INSTITUTO UNIVERSITARIO
SIANI
INGENIERIA COMPUTACIONAL

Fundamentals of ML & DL
CASA course
(12/10/2018)

D. Freire-Obregón, PhD.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Plan for TS2

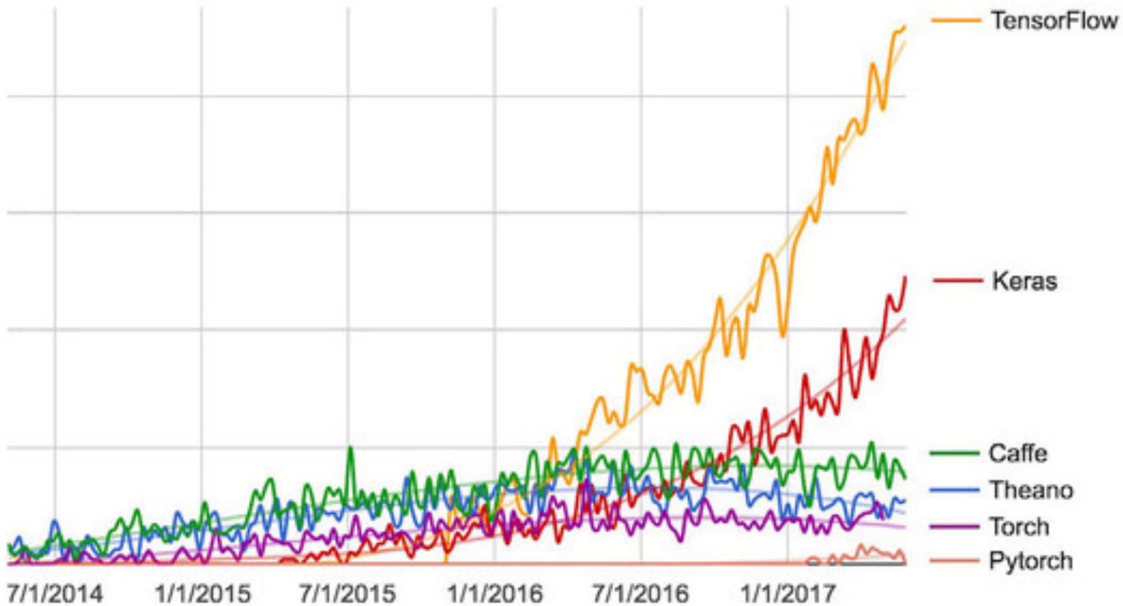
- Frameworks evolution
- Learning representations from data
- Branches of ML
- Overfitting & Underfitting
- Evaluating ML models
- Universal workflow of ML

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

BIP LAB

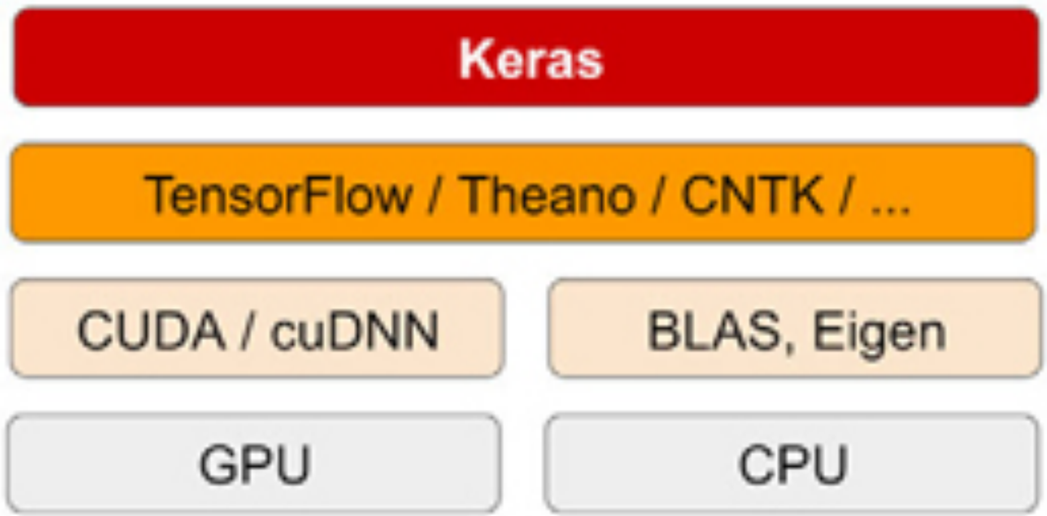
Frameworks evolution



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Frameworks evolution

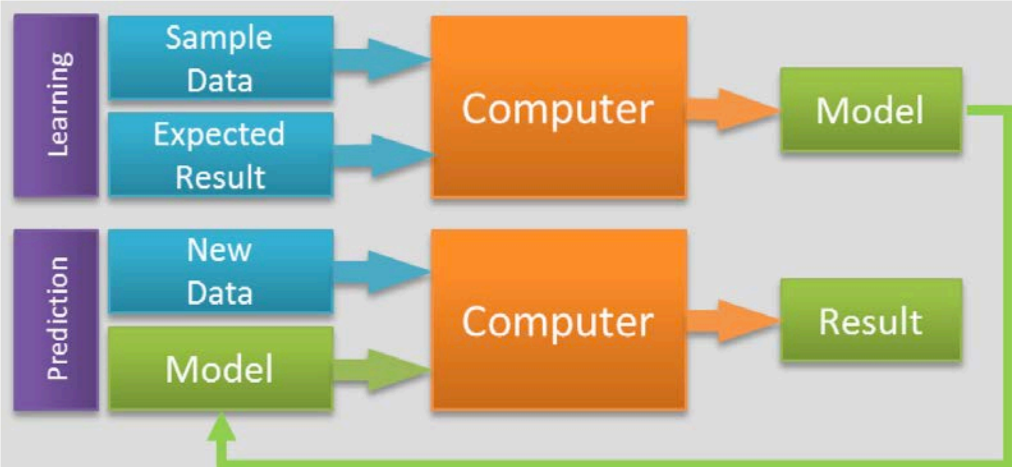


Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

CV & ML

Traditional modeling - CV (feat.ext.)+ML(class.):

Deep learning:



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Learning representations from data

To do ML we need:

1. Input data points.
   - If the task is image tagging, pictures are needed.
   - If the task is speech recog., people speaking samples are needed.
2. Examples of expected output.
   - In the image task, samples already tagged.
   - In the speech recog. Task, transcriptions of the sound files.

Learning representations from data

To do ML we need:

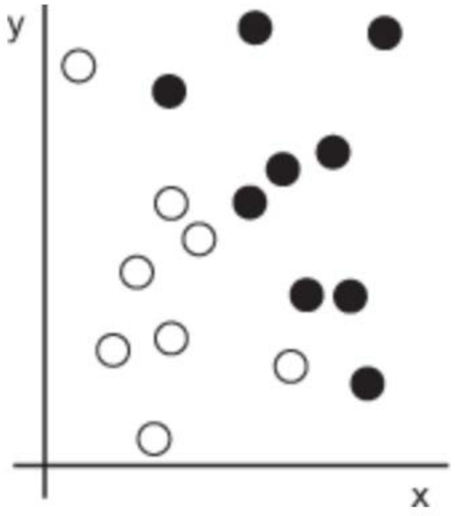
3. A measurement of how good is the algorithm working.
   - Distance between current output and expected output.
   - This measurement is a feedback signal to adjust the algorithm → This adjustment is known as **learning**.

