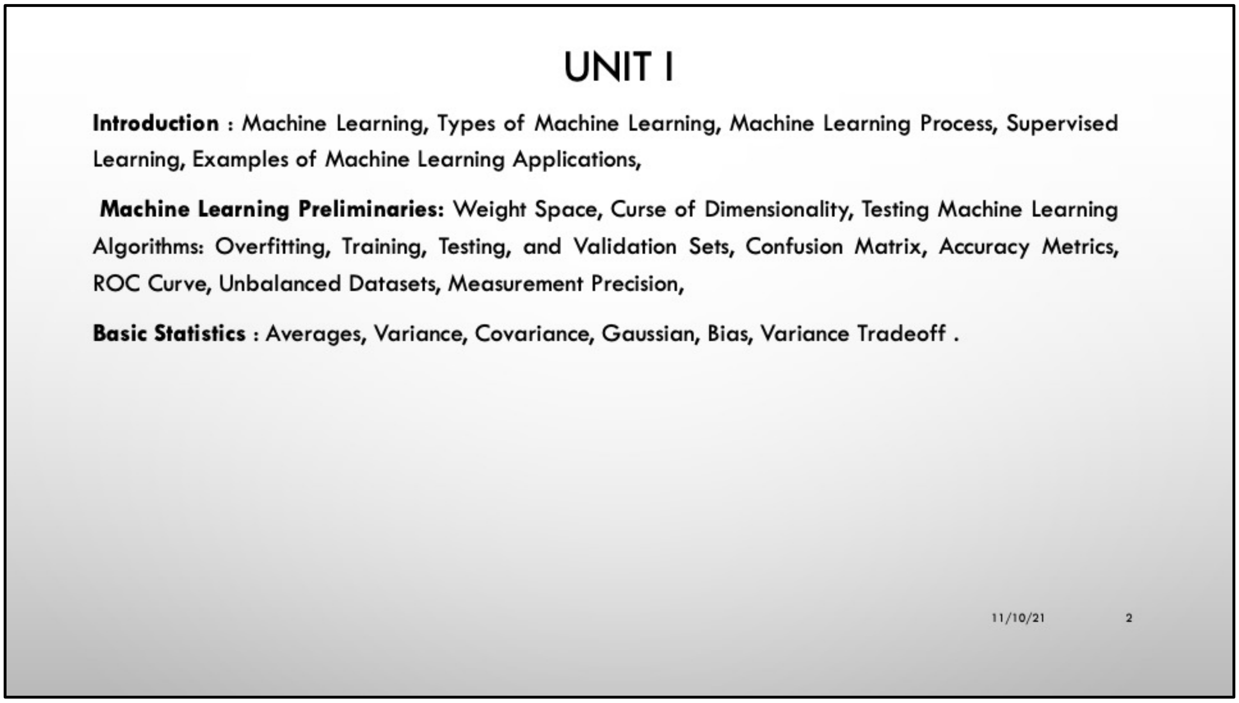


18CSE751 – Introduction to Machine Learning Lecture 2 & 3: Introduction

Dr.Vani Vasudevan

Professor –CSE, NMIT

machine learning, a subset of artificial intelligence, refers to the ability of a computer system to understand large – often huge – amounts of data, without explicit directions, and while doing so adapt and become increasingly smarter. – from UNSW blog
the use of machine learning dramatically reduces the possibility of human error.

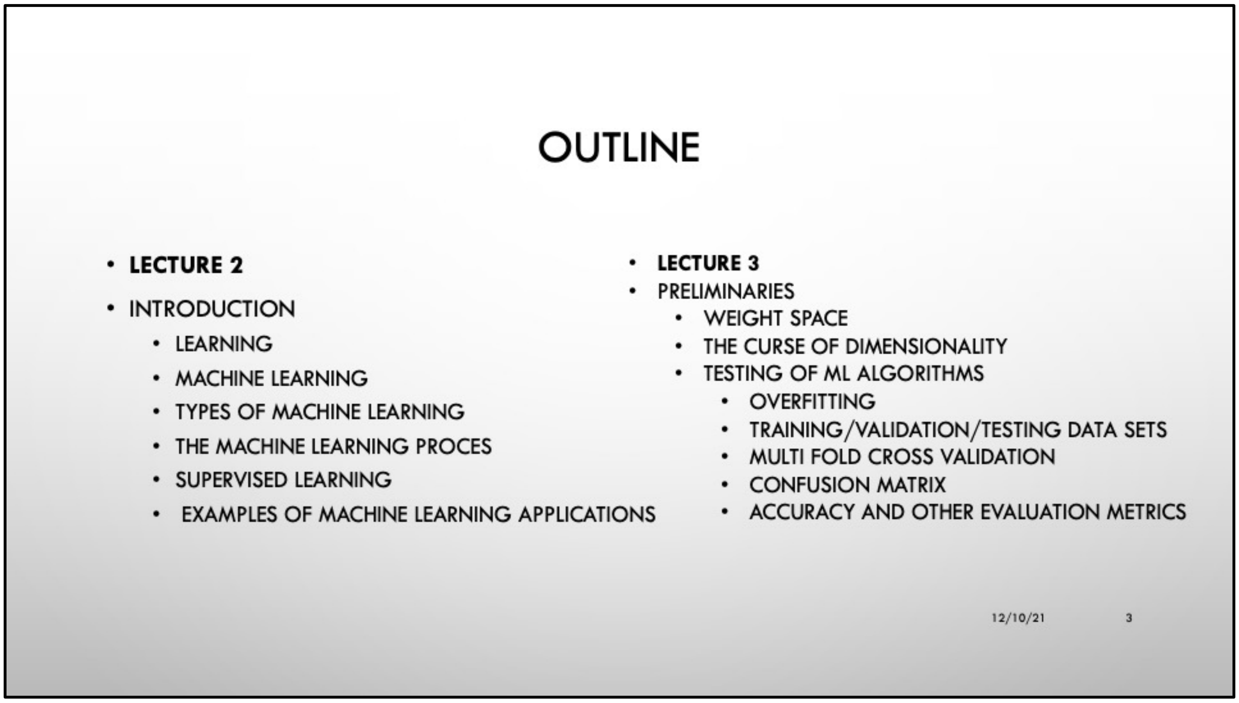


UNIT I

Introduction : Machine Learning, Types of Machine Learning, Machine Learning Process, Supervised Learning, Examples of Machine Learning Applications,

Machine Learning Preliminaries: Weight Space, Curse of Dimensionality, Testing Machine Learning Algorithms: Overfitting, Training, Testing, and Validation Sets, Confusion Matrix, Accuracy Metrics, ROC Curve, Unbalanced Datasets, Measurement Precision,

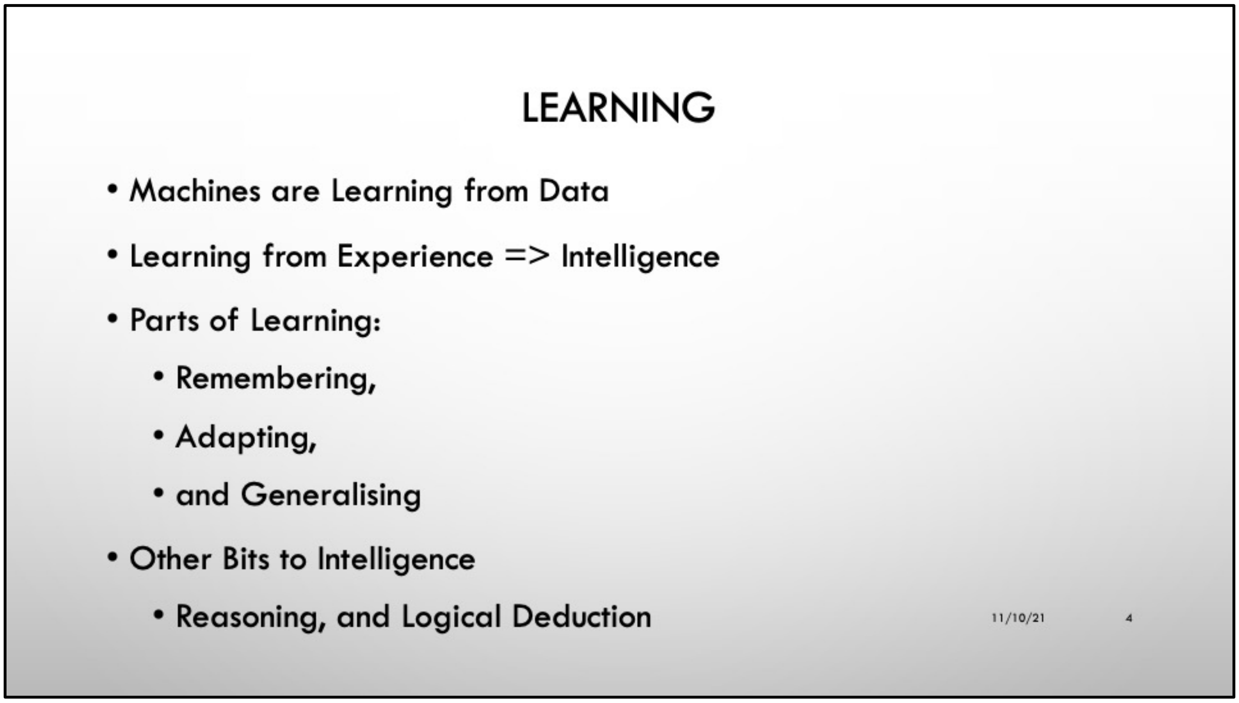
Basic Statistics : Averages, Variance, Covariance, Gaussian, Bias, Variance Tradeoff .



OUTLINE

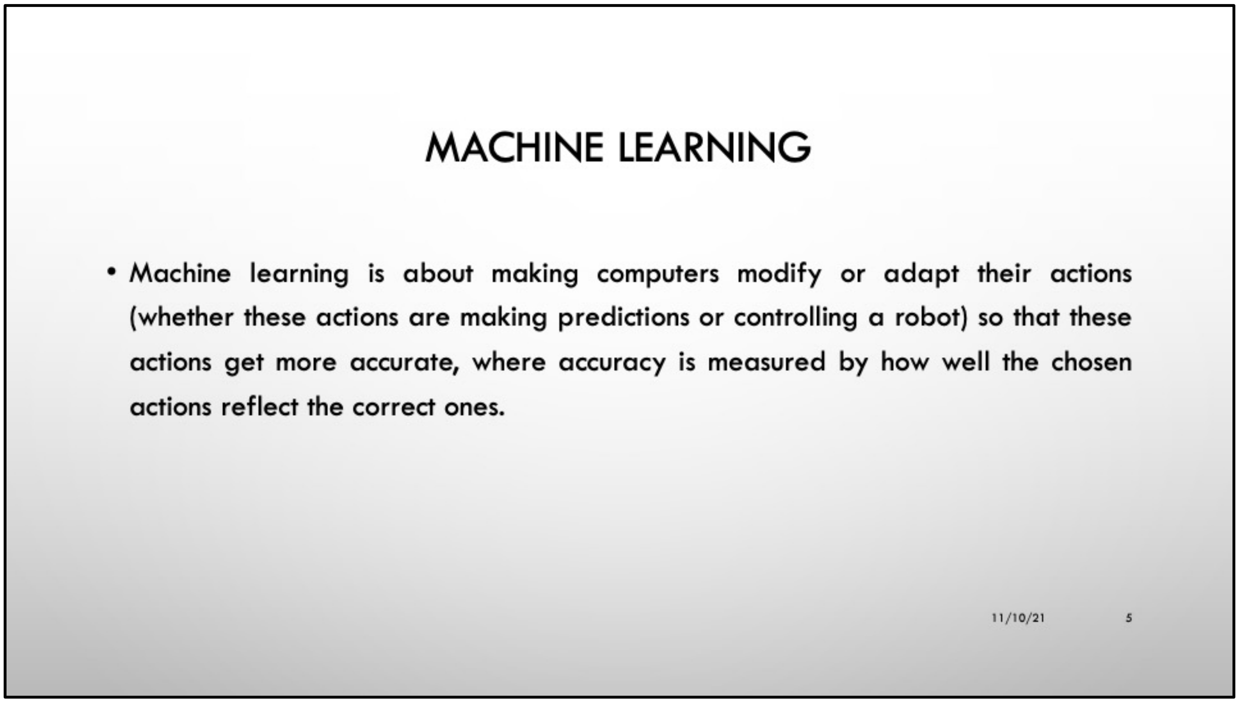
- **LECTURE 2**
- INTRODUCTION
  - LEARNING
  - MACHINE LEARNING
  - TYPES OF MACHINE LEARNING
  - THE MACHINE LEARNING PROCES
  - SUPERVISED LEARNING
  - EXAMPLES OF MACHINE LEARNING APPLICATIONS

- **LECTURE 3**
- PRELIMINARIES
  - WEIGHT SPACE
  - THE CURSE OF DIMENSIONALITY
  - TESTING OF ML ALGORITHMS
    - OVERFITTING
    - TRAINING/VALIDATION/TESTING DATA SETS
    - MULTI FOLD CROSS VALIDATION
    - CONFUSION MATRIX
    - ACCURACY AND OTHER EVALUATION METRICS



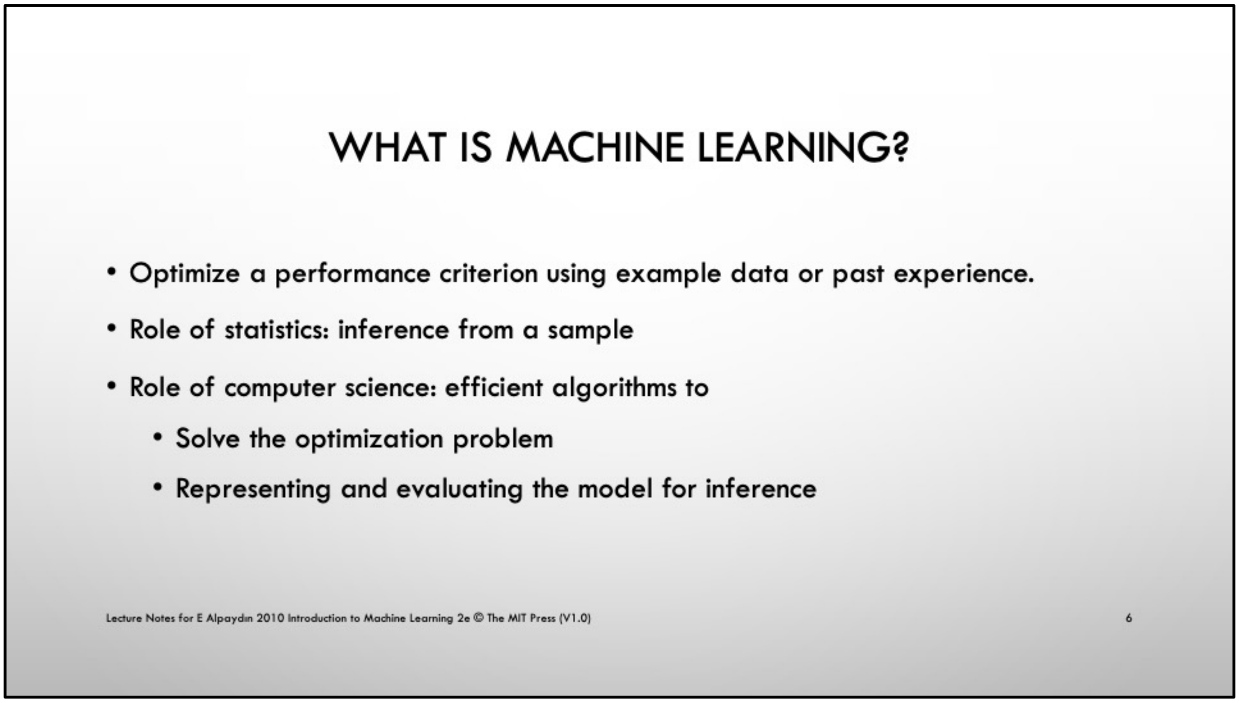
LEARNING

- Machines are Learning from Data
- Learning from Experience => Intelligence
- Parts of Learning:
  - Remembering,
  - Adapting,
  - and Generalising
- Other Bits to Intelligence
  - Reasoning, and Logical Deduction



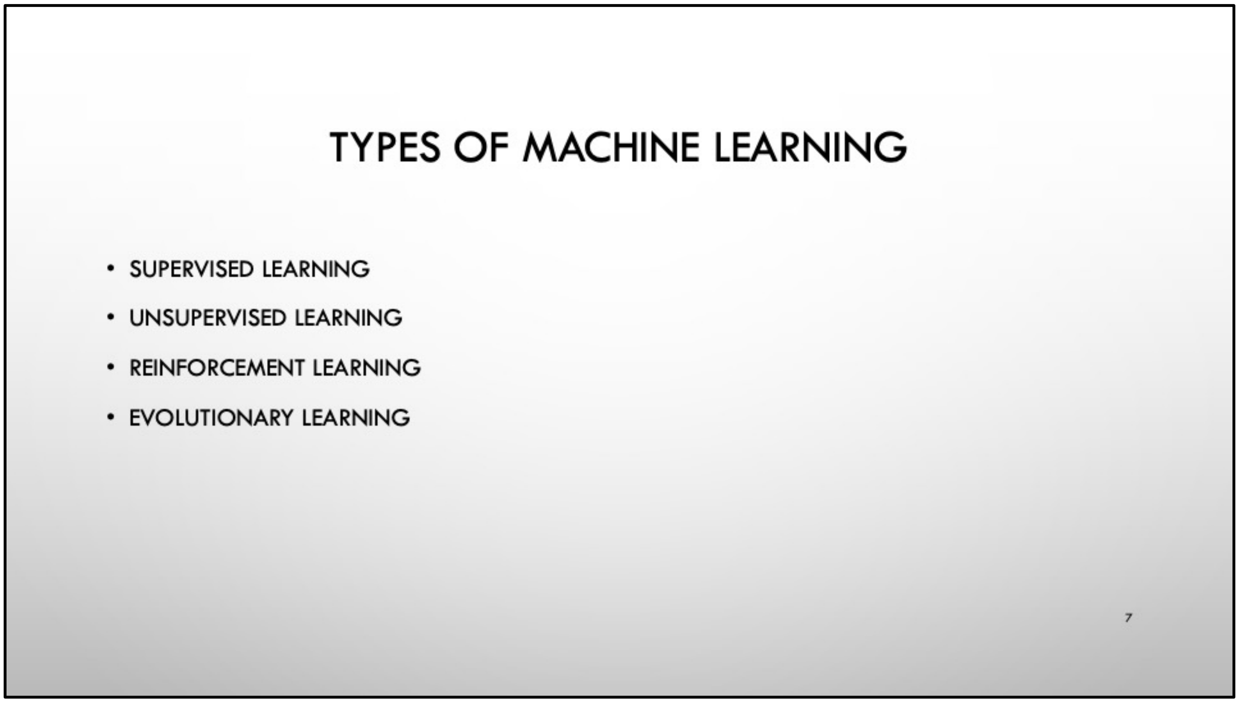
MACHINE LEARNING

- Machine learning is about making computers modify or adapt their actions (whether these actions are making predictions or controlling a robot) so that these actions get more accurate, where accuracy is measured by how well the chosen actions reflect the correct ones.