Problem to solve in ML & DL:

To learn useful representations of the input data

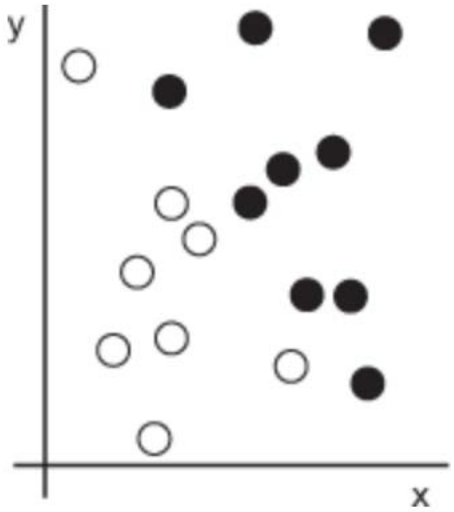
INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Learning representations from data



Lets say we want to develop an algorithm that take coordinates (x,y) of a point and output whether that point is likely to be black or white.

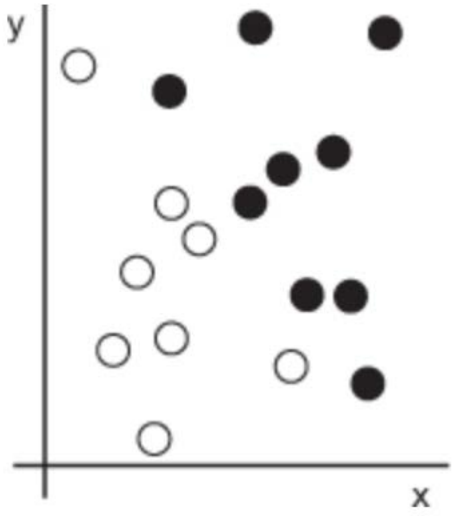
Learning representations from data



Inputs?
Expected outputs?
A measurement?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Learning representations from data



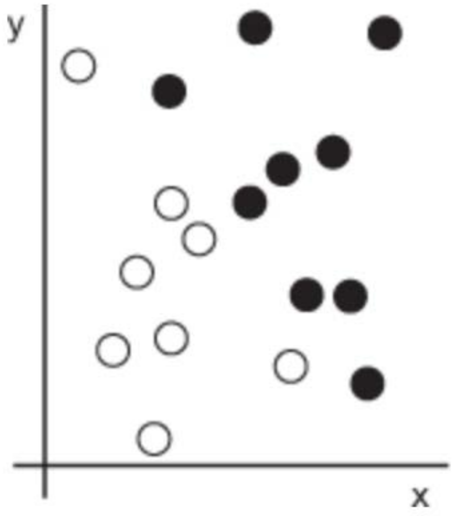
Inputs? Points coordinates.

Expected outputs?

A measurement?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Learning representations from data



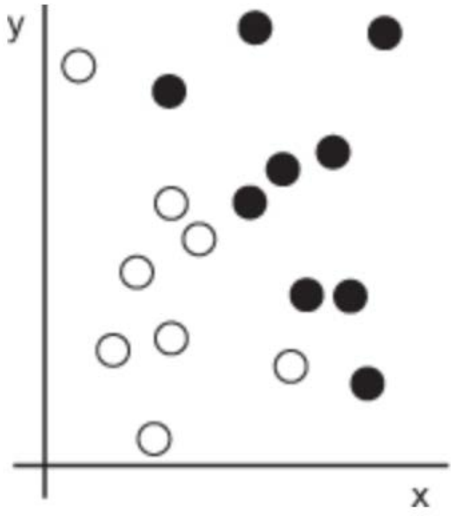
Inputs? Points coordinates.

Expected outputs? Color of our points.

A measurement?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Learning representations from data



Inputs? Points coordinates

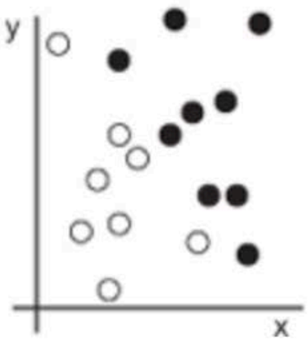
Expected outputs? Color of our points

A measurement? Percentage of points that are being corrected classified

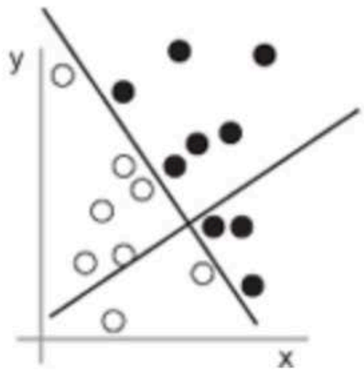
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Learning representations from data

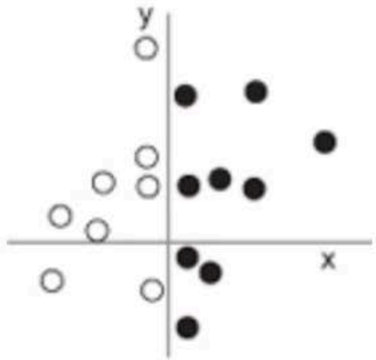
Raw data



Coordinate change



Better representation



A new coordinate systems provides a new representation of our data

With this representation:

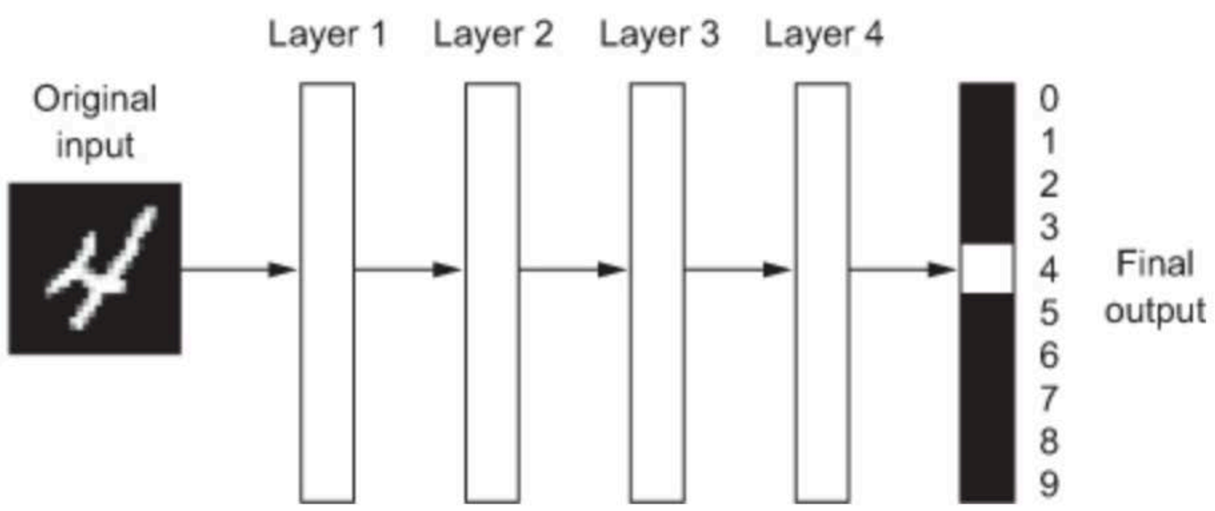
Black points are such that x>0

White points are such that x<0

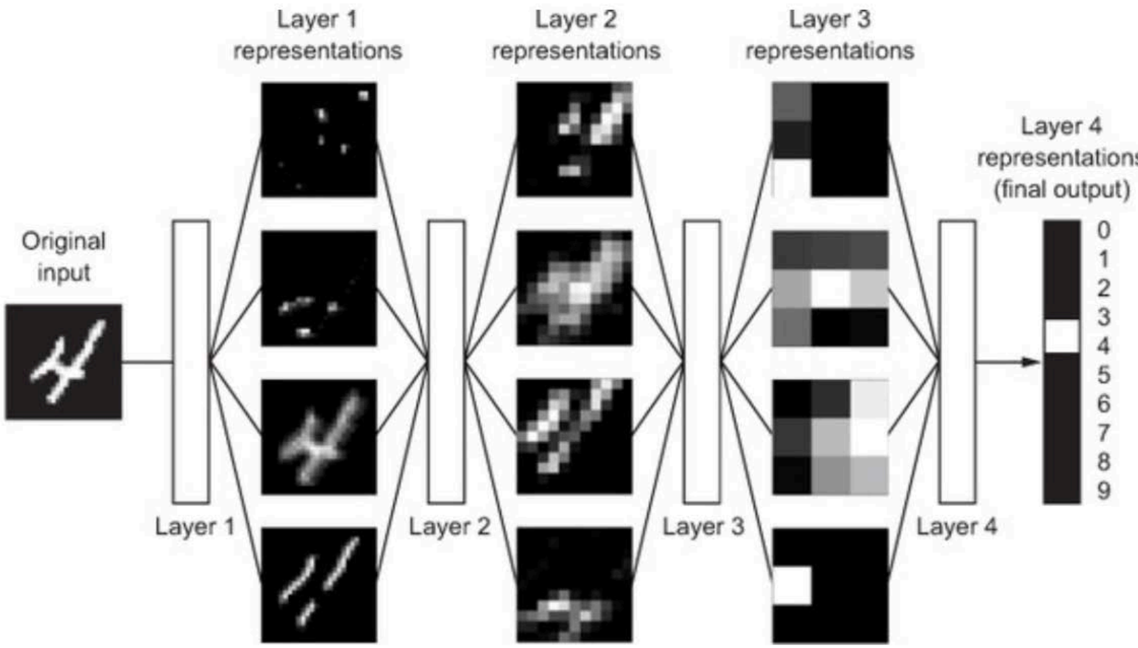
The new representation solves the classification problem.

Deep Learning

Neural network term is a reference to neurobiology.
But DL models are not models in the brain. No evidence.



Deep Learning

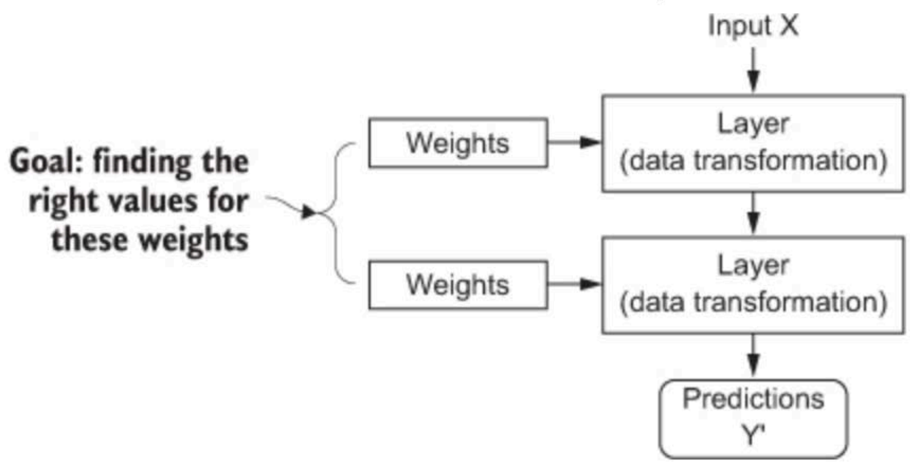


DL is a multistage way to learn data representations.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Deep Learning

The specification of what a layer does to its input data is stored in the layer's weights.



The transformation implemented by a layer is parameterized by its weights.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Deep Learning

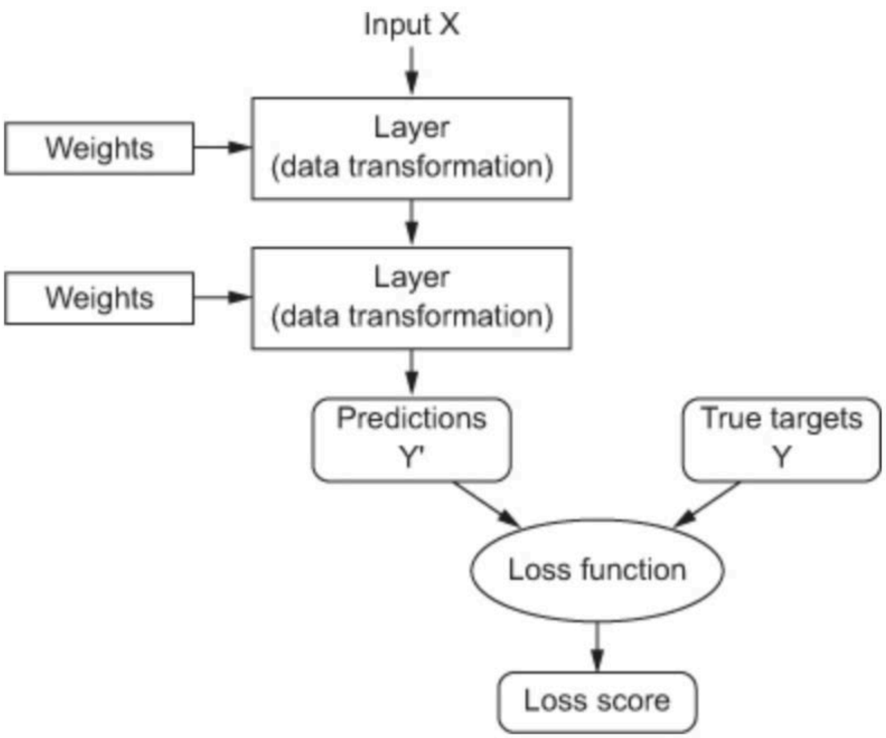
Problem: A deep neural network can contain tens of millions of parameters.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Deep Learning

To control something…we need to observe it.
To control de output of a NN we need to measure how far is it from the target.
Solution: **loss function**

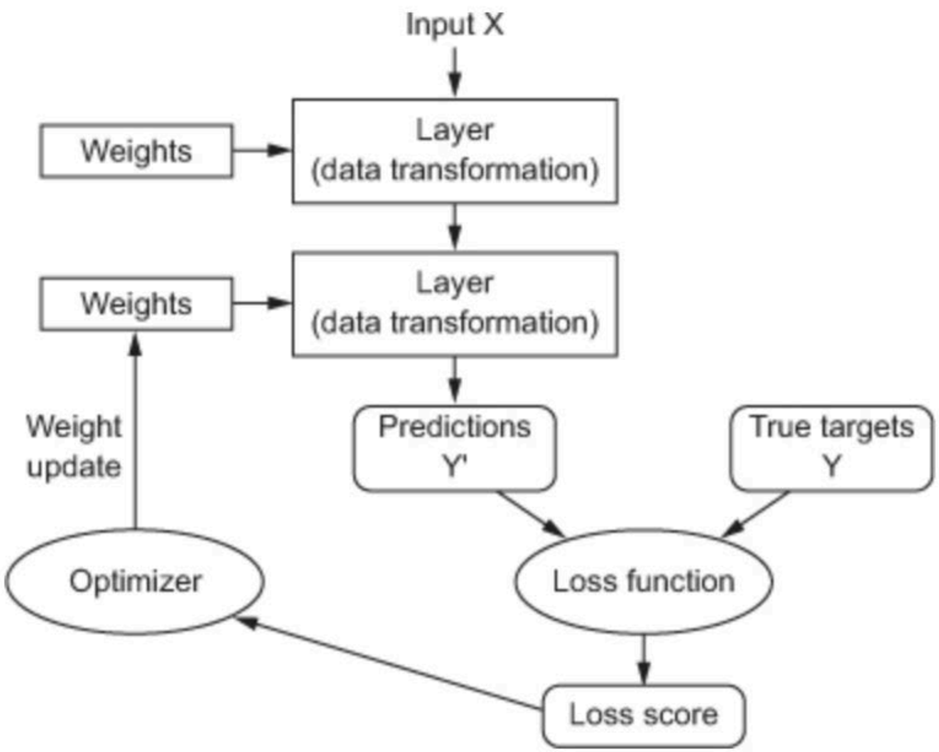
Takes NN predictions and true target and computes a distance score



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Deep Learning

How can we use the loss function information as a feedback signal to adjust weights in a proper way (reduce the loss)?
Solution: **optimizer** which implements **backpropagation**



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Remember LS1

mymodel.compile(optimizer=‘rmsprop’, loss=‘categorical_crossentropy’, metrics=[‘acc’])

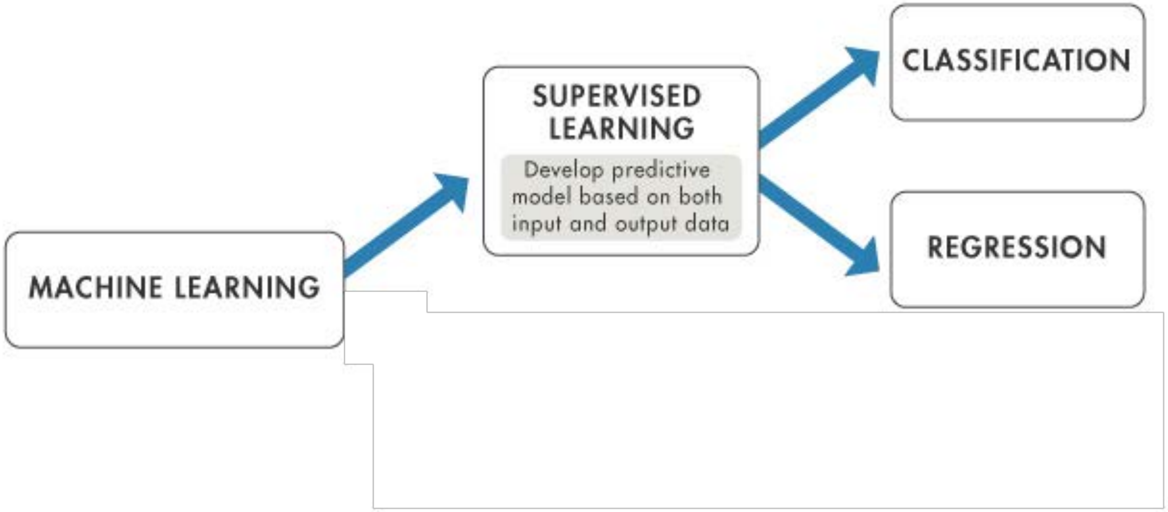
Loss function → How the network is able to measure its performance on the training data.

Optimizer→ The mechanism through which the network will update Itself based on the data it sees and its loss function

Metrics → Measure to monitor during training and testing

Branches of ML: Supervised learning

- Most common case (by far).
- Almost all applications of DL in the spotlight.
- Machine learns to map input data to known targets.



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Branches of ML: Supervised learning

More variants:

•Sequence generation → Given a picture, predict a caption describing it.

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

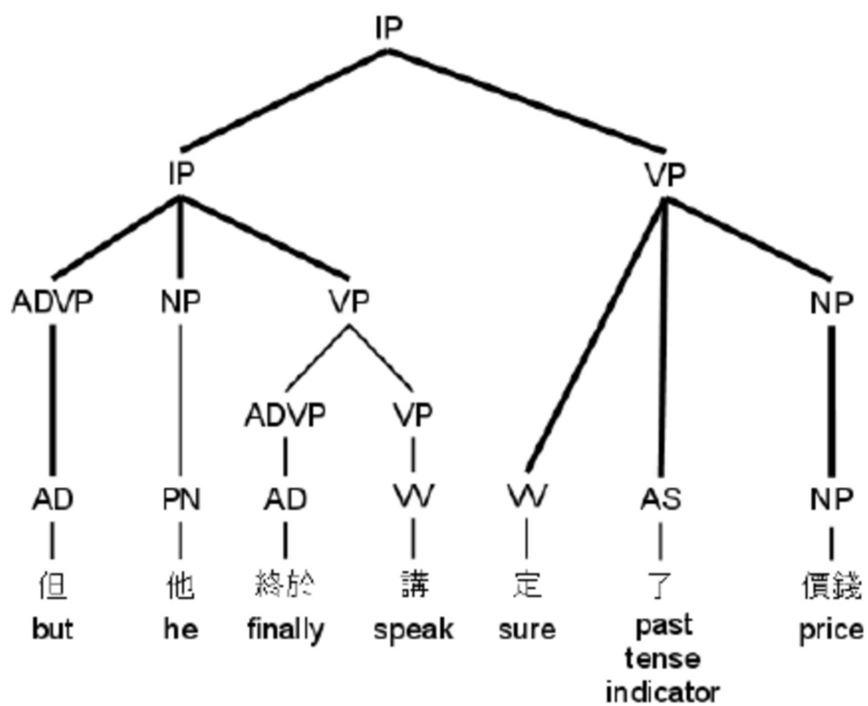
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Branches of ML: Supervised learning

More variants:

•Syntax tree prediction→ Given a sentence, predict its decomposition into a syntax tree.