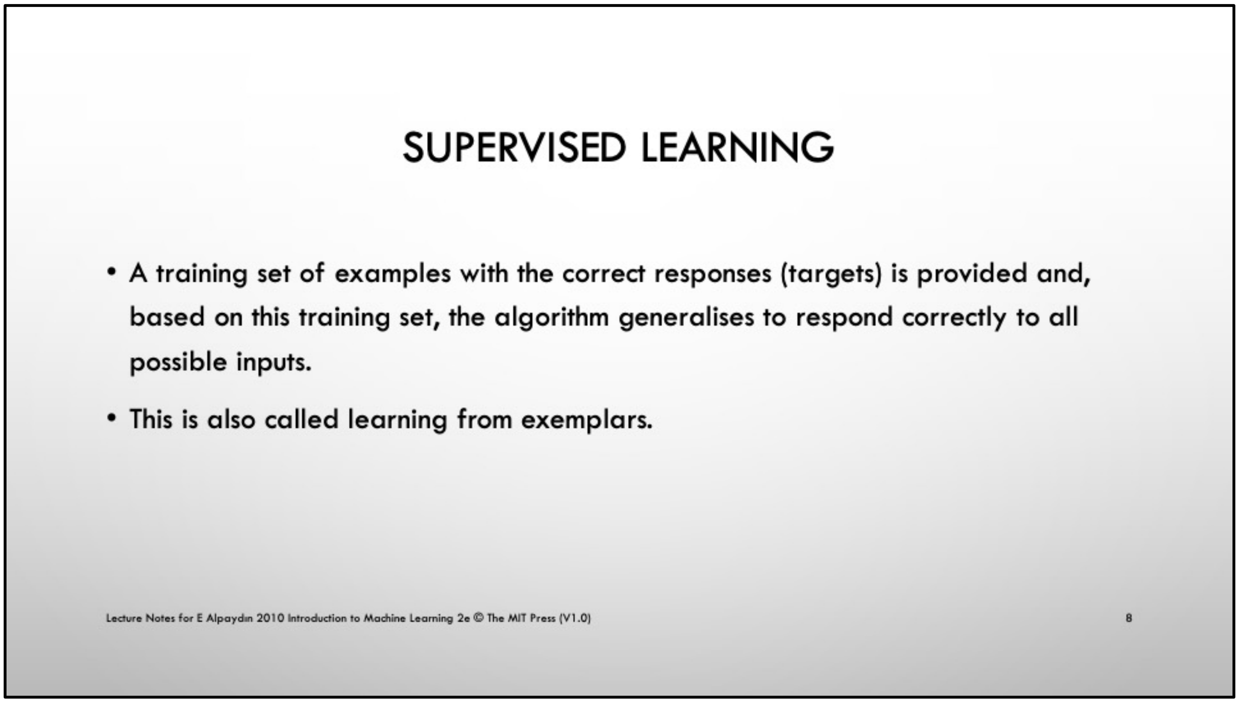
WHAT IS MACHINE LEARNING?

- Optimize a performance criterion using example data or past experience.
- Role of statistics: inference from a sample
- Role of computer science: efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference



TYPES OF MACHINE LEARNING

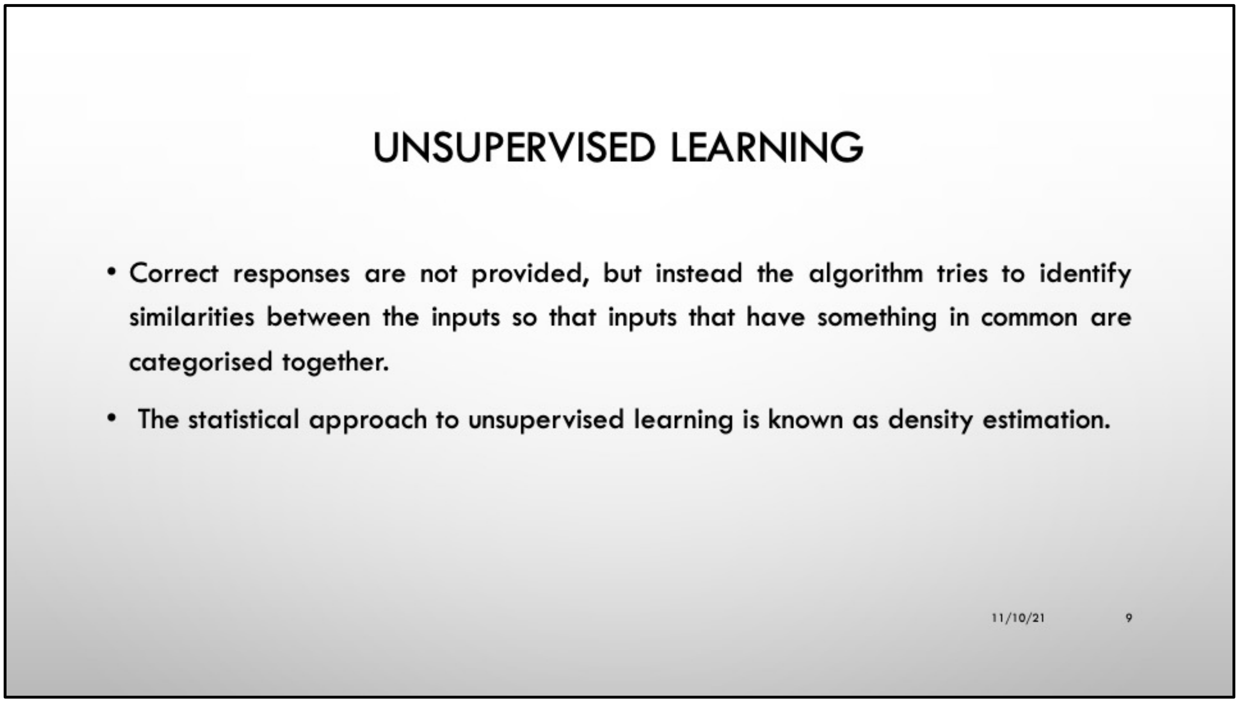
- SUPERVISED LEARNING
- UNSUPERVISED LEARNING
- REINFORCEMENT LEARNING
- EVOLUTIONARY LEARNING



SUPERVISED LEARNING

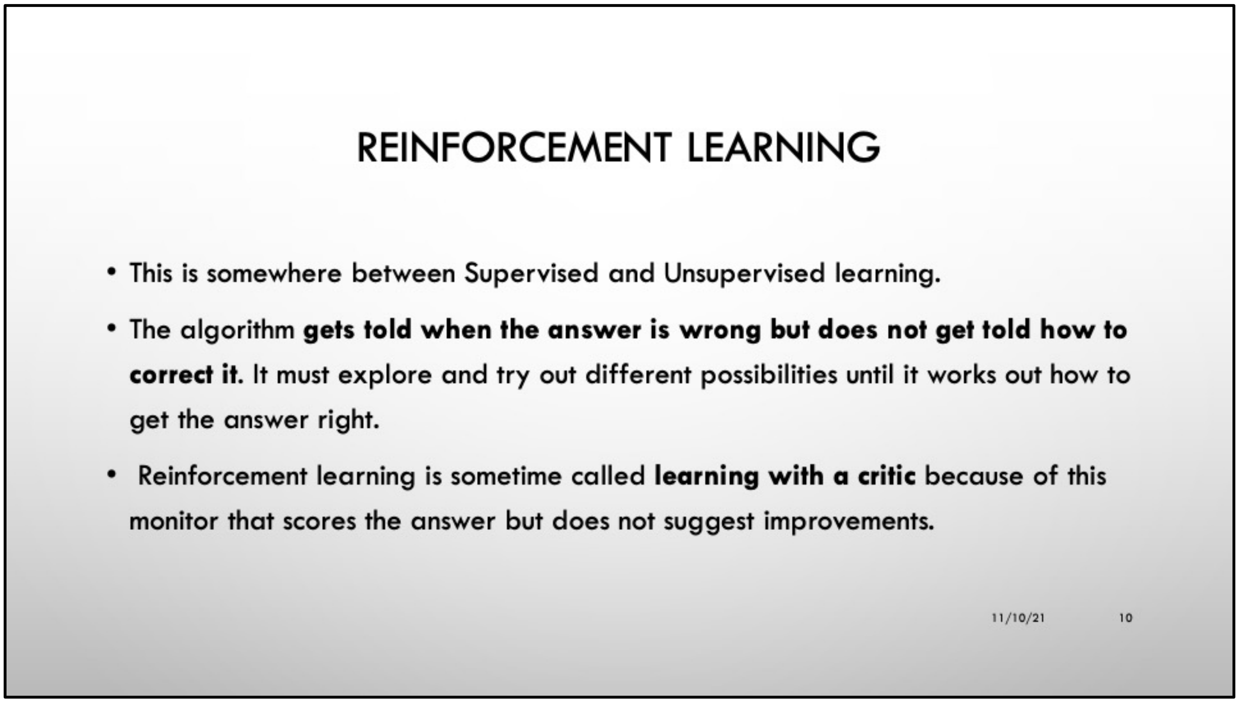
- A training set of examples with the correct responses (targets) is provided and, based on this training set, the algorithm generalises to respond correctly to all possible inputs.
- This is also called learning from exemplars.

Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)



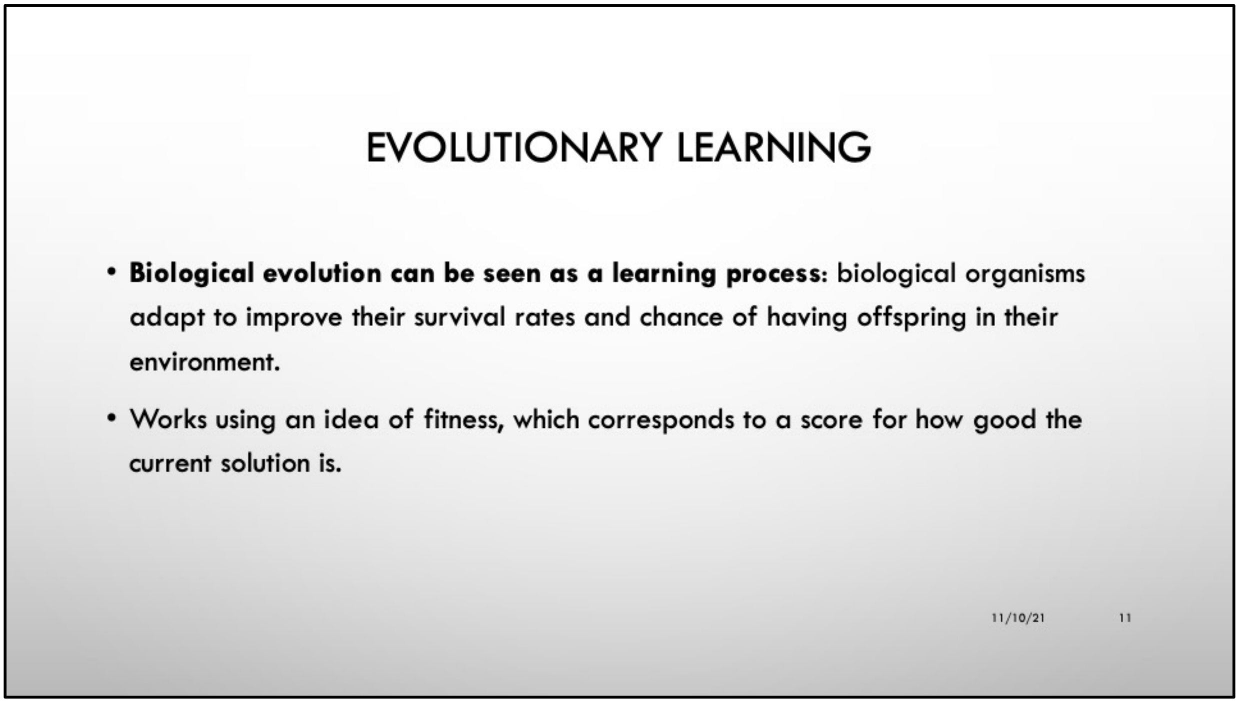
UNSUPERVISED LEARNING

- Correct responses are not provided, but instead the algorithm tries to identify similarities between the inputs so that inputs that have something in common are categorised together.
- The statistical approach to unsupervised learning is known as density estimation.



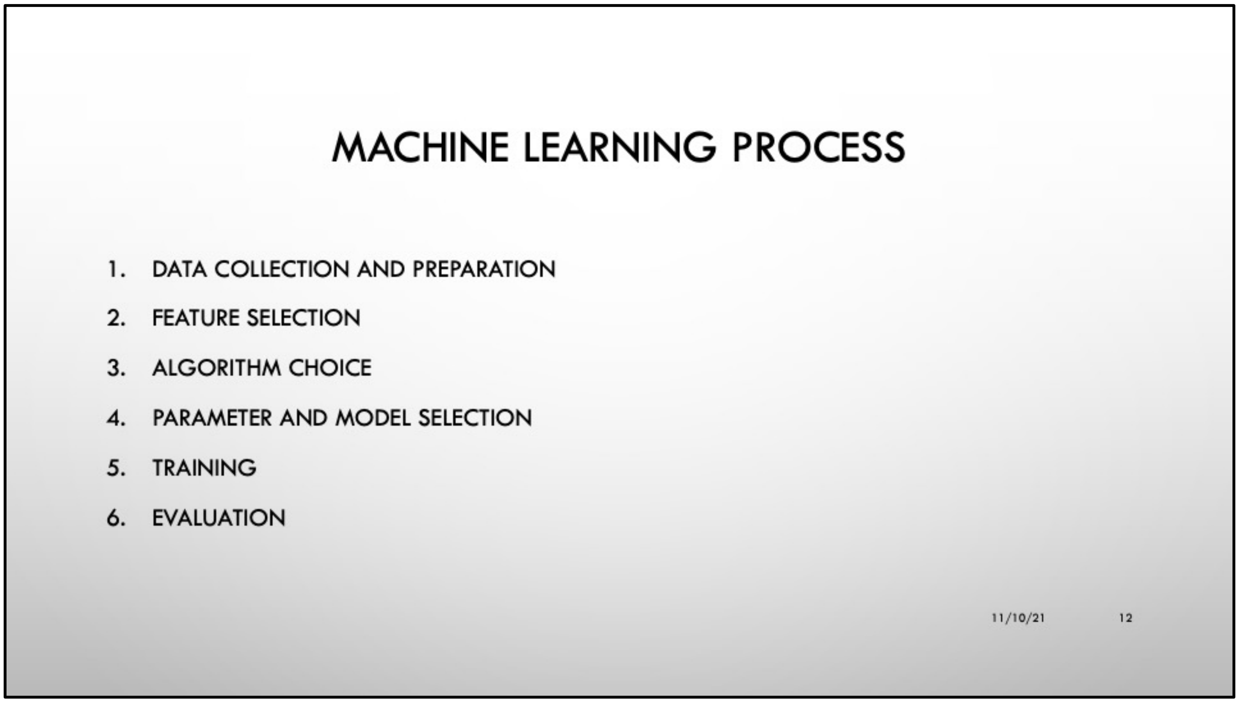
REINFORCEMENT LEARNING

- This is somewhere between Supervised and Unsupervised learning.
- The algorithm **gets told when the answer is wrong but does not get told how to correct it.** It must explore and try out different possibilities until it works out how to get the answer right.
- Reinforcement learning is sometime called **learning with a critic** because of this monitor that scores the answer but does not suggest improvements.