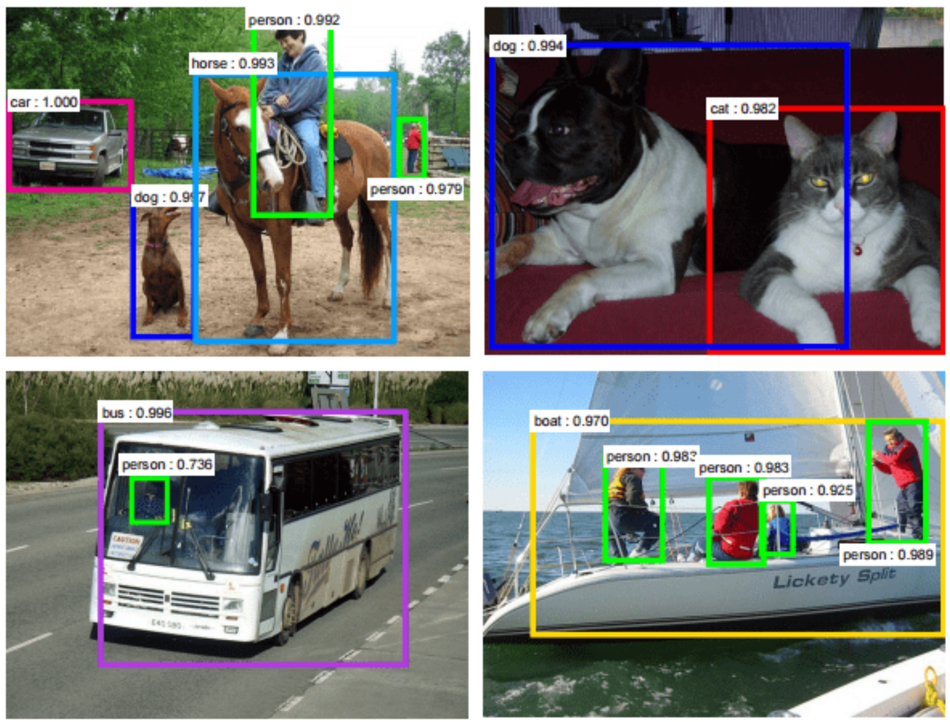


Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Branches of ML: Supervised learning

More variants:

•Object detection → Given a picture, draw a bounding box around certain objects inside the picture.

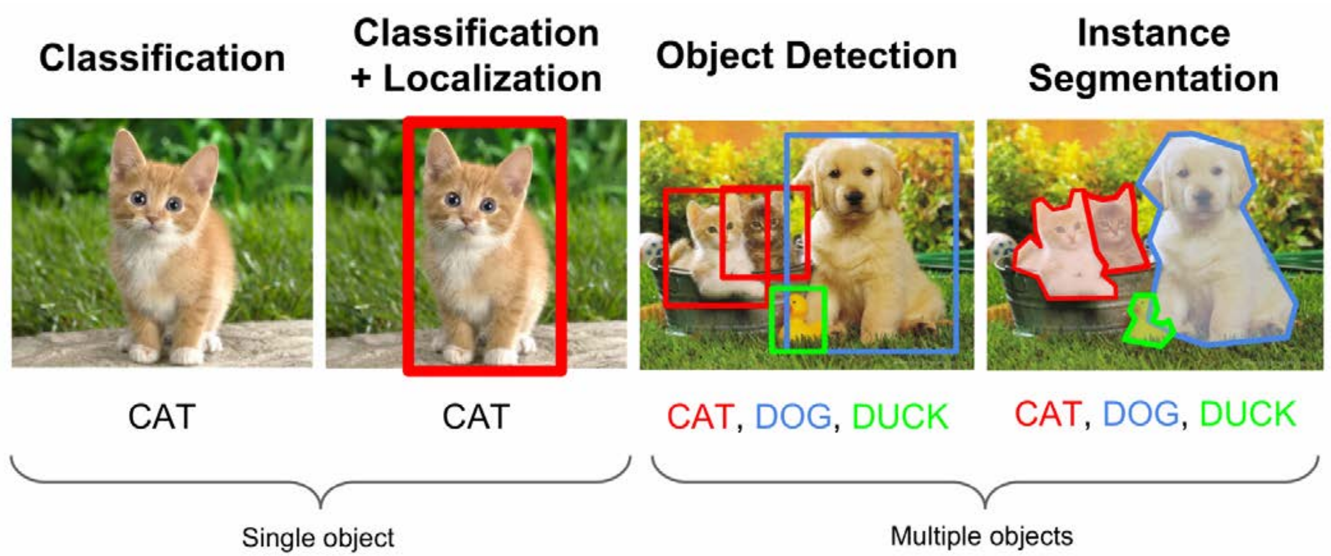


Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Branches of ML: Supervised learning

More variants:

•Image segmentation→ Given a picture, draw a pixel level mask on a specific object.



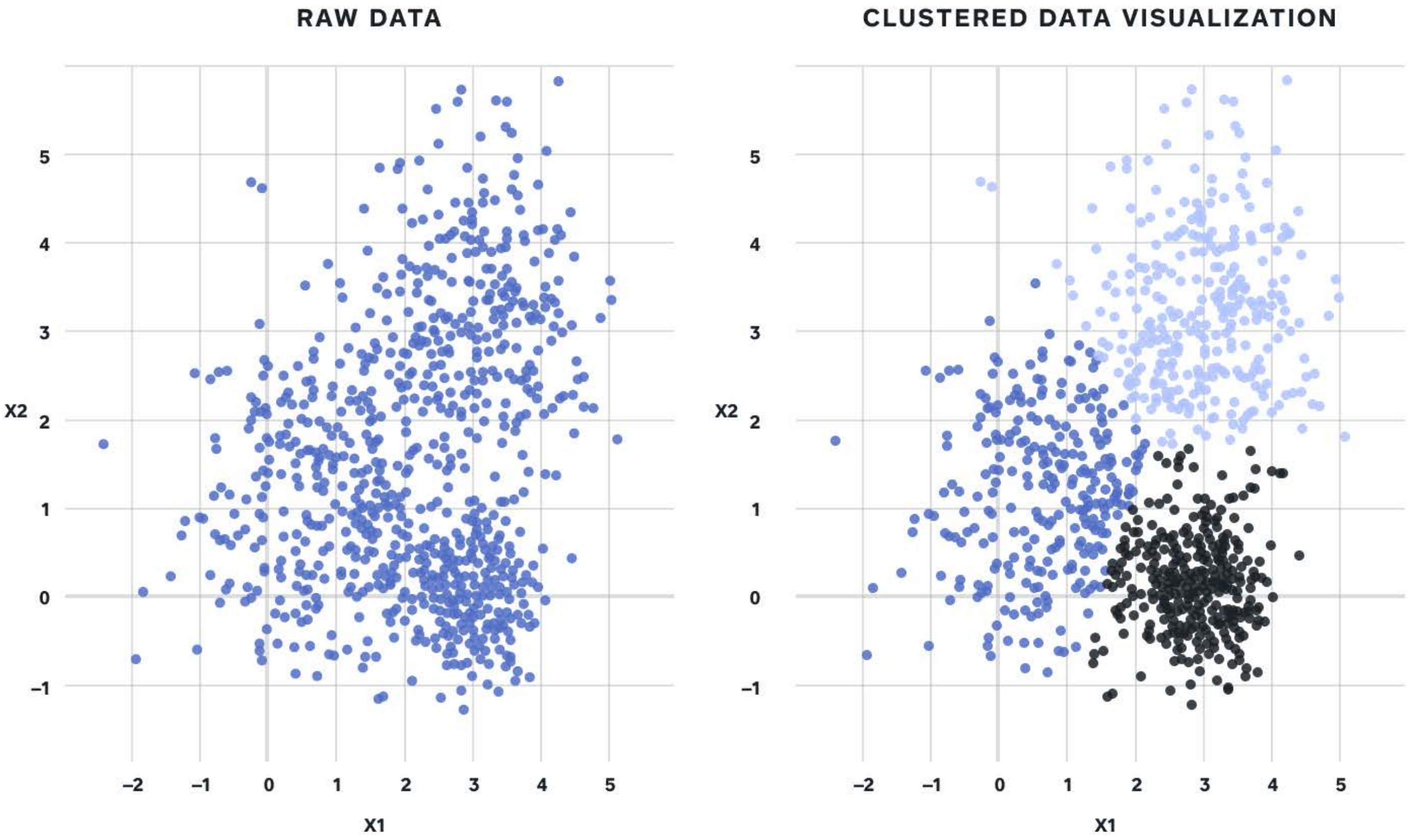
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Branches of ML: Unsupervised learning

- Find interesting transformations of the input data, without the help of any labels.
- Where? Data visualization, data compression, data denoising.
- Used in data analtytics→ Necessary step in better understanding a dataset before doing supervised ML.
- Variants:
  - Dimensionality reduction
  - Clustering

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

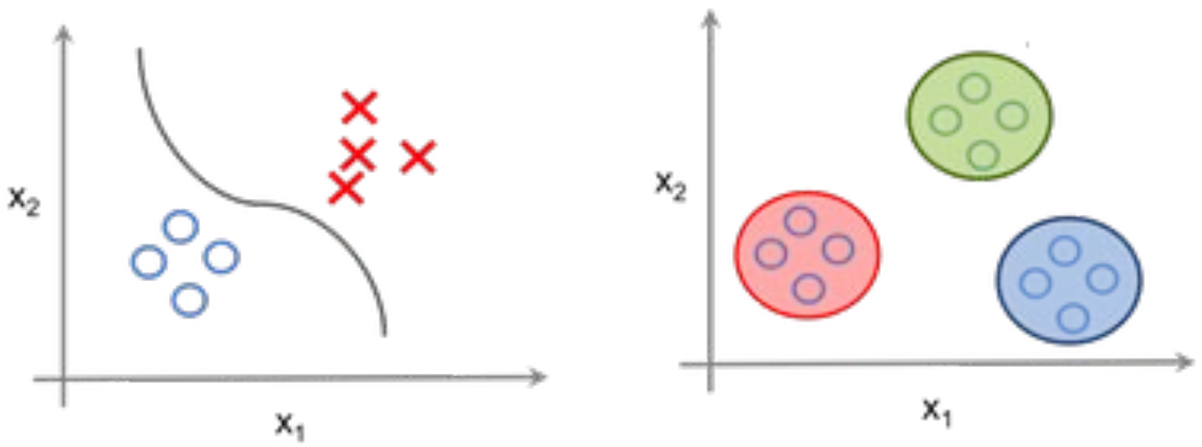
Branches of ML: Unsupervised learning



INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Branches of ML: Unsupervised learning

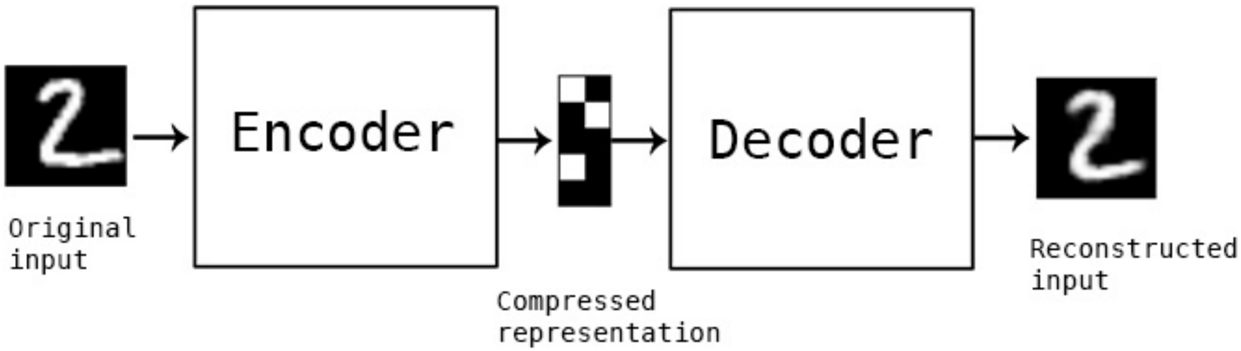
Supervised learning | Unsupervised learning



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

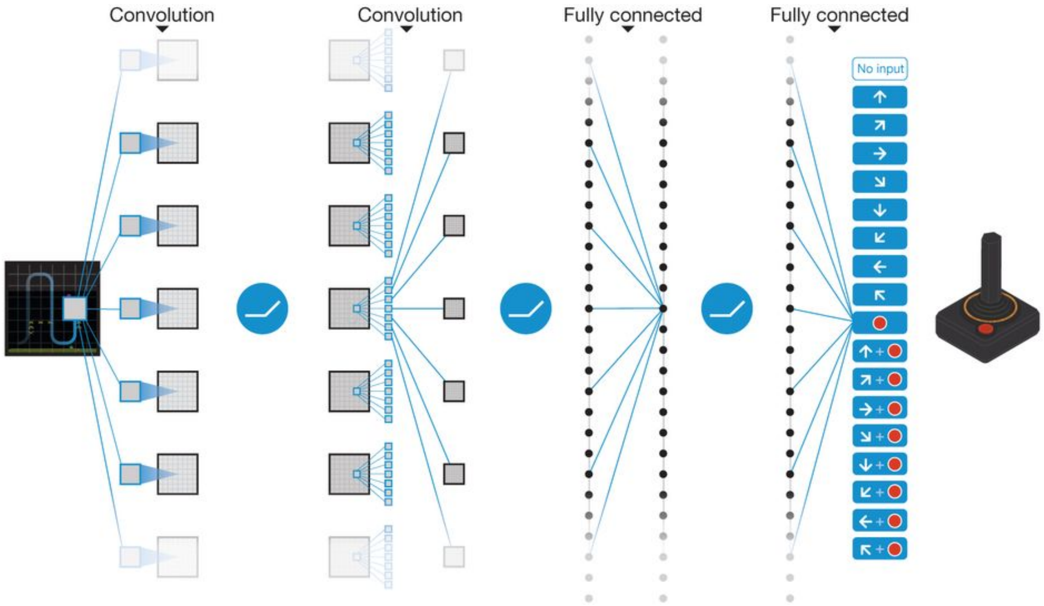
Branches of ML: Self-supervised learning

- Supervised learning without human-annotated labels.
- Autoencoders → Same input is the target.