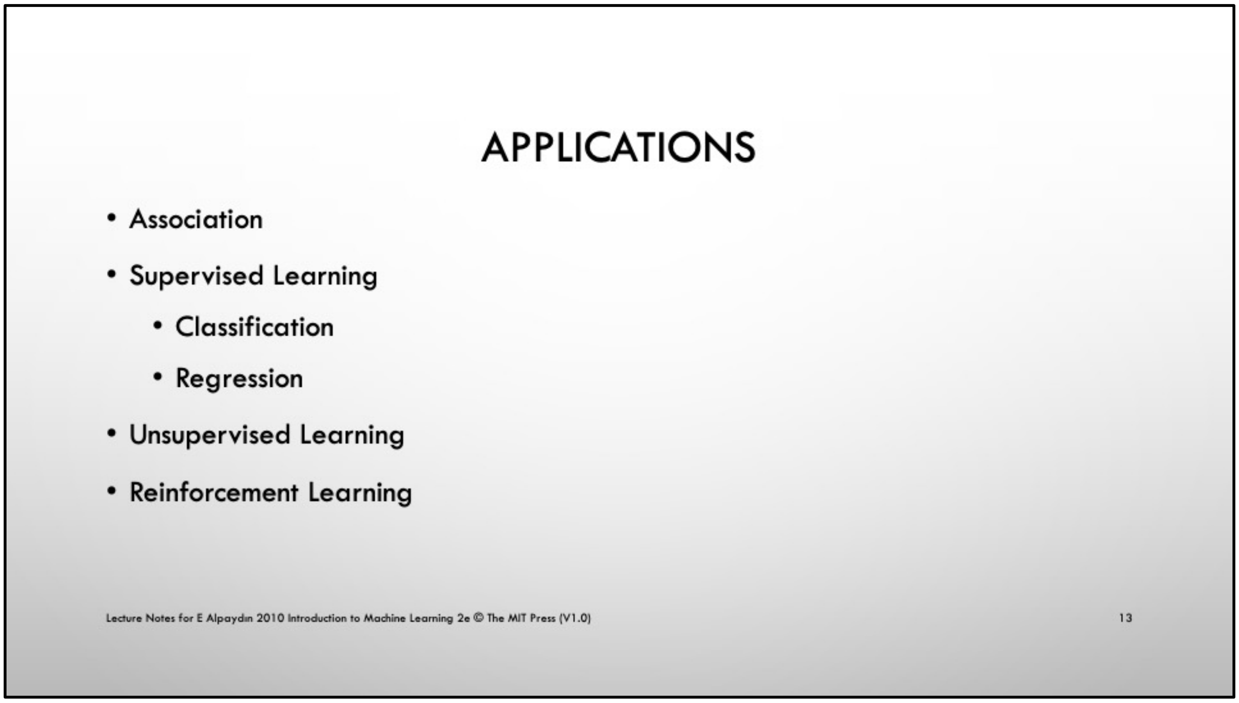
EVOLUTIONARY LEARNING

- **Biological evolution can be seen as a learning process**: biological organisms adapt to improve their survival rates and chance of having offspring in their environment.
- Works using an idea of fitness, which corresponds to a score for how good the current solution is.



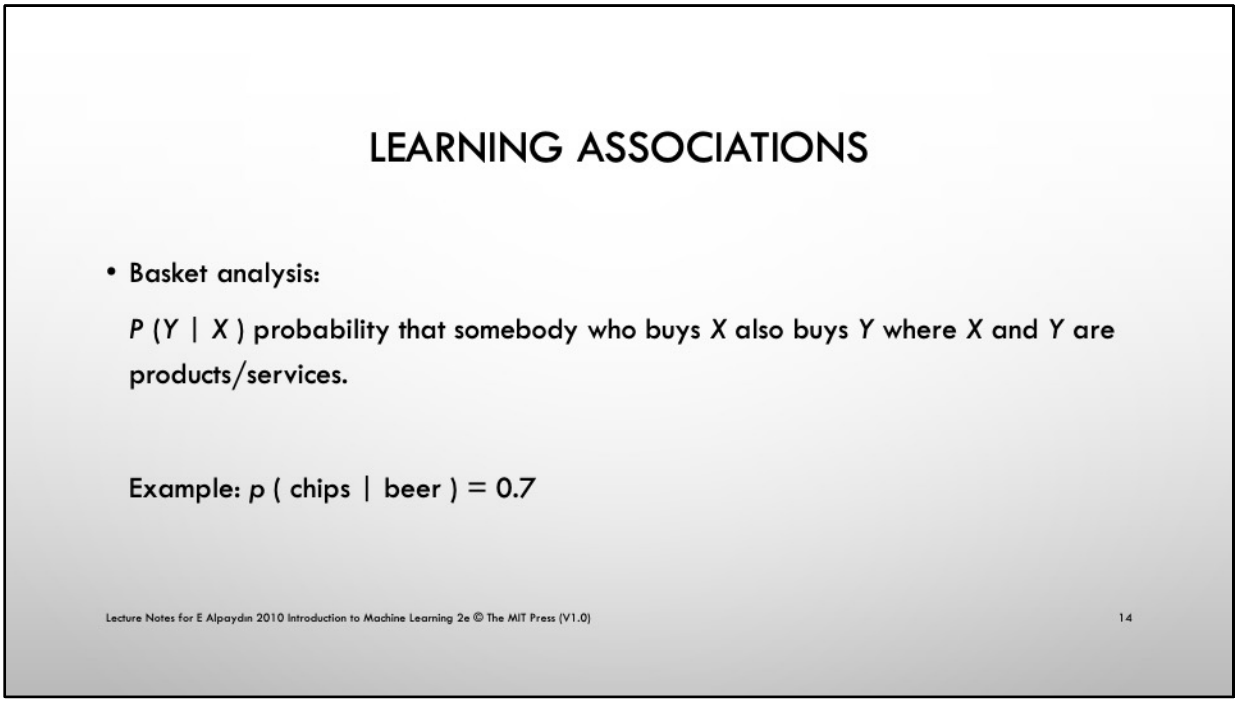
MACHINE LEARNING PROCESS

1. DATA COLLECTION AND PREPARATION
2. FEATURE SELECTION
3. ALGORITHM CHOICE
4. PARAMETER AND MODEL SELECTION
5. TRAINING
6. EVALUATION



APPLICATIONS

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

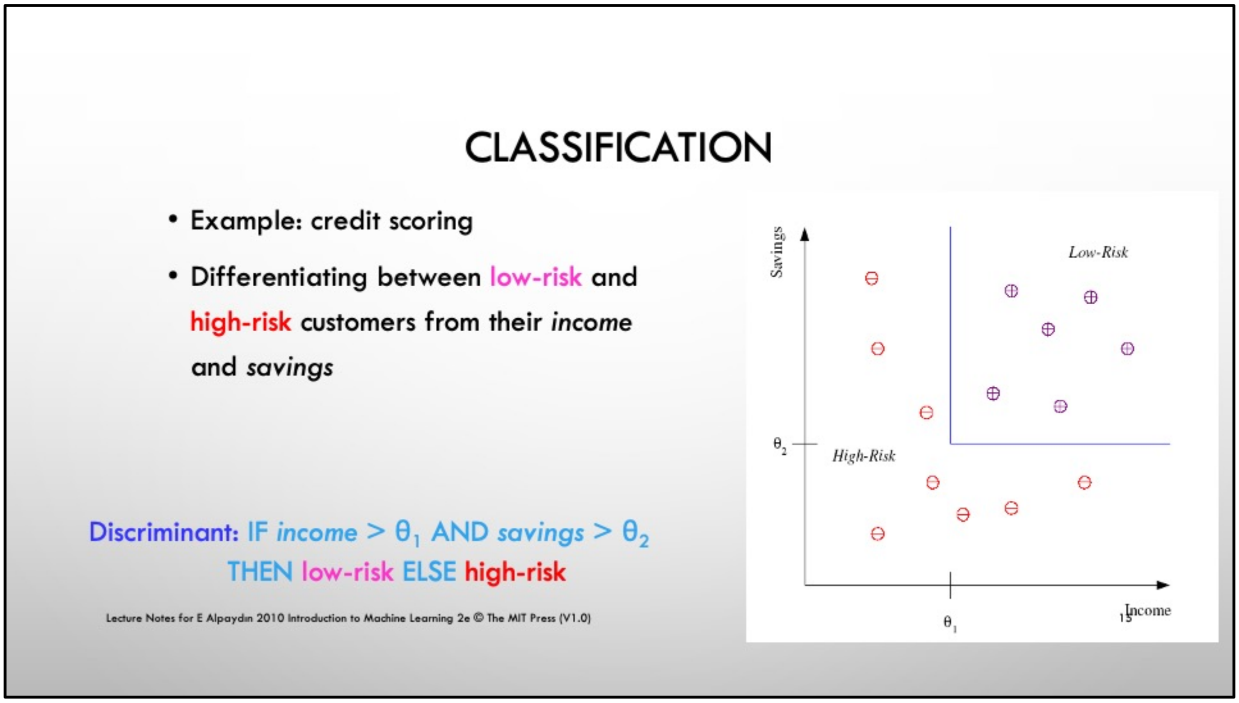


LEARNING ASSOCIATIONS

- Basket analysis:

  P (*Y* | *X*) probability that somebody who buys *X* also buys *Y* where *X* and *Y* are products/services.

  Example: *p* (chips | beer) = 0.7

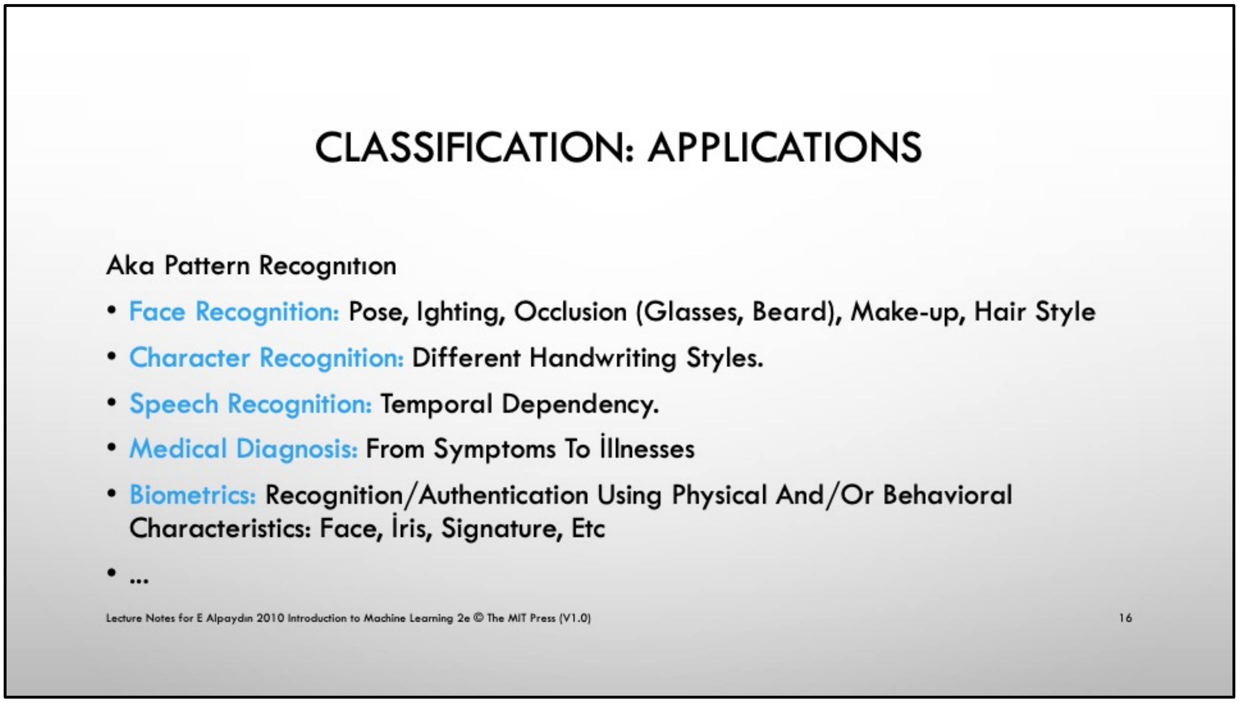


CLASSIFICATION

- Example: credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*

Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)

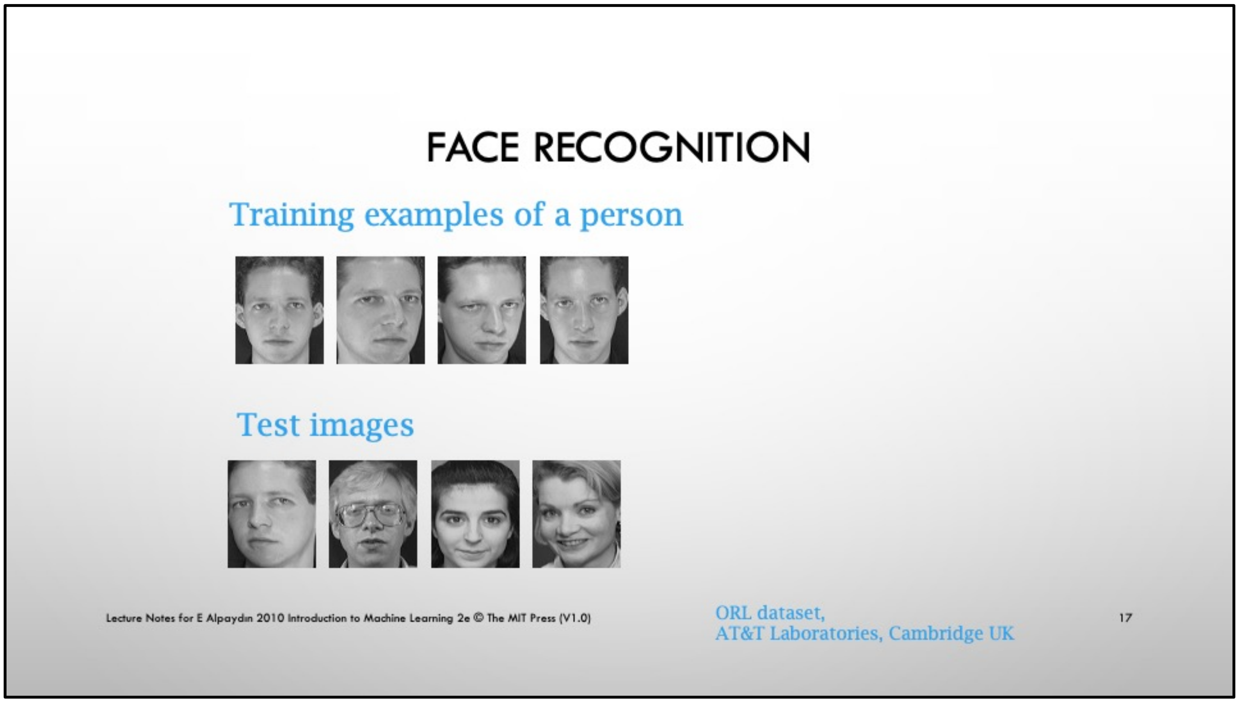


CLASSIFICATION: APPLICATIONS

Aka Pattern Recognıtıon

- Face Recognition: Pose, lghting, Occlusion (Glasses, Beard), Make-up, Hair Style
- Character Recognition: Different Handwriting Styles.
- Speech Recognition: Temporal Dependency.
- Medical Diagnosis: From Symptoms To İllnesses
- Biometrics: Recognition/Authentication Using Physical And/Or Behavioral Characteristics: Face, İris, Signature, Etc
- ...

Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0) 16



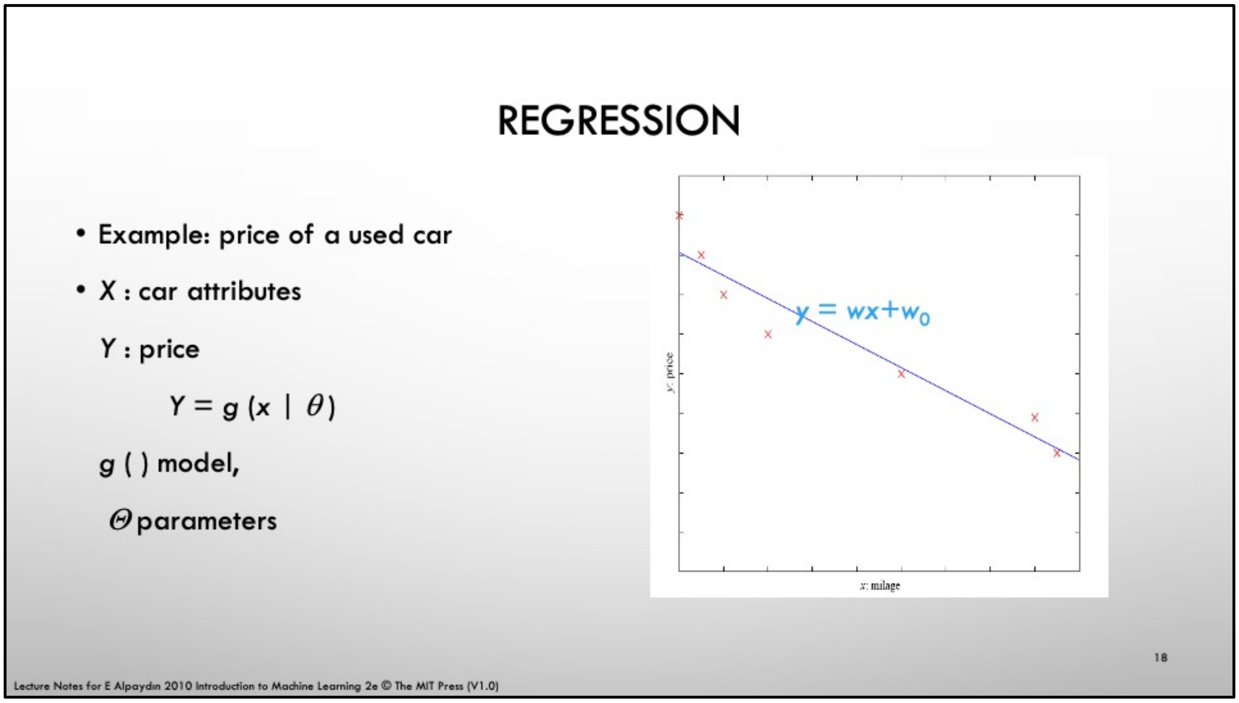
FACE RECOGNITION

Training examples of a person

Test images

Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)

ORL dataset,
AT&T Laboratories, Cambridge UK



REGRESSION

- Example: price of a used car
- X : car attributes

  Y : price

  $Y = g(x \mid \theta)$

  $g(\,)$ model,

  Θ parameters

18

Lecture Notes for E Alpaydin 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)

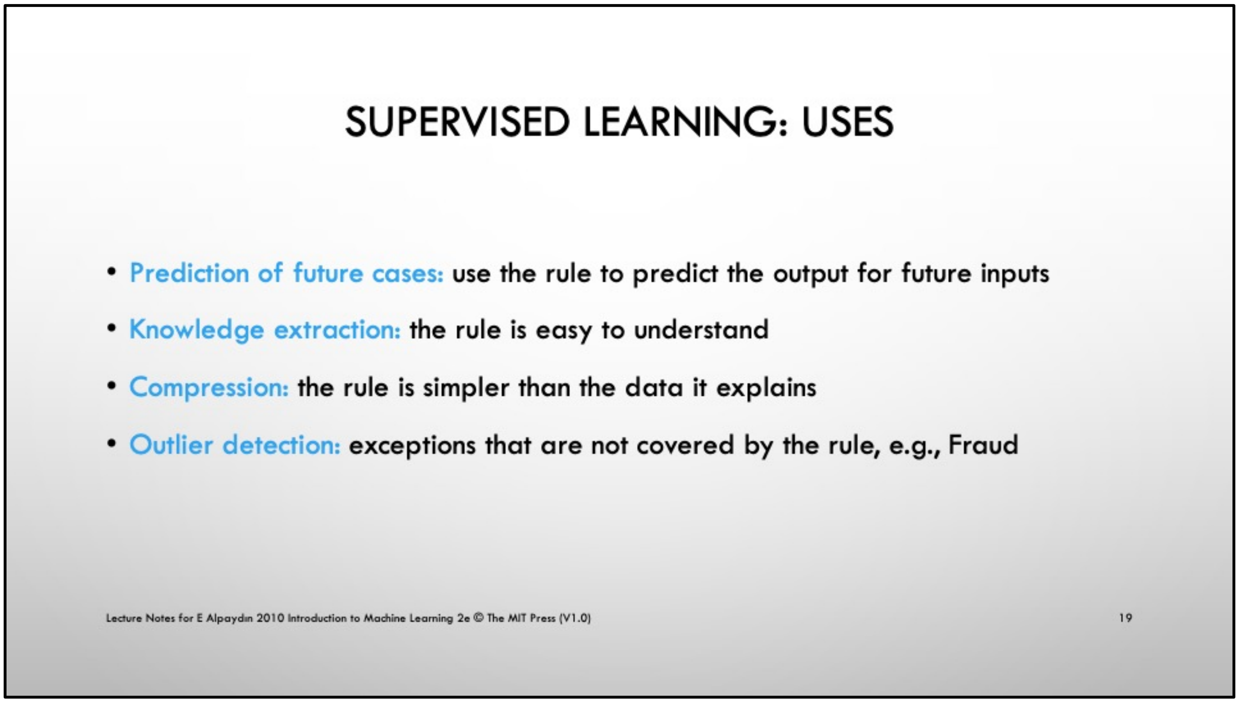
Identify correct regression technique to use in the given situations

1. The medical team at a large children's hospital wants to determine whether or not babies will be breastfeeding at the time of release based on their gestational age in weeks at the time of birth.

2. Based on the health score rankings of the top 32 cereals, Breakfast cereal manufacturing company wants to identify the best combination of sugar, fat, and fiber per serving that yields their high health score.

3. Based on the historical data, a design team at a car manufacturing wants to place a proposed design into four categories, from smallest to largest. So, we'll have small, compact, medium or large based on the overall weight and dimensions of a car.

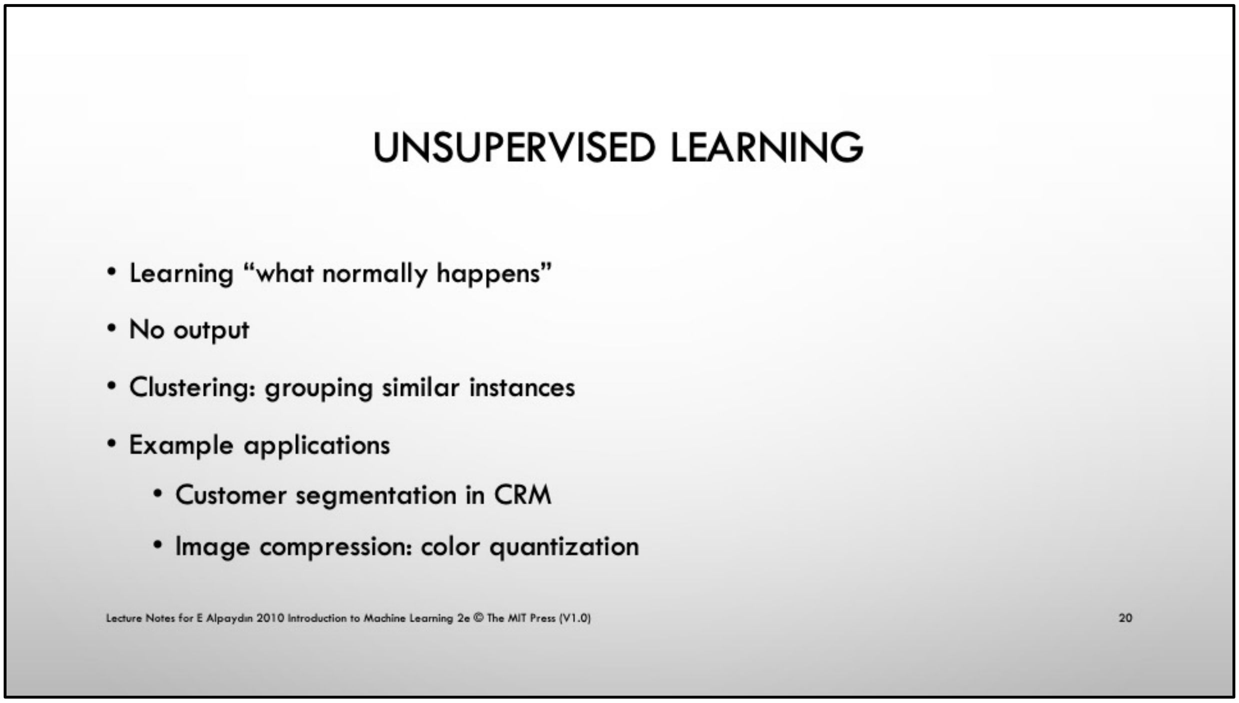
4. Based on the car mileage, the car manufacturer wants to estimate the sales price of the car.



SUPERVISED LEARNING: USES

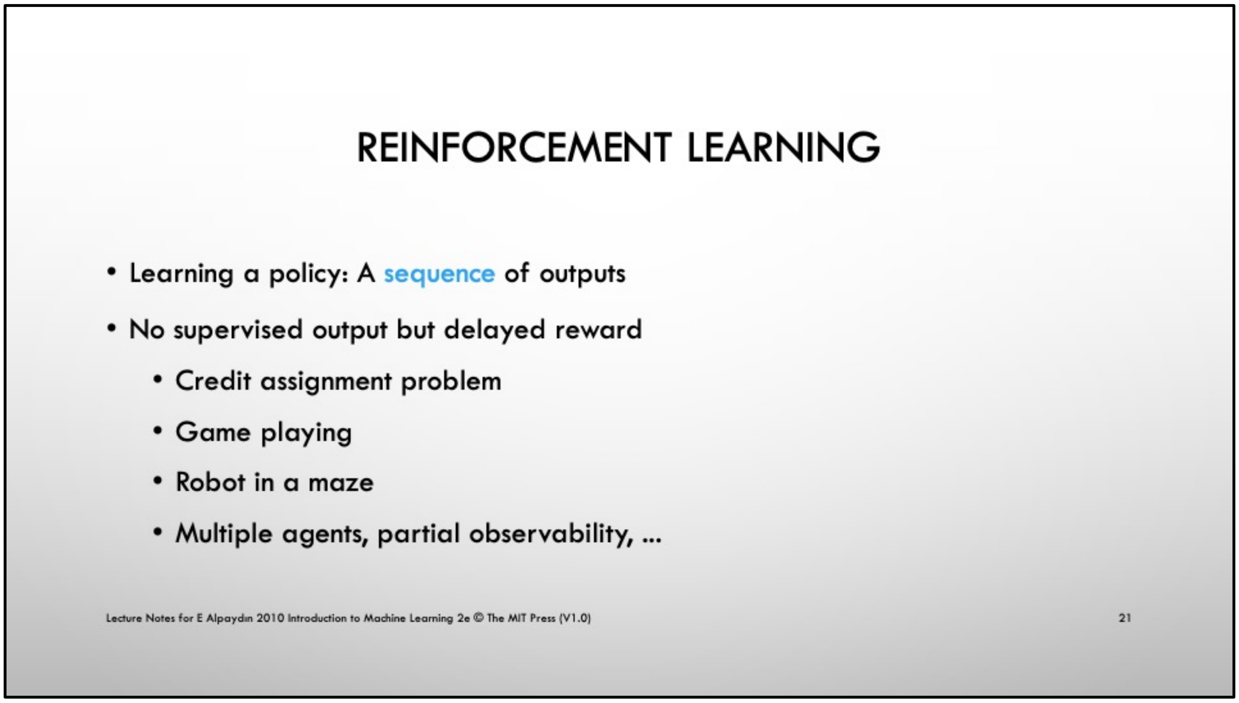
- Prediction of future cases: use the rule to predict the output for future inputs
- Knowledge extraction: the rule is easy to understand
- Compression: the rule is simpler than the data it explains
- Outlier detection: exceptions that are not covered by the rule, e.g., Fraud

Lecture Notes for E Alpaydın 2010 Introduction to Machine Learning 2e © The MIT Press (V1.0)



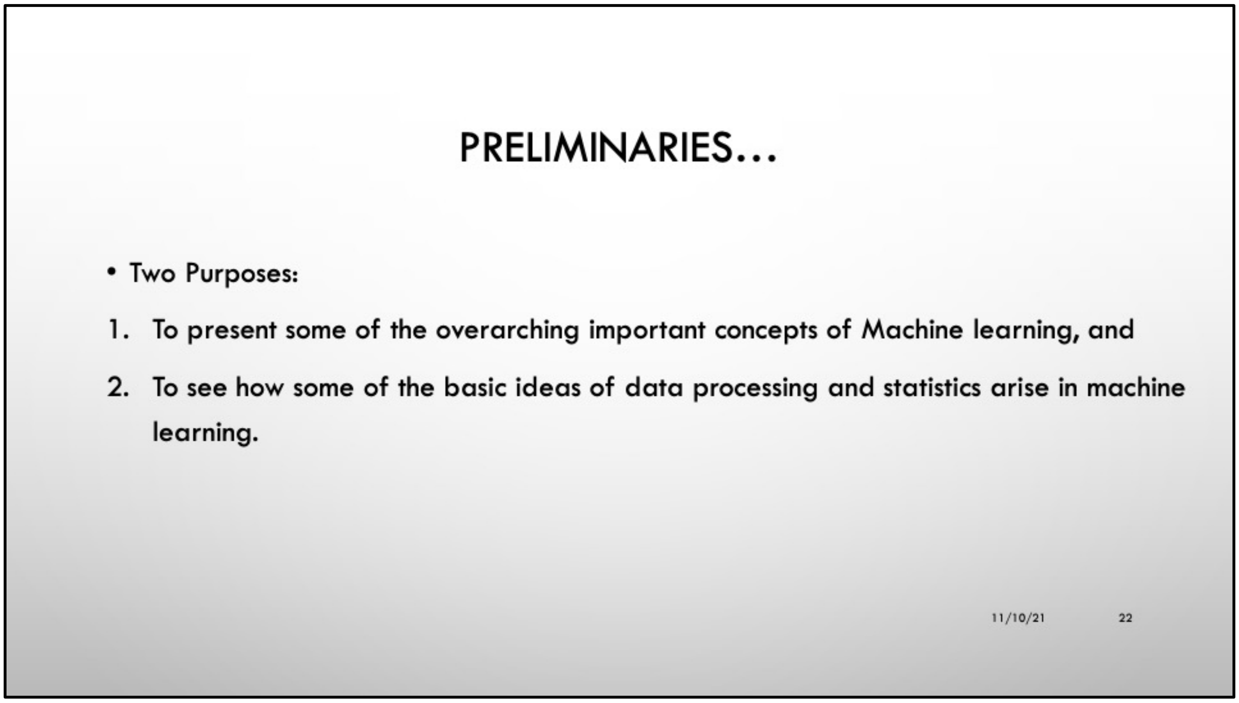
UNSUPERVISED LEARNING

- Learning “what normally happens”
- No output
- Clustering: grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: color quantization



REINFORCEMENT LEARNING

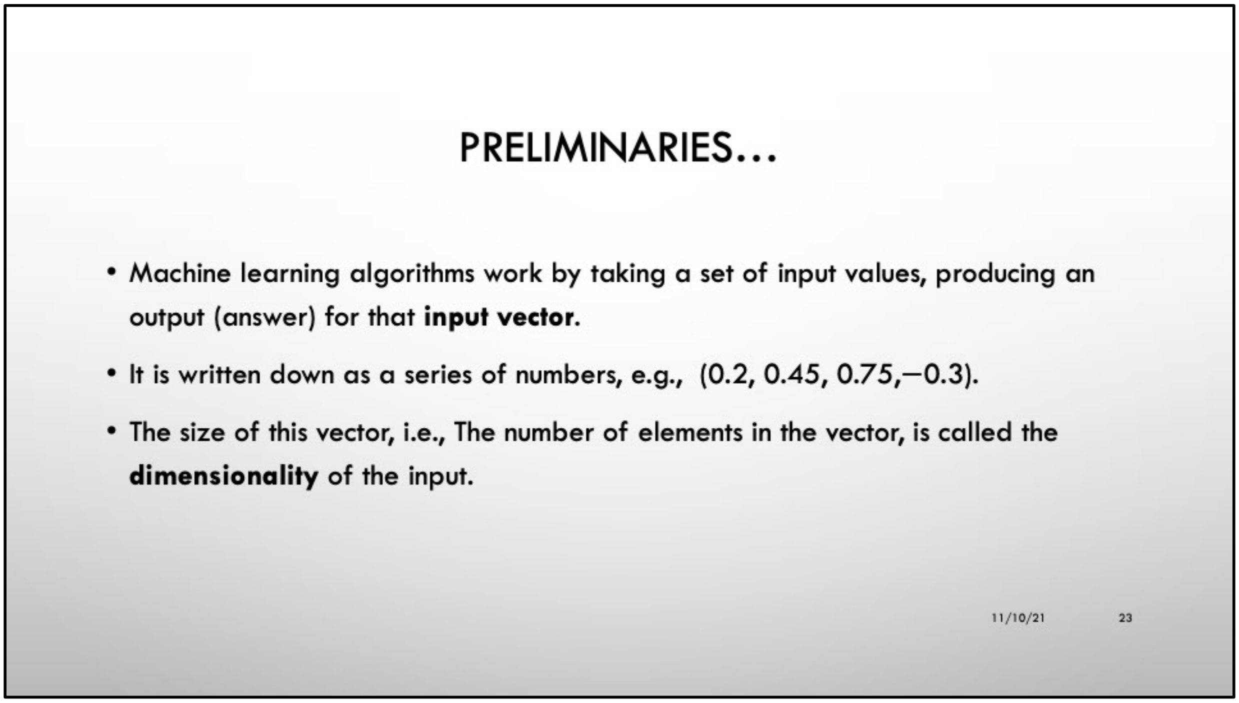
- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...