Branches of ML: Reinforcement learning

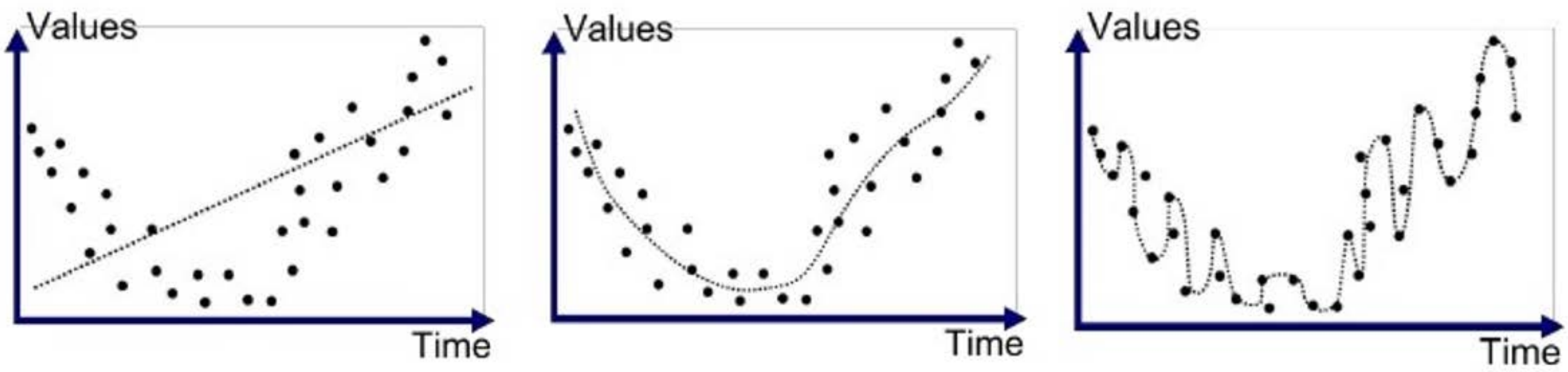
- An agent receives information about its environment and learns to choose actions that maximize some reward. https://www.youtube.com/watch?v=V1eYniJ0Rnk



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Overfitting & Underfitting

Generalization is the key



Underfitted Good Fit/Robust Overfitted

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

INSTITUTO UNIVERSITARIO
SIANI
INGENIERIA COMPUTACIONAL

Overfitting & Underfitting

Among others..

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Overfitting & Underfitting

Underfitting

Milan!

Tottenham!

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Overfitting & Underfitting

Overfitting

New test samples..

Madrid!

Madrid!

Sassuolo!

Generalization → To perform well on never-before-seen data.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

So far, we have splitted datasets into training and test samples.

Why dont we evaluate the model on the train set?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

Why dont we evaluate the model on the train set?

Female!!!

Model overfits quickly.
No generalization

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

Awesome…then we will have two sets, training and test sets.

Eeeem….not so fast.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

If we train and test with the same dataset → Overfitting
Weights are tunned based on the same data…no learning.

When we create a model…we tune hyperparameters.

- Number of layers
- Size of layers (units per layer)
- Etc.

If we do this tunning based on the test…we are still overfitting.

Information leaks → Every time we tune a hyperparameter based on the test data performance, some information about the test dataset leaks into the model….we are creating our model using test data!!

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Evaluating ML models

RULE 1: Data is aways raw…preprocessing is necessary.
RULE 2: Split your dataset into training, validation and test sets.

Evaluating ML models

RULE 2: Split your dataset into training, validation and test sets.

Where is the validation set?
Extract it from the training set…20% validation - 80% training.
Be careful with data representativeness.

Training set is ordered: [0 1 2 3 4 5 6 7 8 9]

Test set….**forget about it!** Don't touch it until the end.

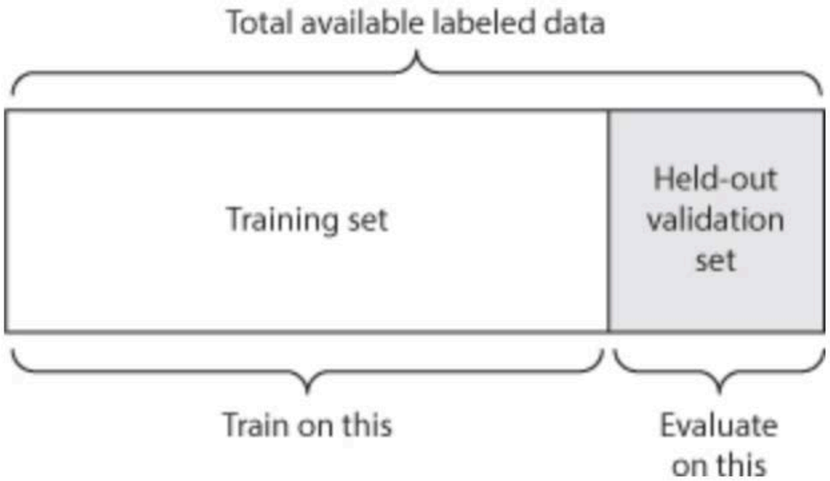
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

RULE 2: Split your dataset into training, validation and test sets.

Approaches:

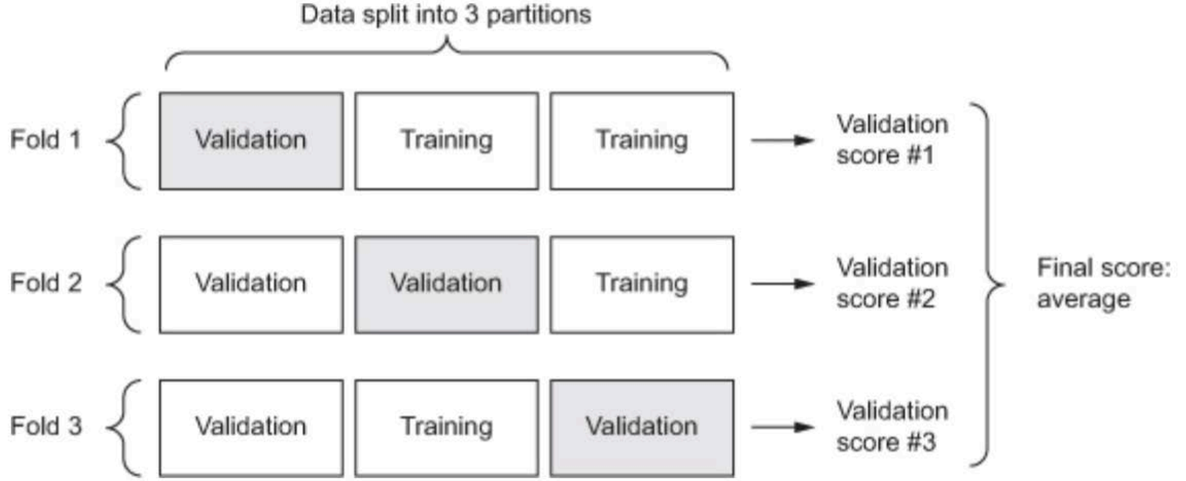
Simple hold out:



Evaluating ML models

RULE 2: Split your dataset into training, validation and test sets.