PRELIMINARIES…

- Two Purposes:

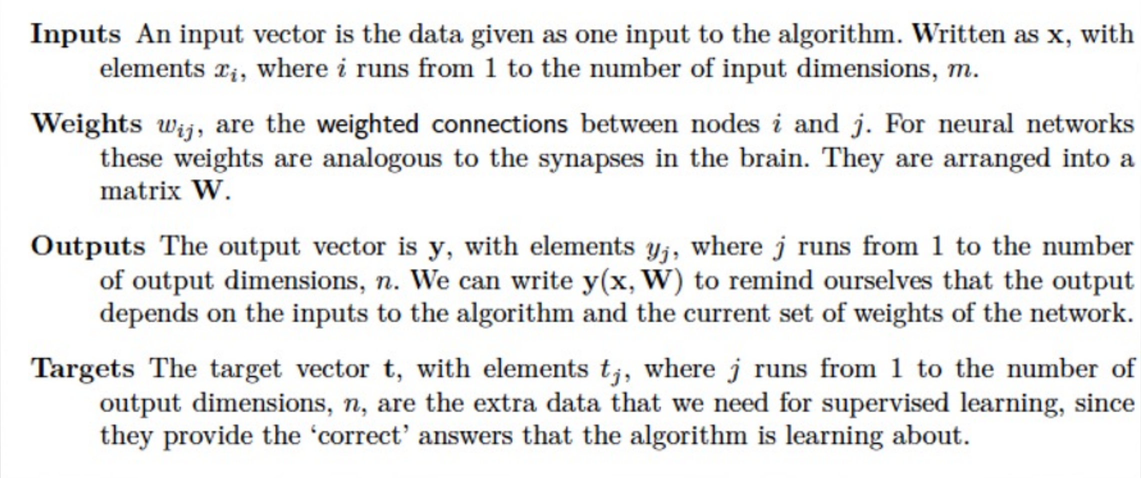
1. To present some of the overarching important concepts of Machine learning, and
2. To see how some of the basic ideas of data processing and statistics arise in machine learning.



PRELIMINARIES…

- Machine learning algorithms work by taking a set of input values, producing an output (answer) for that **input vector.**
- It is written down as a series of numbers, e.g., (0.2, 0.45, 0.75,−0.3).
- The size of this vector, i.e., The number of elements in the vector, is called the **dimensionality** of the input.

PRELIMINARIES…

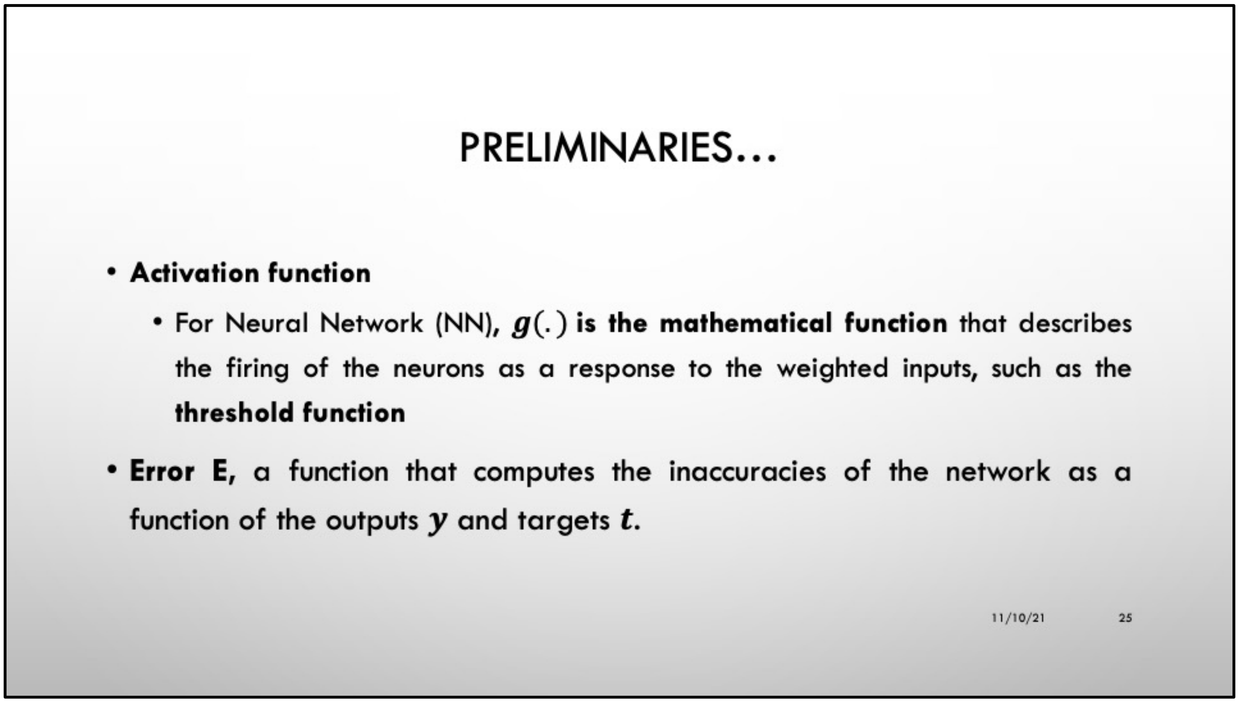


Inputs An input vector is the data given as one input to the algorithm. Written as $\mathbf{x}$, with elements x_i, where i runs from 1 to the number of input dimensions, m.

Weights w_{ij}, are the weighted connections between nodes i and j. For neural networks these weights are analogous to the synapses in the brain. They are arranged into a matrix $\mathbf{W}$.

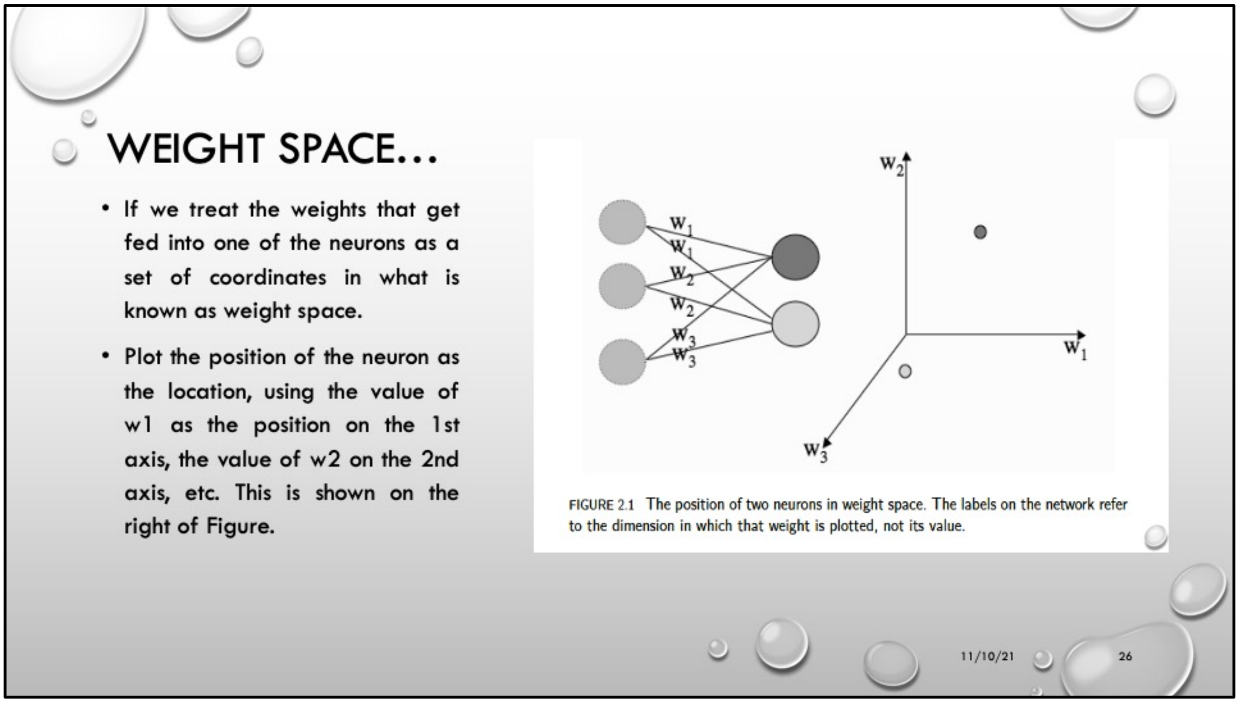
Outputs The output vector is $\mathbf{y}$, with elements y_j, where j runs from 1 to the number of output dimensions, n. We can write $\mathbf{y}(\mathbf{x}, \mathbf{W})$ to remind ourselves that the output depends on the inputs to the algorithm and the current set of weights of the network.

Targets The target vector $\mathbf{t}$, with elements t_j, where j runs from 1 to the number of output dimensions, n, are the extra data that we need for supervised learning, since they provide the 'correct' answers that the algorithm is learning about.



PRELIMINARIES…

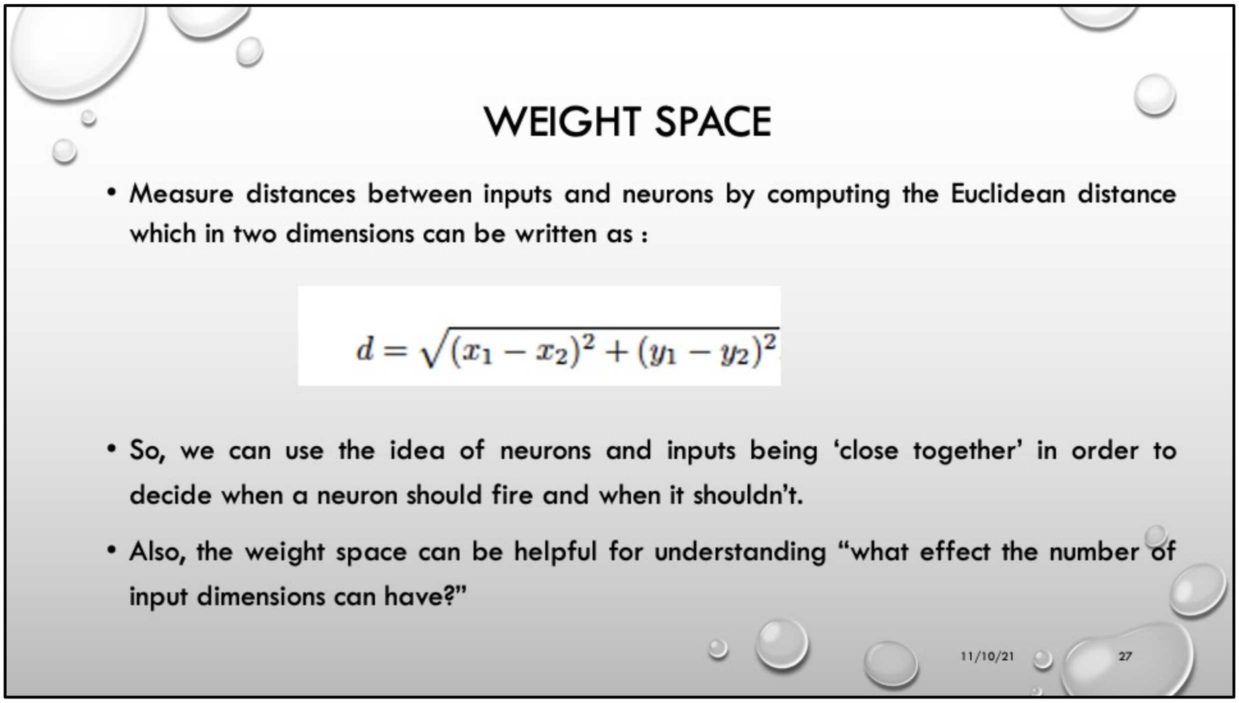
- **Activation function**
  - For Neural Network (NN), $g(.)$ **is the mathematical function** that describes the firing of the neurons as a response to the weighted inputs, such as the **threshold function**
- **Error E,** a function that computes the inaccuracies of the network as a function of the outputs y and targets t.



WEIGHT SPACE…

- If we treat the weights that get fed into one of the neurons as a set of coordinates in what is known as weight space.
- Plot the position of the neuron as the location, using the value of w1 as the position on the 1st axis, the value of w2 on the 2nd axis, etc. This is shown on the right of Figure.

FIGURE 2.1 The position of two neurons in weight space. The labels on the network refer to the dimension in which that weight is plotted, not its value.



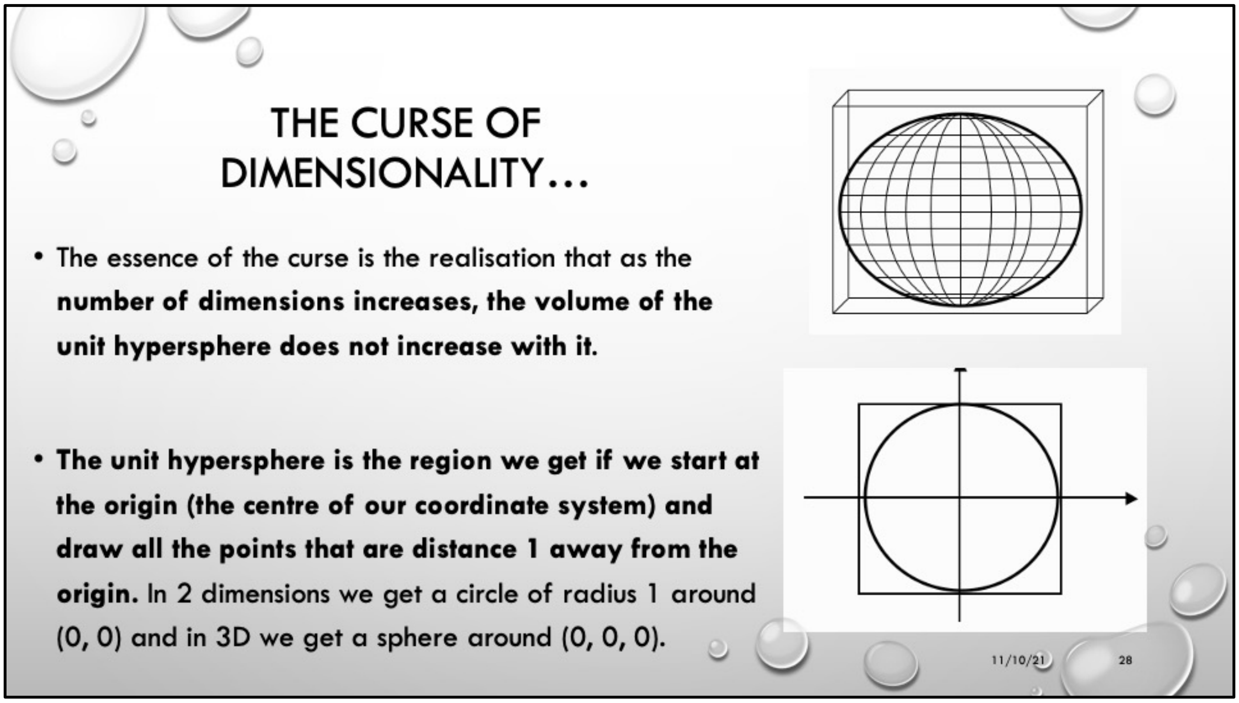
WEIGHT SPACE

- Measure distances between inputs and neurons by computing the Euclidean distance which in two dimensions can be written as :

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- So, we can use the idea of neurons and inputs being 'close together' in order to decide when a neuron should fire and when it shouldn't.
- Also, the weight space can be helpful for understanding "what effect the number of input dimensions can have?"

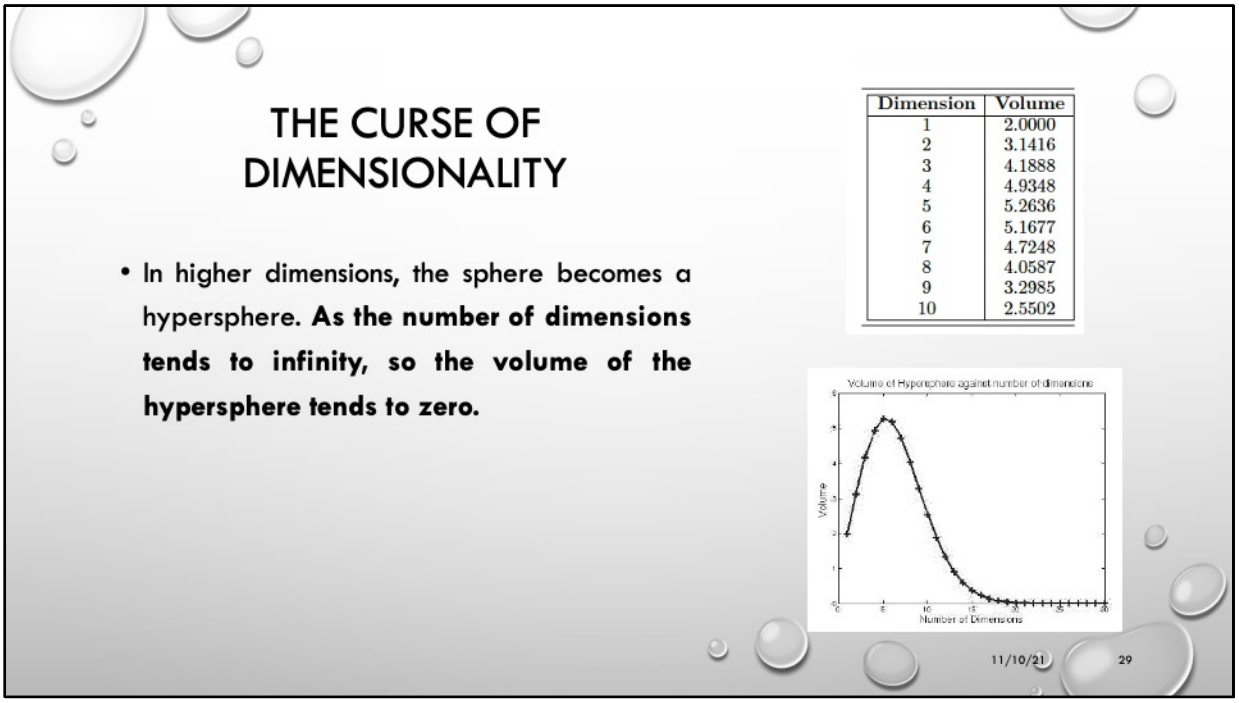
11/10/21 27



THE CURSE OF DIMENSIONALITY…

- The essence of the curse is the realisation that as the **number of dimensions increases, the volume of the unit hypersphere does not increase with it.**

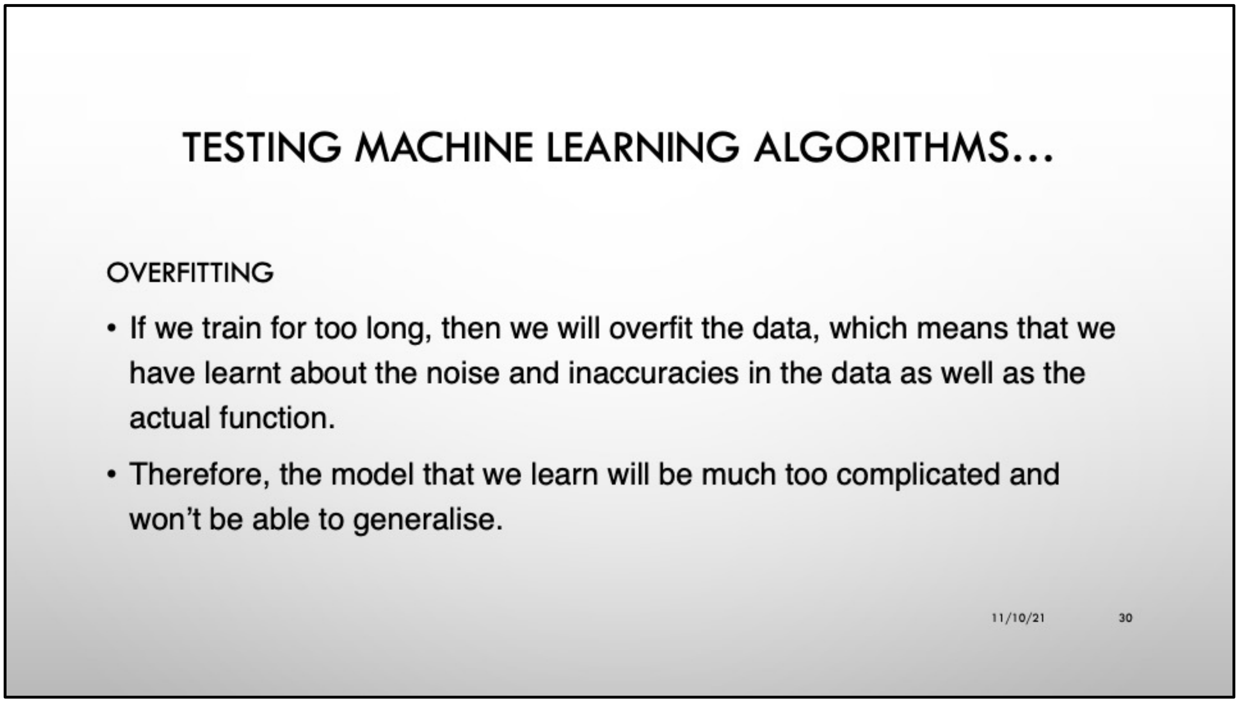
- **The unit hypersphere is the region we get if we start at the origin (the centre of our coordinate system) and draw all the points that are distance 1 away from the origin.** In 2 dimensions we get a circle of radius 1 around (0, 0) and in 3D we get a sphere around (0, 0, 0).



THE CURSE OF DIMENSIONALITY

- In higher dimensions, the sphere becomes a hypersphere. **As the number of dimensions tends to infinity, so the volume of the hypersphere tends to zero.**

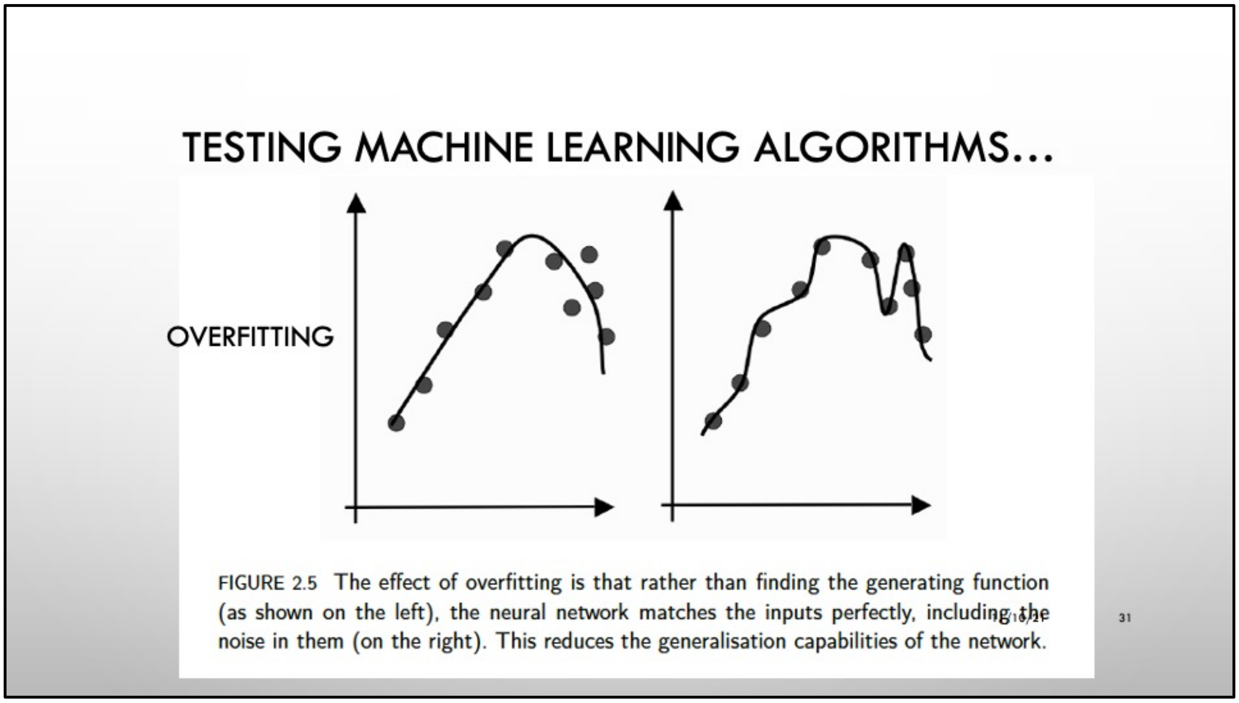
Dimension	Volume
1	2.0000
2	3.1416
3	4.1888
4	4.9348
5	5.2636
6	5.1677
7	4.7248
8	4.0587
9	3.2985
10	2.5502



TESTING MACHINE LEARNING ALGORITHMS…

OVERFITTING

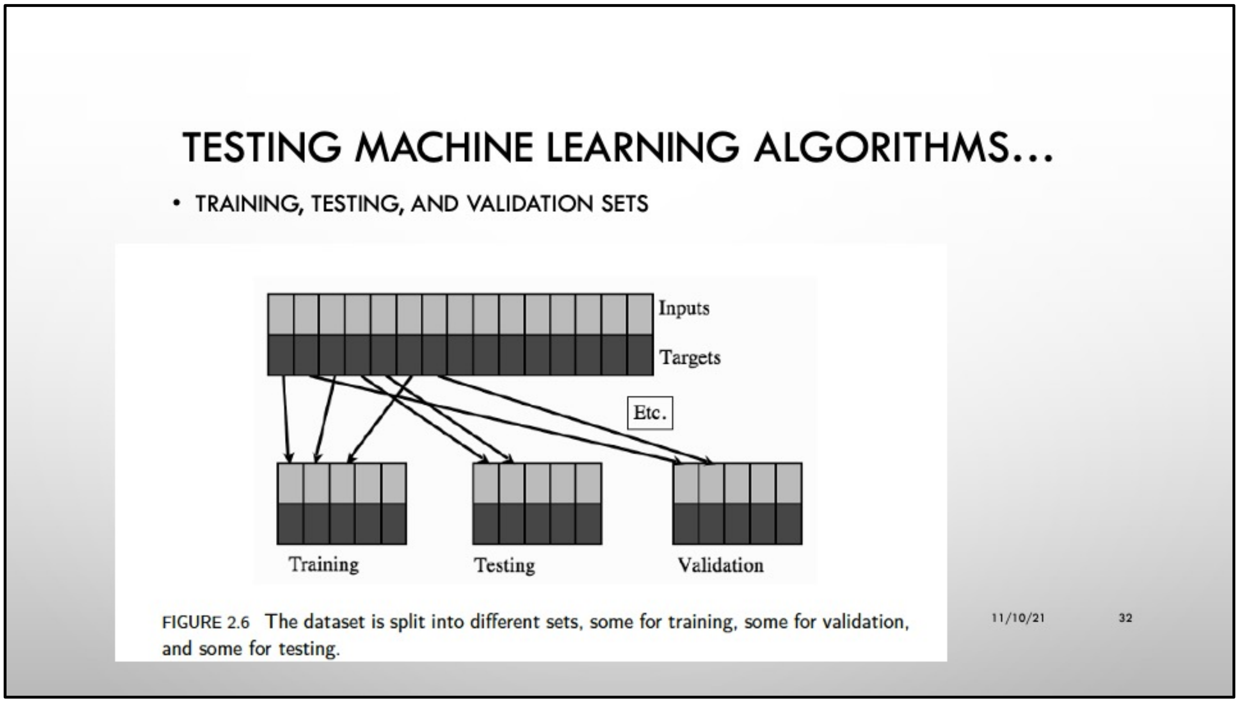
- If we train for too long, then we will overfit the data, which means that we have learnt about the noise and inaccuracies in the data as well as the actual function.
- Therefore, the model that we learn will be much too complicated and won't be able to generalise.



TESTING MACHINE LEARNING ALGORITHMS…

OVERFITTING

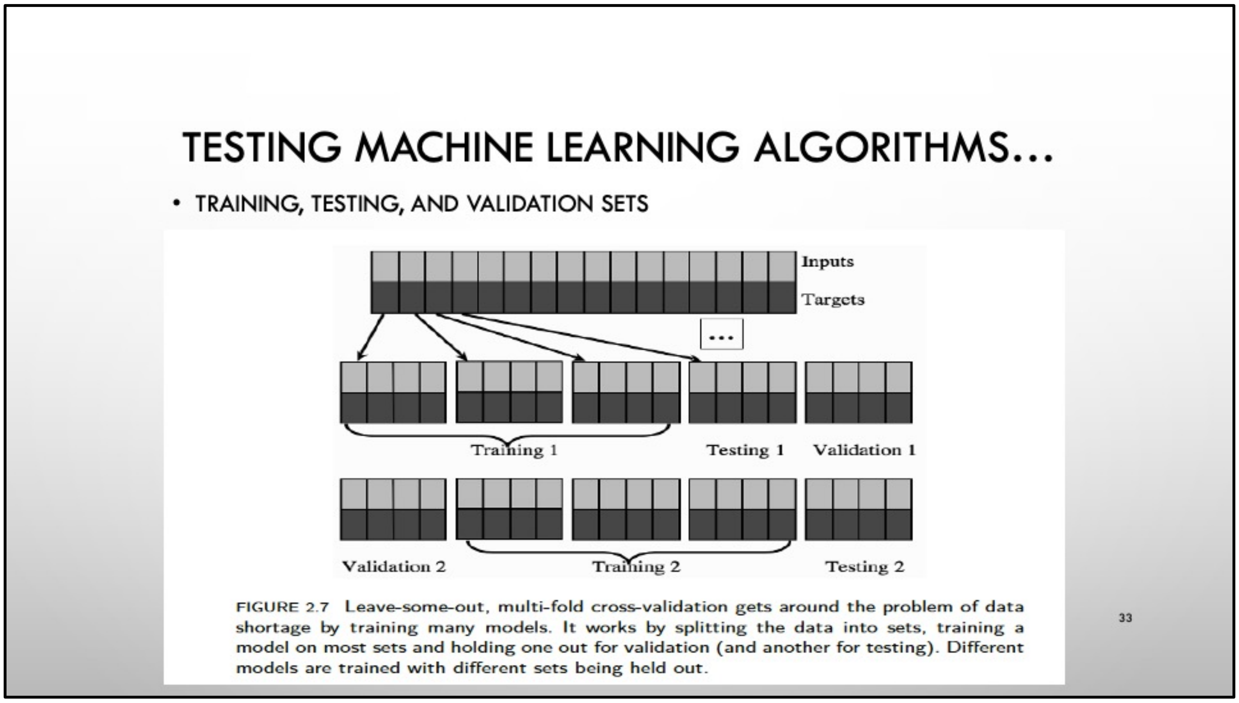
FIGURE 2.5 The effect of overfitting is that rather than finding the generating function (as shown on the left), the neural network matches the inputs perfectly, including the noise in them (on the right). This reduces the generalisation capabilities of the network.



TESTING MACHINE LEARNING ALGORITHMS…

- TRAINING, TESTING, AND VALIDATION SETS

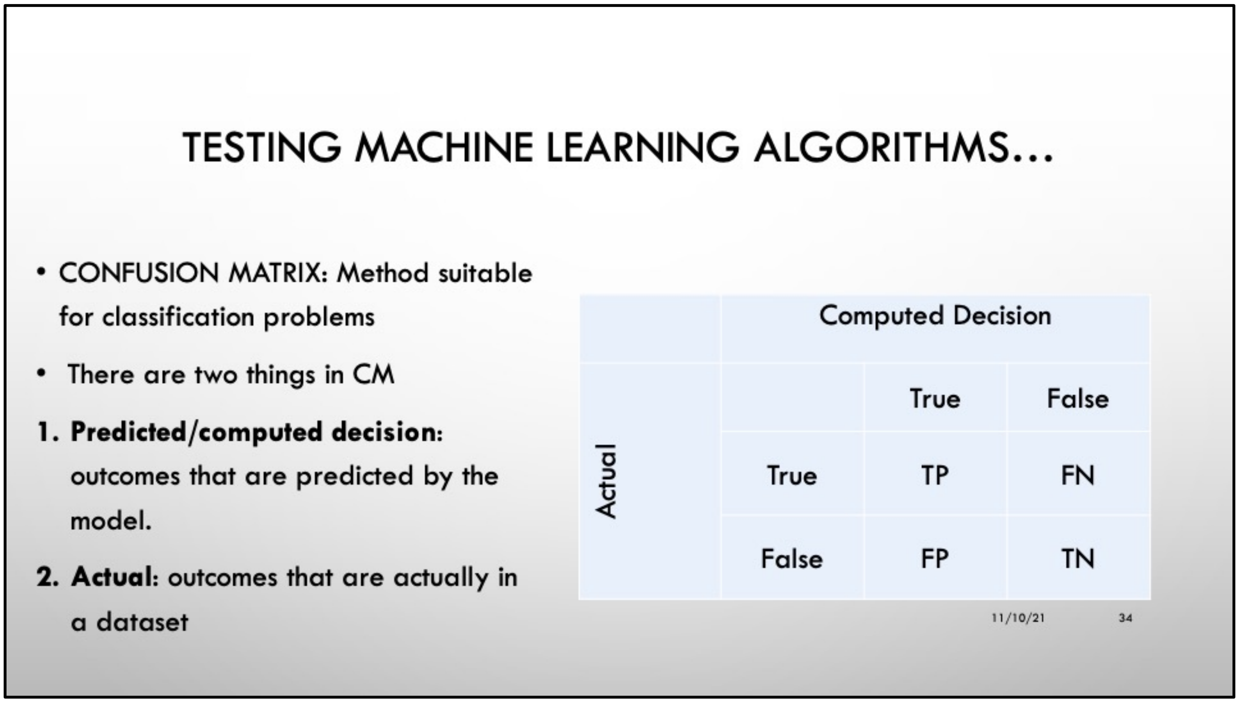
FIGURE 2.6 The dataset is split into different sets, some for training, some for validation, and some for testing.



TESTING MACHINE LEARNING ALGORITHMS…

- TRAINING, TESTING, AND VALIDATION SETS

FIGURE 2.7 Leave-some-out, multi-fold cross-validation gets around the problem of data shortage by training many models. It works by splitting the data into sets, training a model on most sets and holding one out for validation (and another for testing). Different models are trained with different sets being held out.

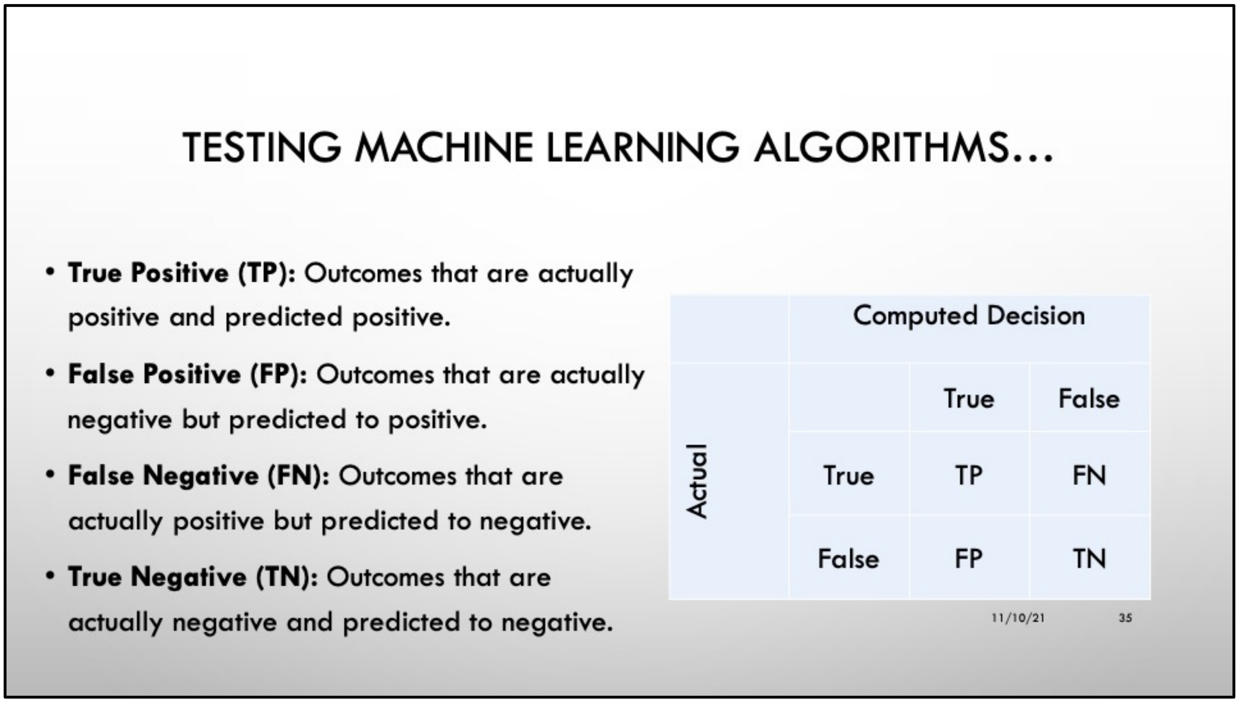


TESTING MACHINE LEARNING ALGORITHMS…

- CONFUSION MATRIX: Method suitable for classification problems
- There are two things in CM

1. **Predicted/computed decision**: outcomes that are predicted by the model.
2. **Actual**: outcomes that are actually in a dataset

		Computed Decision	
		True	False
Actual	True	TP	FN
	False	FP	TN



TESTING MACHINE LEARNING ALGORITHMS…

- **True Positive (TP):** Outcomes that are actually positive and predicted positive.
- **False Positive (FP):** Outcomes that are actually negative but predicted to positive.
- **False Negative (FN):** Outcomes that are actually positive but predicted to negative.
- **True Negative (TN):** Outcomes that are actually negative and predicted to negative.

		Computed Decision	
		True	False
Actual	True	TP	FN
	False	FP	TN

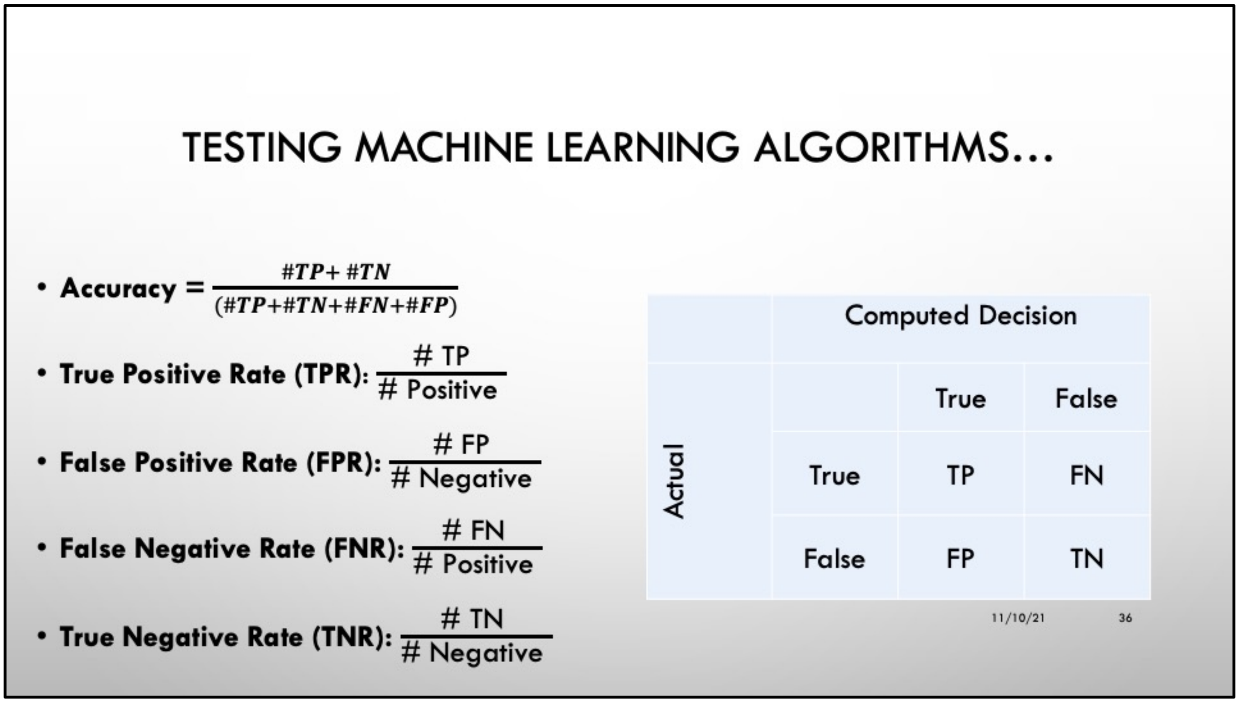
11/10/21 35

True positives .TP/: These refer to the positive tuples that were correctly labeled by the classifier. Let TP be the number of true positives.

True negatives .TN/: These are the negative tuples that were correctly labeled by the classifier. Let TN be the number of true negatives.

False positives .FP/: These are the negative tuples that were incorrectly labeled as positive (e.g., tuples of class buys computer = no for which the classifier predicted buys computer = yes). Let FP be the number of false positives.

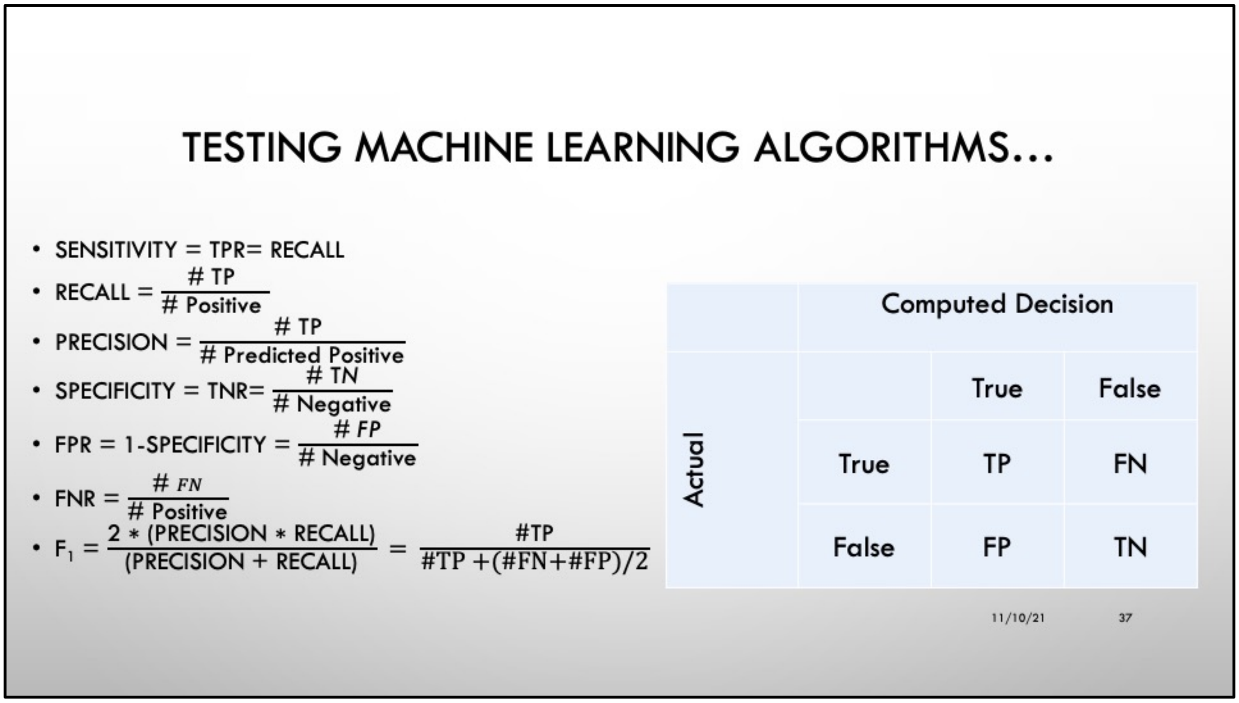
False negatives .FN/: These are the positive tuples that were mislabeled as negative (e.g., tuples of class buys computer= yes for which the classifier predicted buys computer = no). Let FN be the number of false negatives.



TESTING MACHINE LEARNING ALGORITHMS…

- **Accuracy =** $\frac{\#TP+\#TN}{(\#TP+\#TN+\#FN+\#FP)}$
- **True Positive Rate (TPR):** $\frac{\#\text{ TP}}{\#\text{ Positive}}$
- **False Positive Rate (FPR):** $\frac{\#\text{ FP}}{\#\text{ Negative}}$
- **False Negative Rate (FNR):** $\frac{\#\text{ FN}}{\#\text{ Positive}}$
- **True Negative Rate (TNR):** $\frac{\#\text{ TN}}{\#\text{ Negative}}$

		Computed Decision	
Actual		True	False
	True	TP	FN
	False	FP	TN



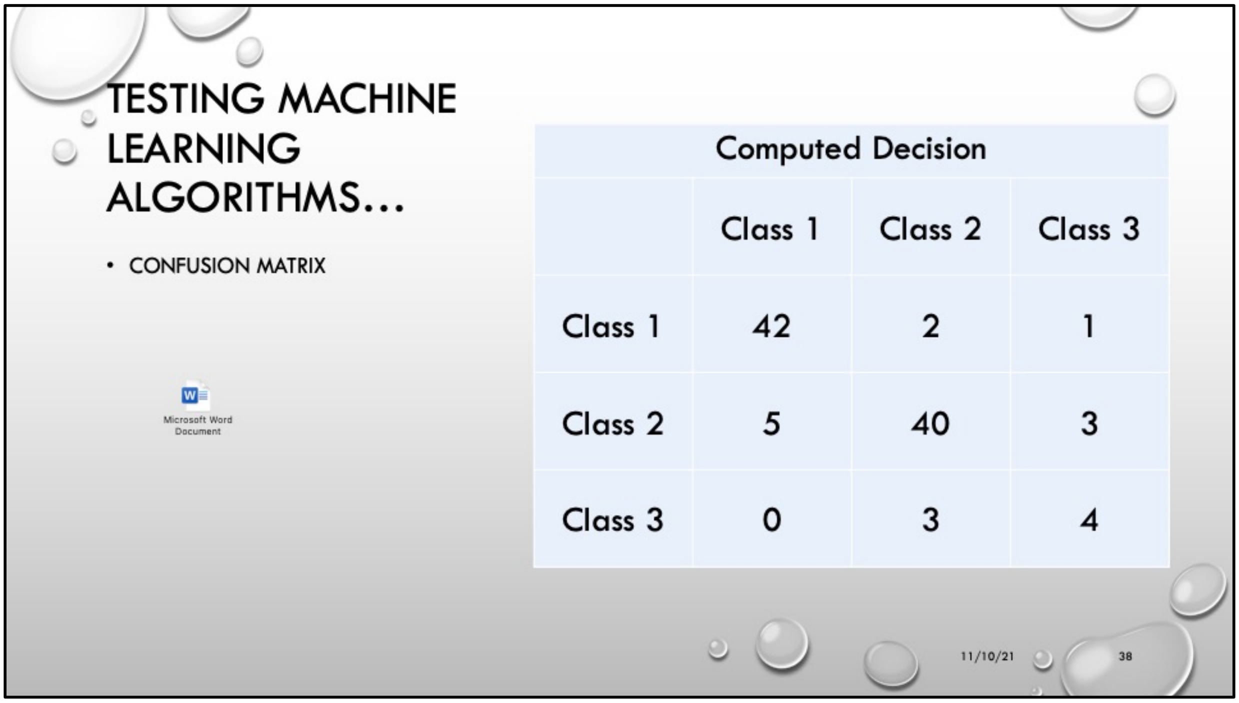
TESTING MACHINE LEARNING ALGORITHMS…

- SENSITIVITY = TPR= RECALL
- RECALL = $\frac{\text{\# TP}}{\text{\# Positive}}$
- PRECISION = $\frac{\text{\# TP}}{\text{\# Predicted Positive}}$
- SPECIFICITY = TNR= $\frac{\text{\# TN}}{\text{\# Negative}}$
- FPR = 1-SPECIFICITY = $\frac{\# \mathit{FP}}{\text{\# Negative}}$
- FNR = $\frac{\# \mathit{FN}}{\text{\# Positive}}$
- $F_1 = \frac{2 * (\text{PRECISION} * \text{RECALL})}{(\text{PRECISION} + \text{RECALL})} = \frac{\#\text{TP}}{\#\text{TP} + (\#\text{FN} + \#\text{FP})/2}$

	Computed Decision		
		True	False
Actual	True	TP	FN
	False	FP	TN

11/10/21 37

Reference : Han and Kamber   Pg: 365



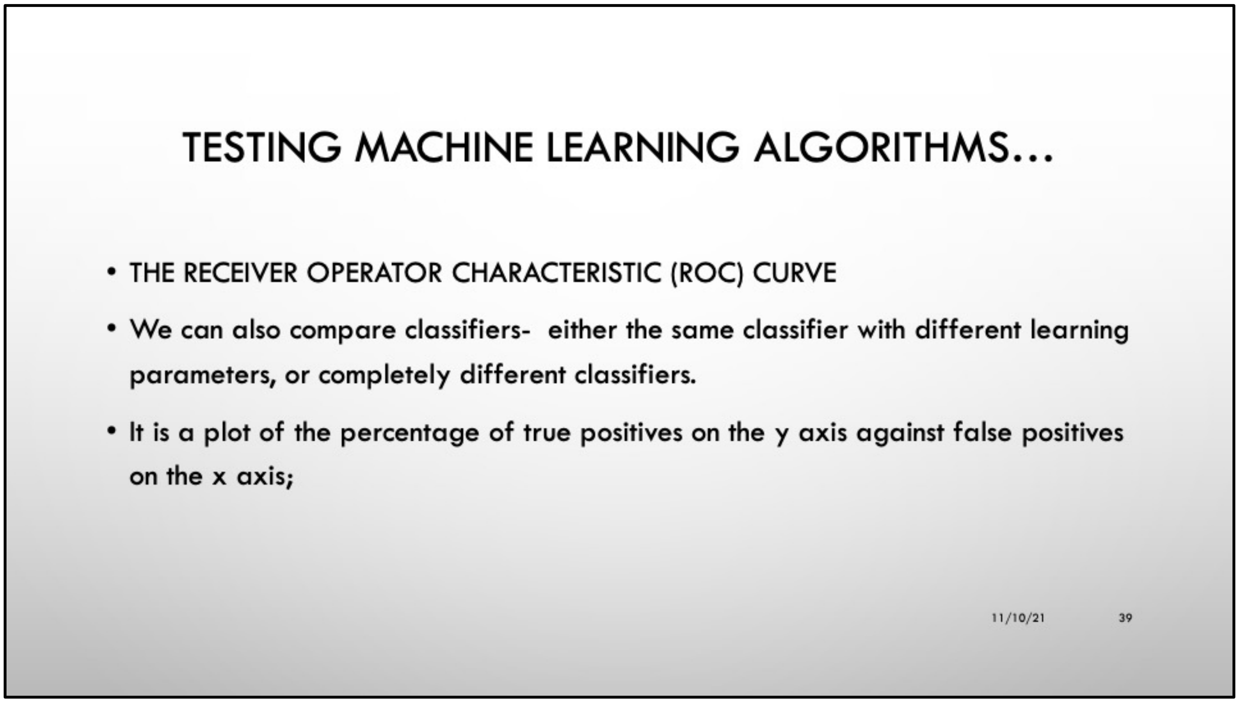
TESTING MACHINE LEARNING ALGORITHMS…

- CONFUSION MATRIX

Microsoft Word Document

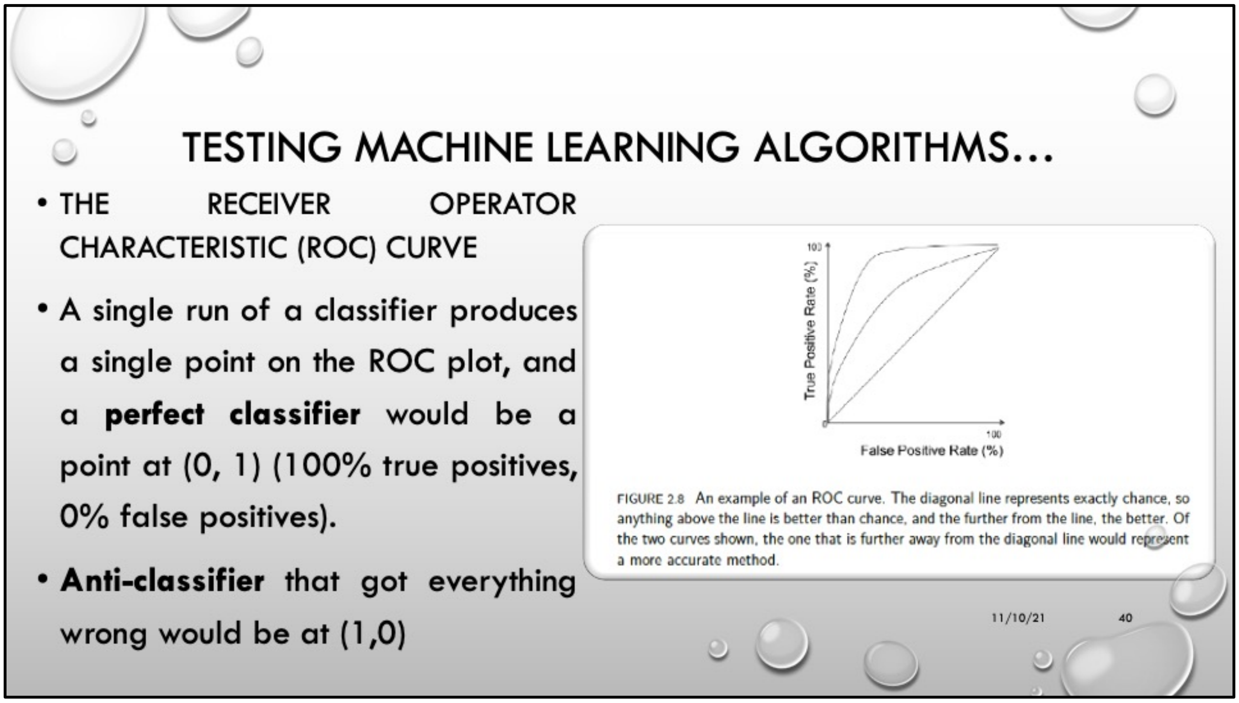
	Computed Decision		
	Class 1	Class 2	Class 3
Class 1	42	2	1
Class 2	5	40	3
Class 3	0	3	4

11/10/21 38



TESTING MACHINE LEARNING ALGORITHMS…

- THE RECEIVER OPERATOR CHARACTERISTIC (ROC) CURVE
- We can also compare classifiers- either the same classifier with different learning parameters, or completely different classifiers.
- It is a plot of the percentage of true positives on the y axis against false positives on the x axis;



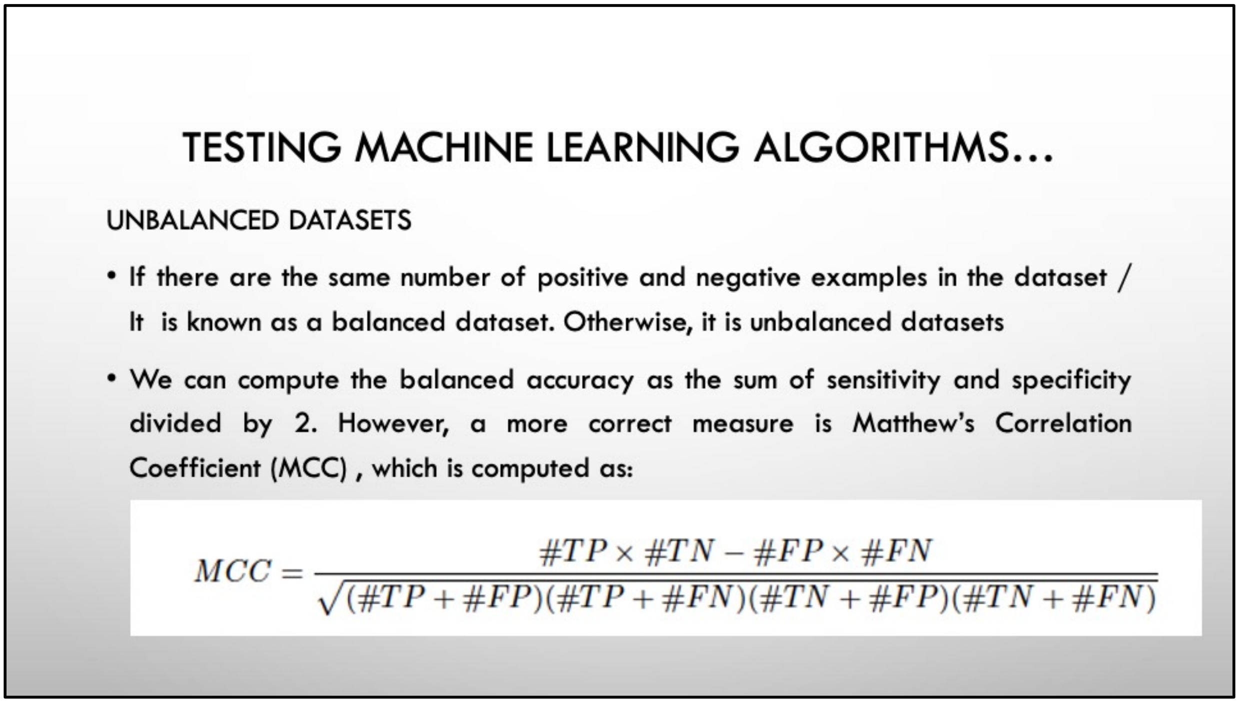
TESTING MACHINE LEARNING ALGORITHMS…

- THE RECEIVER OPERATOR CHARACTERISTIC (ROC) CURVE
- A single run of a classifier produces a single point on the ROC plot, and a **perfect classifier** would be a point at (0, 1) (100% true positives, 0% false positives).
- **Anti-classifier** that got everything wrong would be at (1,0)

FIGURE 2.8 An example of an ROC curve. The diagonal line represents exactly chance, so anything above the line is better than chance, and the further from the line, the better. Of the two curves shown, the one that is further away from the diagonal line would represent a more accurate method.

11/10/21 40

In order to compare classifiers, or choices of parameters settings for the same classifier, you could just compute the point that is furthest from the ‘chance’ line along the diagonal.
However, it is normal to compute the area under the curve (AUC) instead. If you only have one point for each classifier, the curve is the trapezoid that runs from (0,0) up to the point and then from there to (1,1). If there are more points (based on more runs of the classifier, such as trained and/or tested on different datasets), then they are just included in order along the diagonal line.

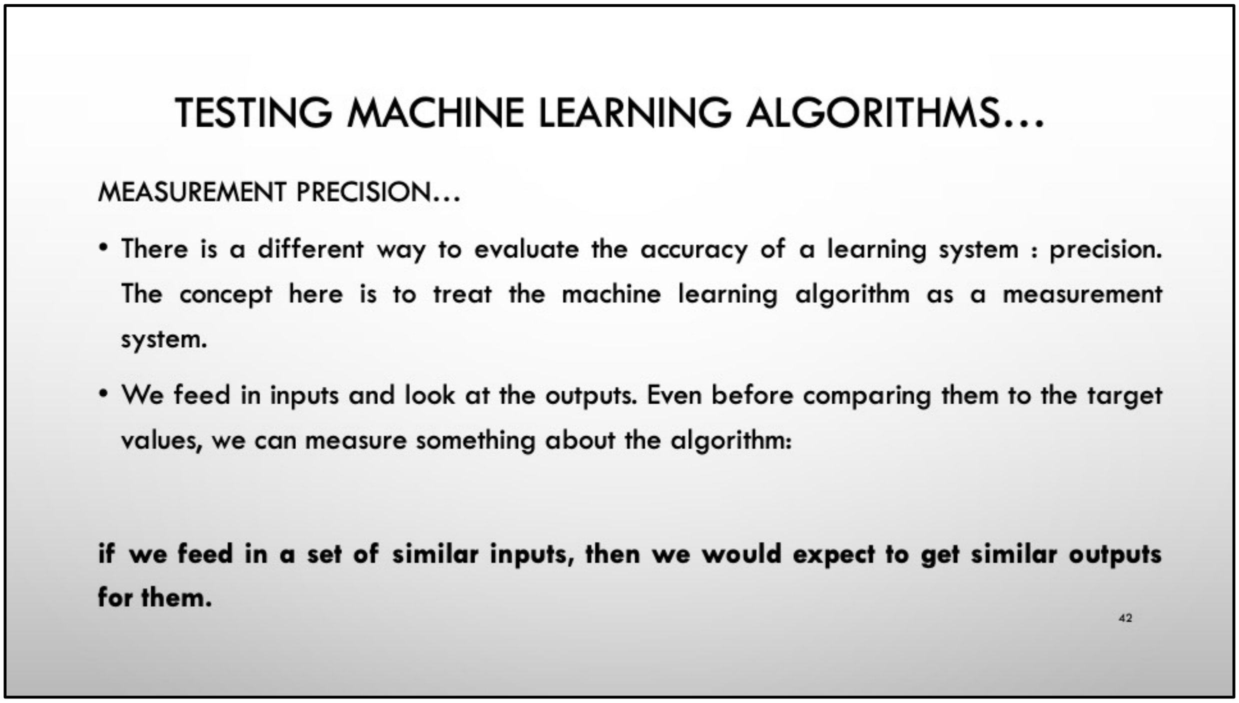


TESTING MACHINE LEARNING ALGORITHMS…

UNBALANCED DATASETS

- If there are the same number of positive and negative examples in the dataset / It is known as a balanced dataset. Otherwise, it is unbalanced datasets
- We can compute the balanced accuracy as the sum of sensitivity and specificity divided by 2. However, a more correct measure is Matthew’s Correlation Coefficient (MCC) , which is computed as:

$$MCC = \frac{\#TP \times \#TN - \#FP \times \#FN}{\sqrt{(\#TP + \#FP)(\#TP + \#FN)(\#TN + \#FP)(\#TN + \#FN)}}$$



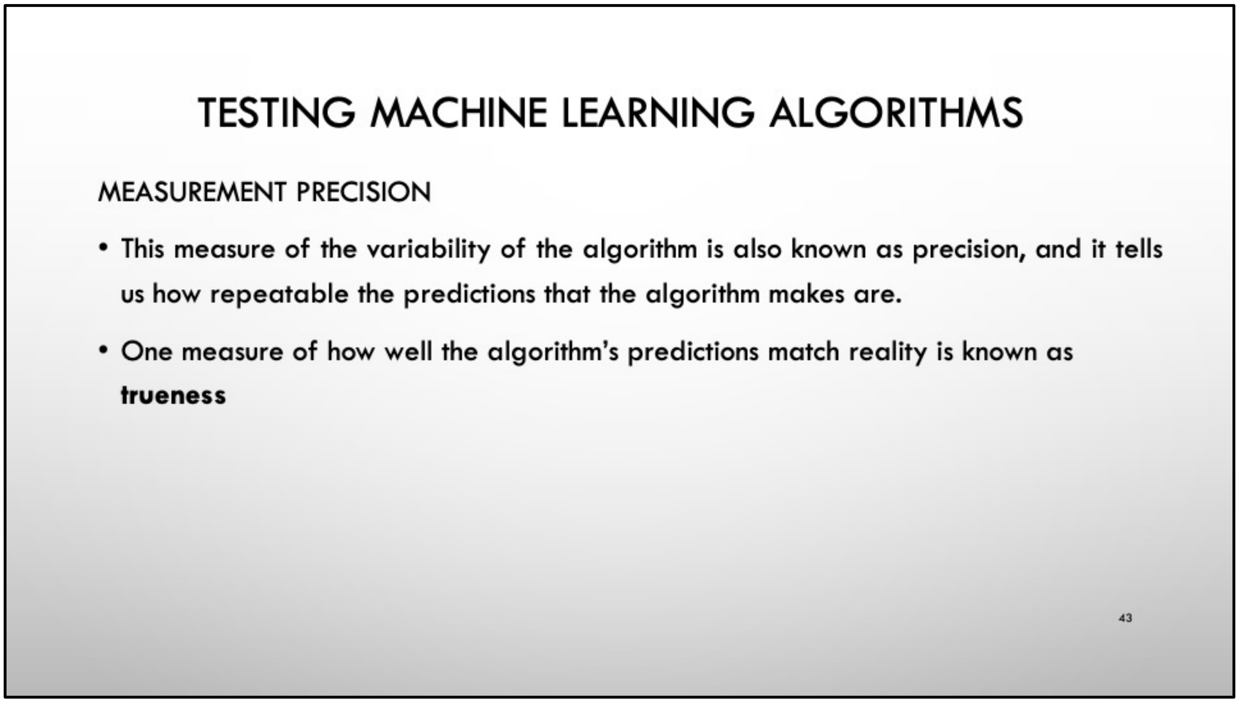
TESTING MACHINE LEARNING ALGORITHMS…

MEASUREMENT PRECISION…

- There is a different way to evaluate the accuracy of a learning system : precision. The concept here is to treat the machine learning algorithm as a measurement system.
- We feed in inputs and look at the outputs. Even before comparing them to the target values, we can measure something about the algorithm:

if we feed in a set of similar inputs, then we would expect to get similar outputs for them.

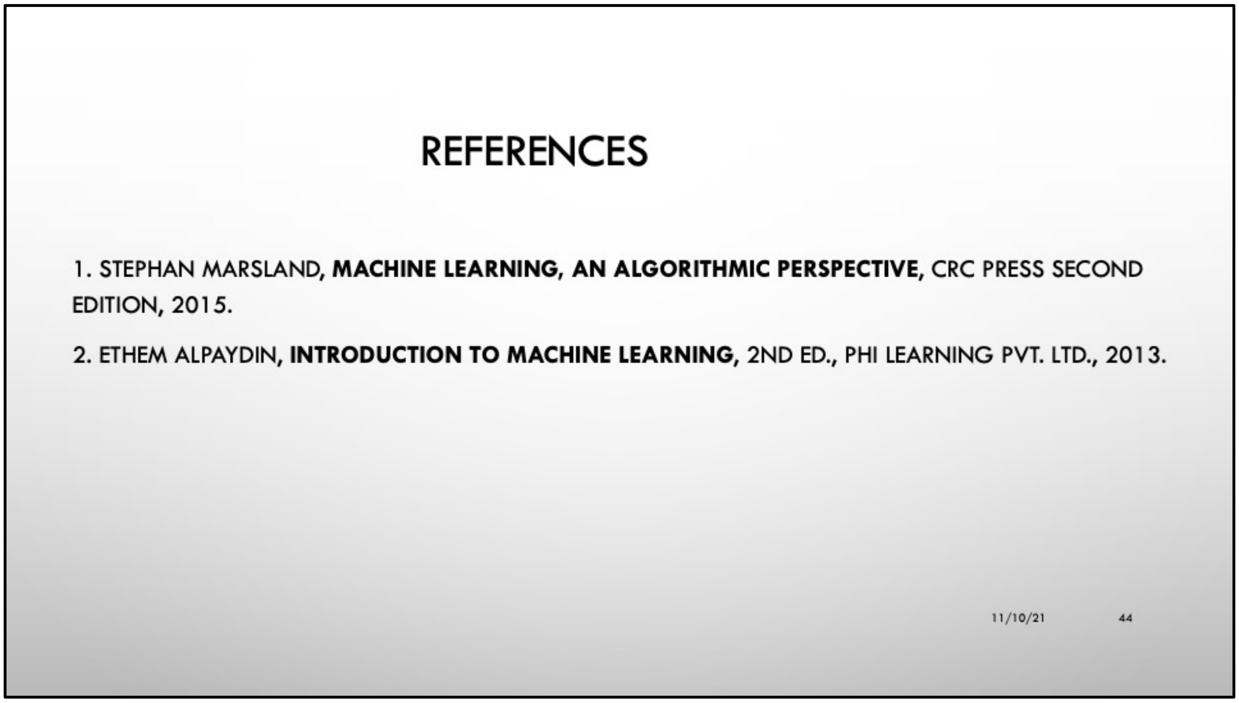
42



TESTING MACHINE LEARNING ALGORITHMS

MEASUREMENT PRECISION

- This measure of the variability of the algorithm is also known as precision, and it tells us how repeatable the predictions that the algorithm makes are.
- One measure of how well the algorithm's predictions match reality is known as **trueness**



REFERENCES

1. STEPHAN MARSLAND, **MACHINE LEARNING, AN ALGORITHMIC PERSPECTIVE**, CRC PRESS SECOND EDITION, 2015.

2. ETHEM ALPAYDIN, **INTRODUCTION TO MACHINE LEARNING**, 2ND ED., PHI LEARNING PVT. LTD., 2013.