Approaches:
K-Fold Cross Validation



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Evaluating ML models

RULE 2: Split your dataset into training, validation and test sets.

Approaches:

Iterated K-Fold validation with suffling →

- When we have very little available data.
- Apply K-Fold multiple times…shuffling data every time before splitting.
- PxK models are evaluated (expensive).

Universal workflow of ML:
1. Defining the problem and assembling the datasets

Define the problem at hand:
- What will your input data be? What is your question?
  - Bear in mind: no data, no business.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

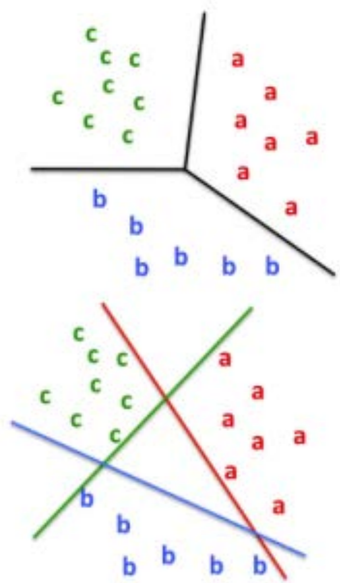
INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

Universal workflow of ML:

1. Defining the problem and assembling the datasets

Define the problem at hand:

- What will your input data be? What is your question?
  - Bear in mind: no data, no business.
- What type of problem are you facing?
  - Identifying the problem type will guide your choice of model architecture, loss function, etc.



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Universal workflow of ML:
1. Defining the problem and assembling the datasets

Define the problem at hand:

- What will your input data be? What is your question?
  - Bear in mind: no data, no business.
- What type of problem are you facing?
  - Identifying the problem type will guide your choice of model architecture, loss function, etc.

Do not move to the next stage until you know:

What your inputs/outputs are

You are making two hypotheses:

1. You hypothesize that your outputs can be predicted (given your inputs).
2. You hypothesize that you have enough available data.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
1. Defining the problem and assembling the datasets

Not all problems can be solved.

You may have inputs X and targets Y, but that doesn't mean X contains enough information to predict Y.

Tell me an example of an impossible problem (yet).

Universal workflow of ML:
1. Defining the problem and assembling the datasets

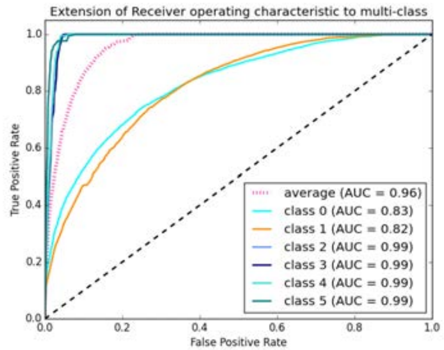
The stock market

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
2. Choosing a measure of success

To control something, you need to observe it.
To achieve success, you need to define what you mean by success

Metric for success → Guides the choice of a loss function
For balanced-classification problems → ROC AUC



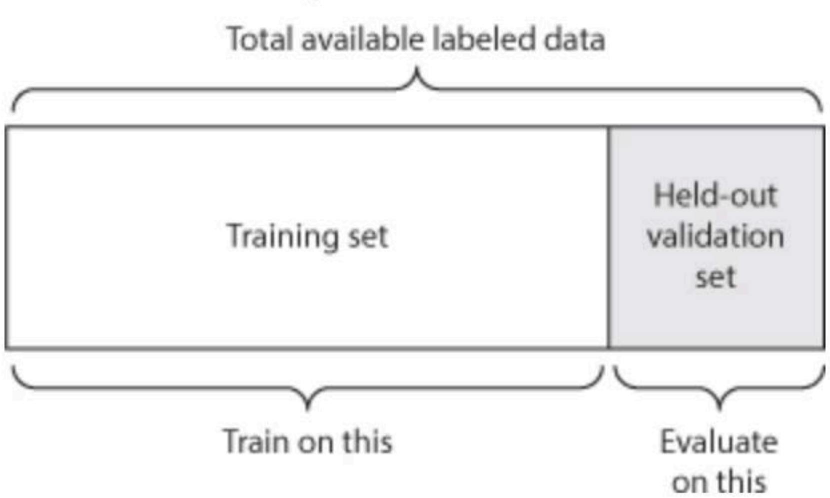
Remember:
model.compile(optimizer='rmsprop', loss='categorical_crossentropy', **metrics=['acc']**)

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
3. Deciding on an evaluation protocol

You must establish how you will measure you current progress:

- Hold-out validation set → When you have plenty of data

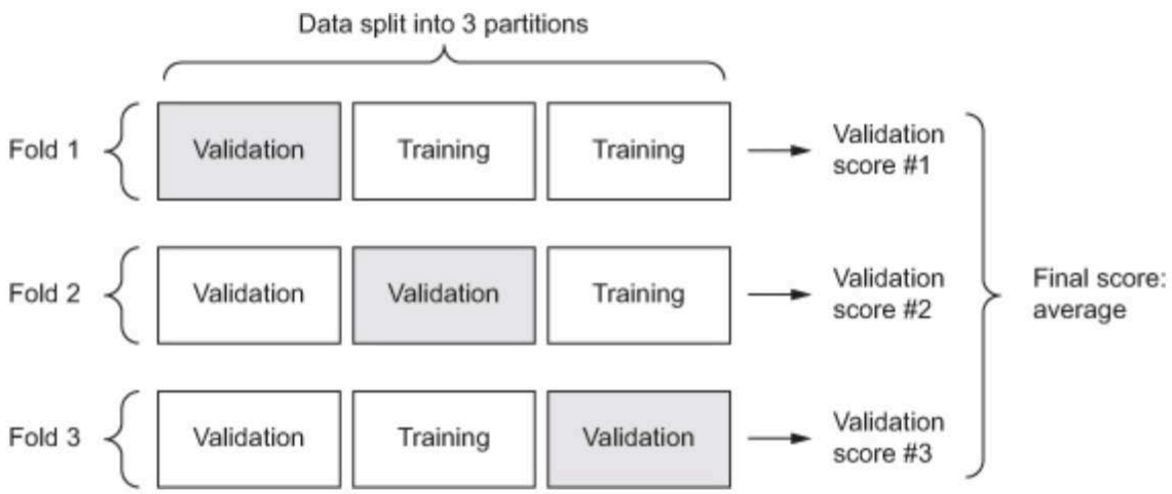


Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
3. Deciding on an evaluation protocol

You must establish how you will measure you current progress:

- Hold-out validation set → When you have plenty of data
- K-Fold cross-validation → Too few samples for hold-out validation to be reliable.



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Universal workflow of ML:
3. Deciding on an evaluation protocol

You must establish how you will measure you current progress:

- Hold-out validation set → When you have plenty of data
- K-Fold cross-validation → Too few samples for hold-out validation to be reliable.
- Iterated K-Fold validation → For performing highly accurate model evaluation when little data is available.

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
4. Preparing your data

You must format your data in a way that can be fed into a DL/ML model.

1- Data vectorization:

- All inputs in a NN must be tensors (floating point (float32) o integers)
- No matter your raw data is (images, sounds, text)….turn it into tensors.

- Wait!!! What are tensors?

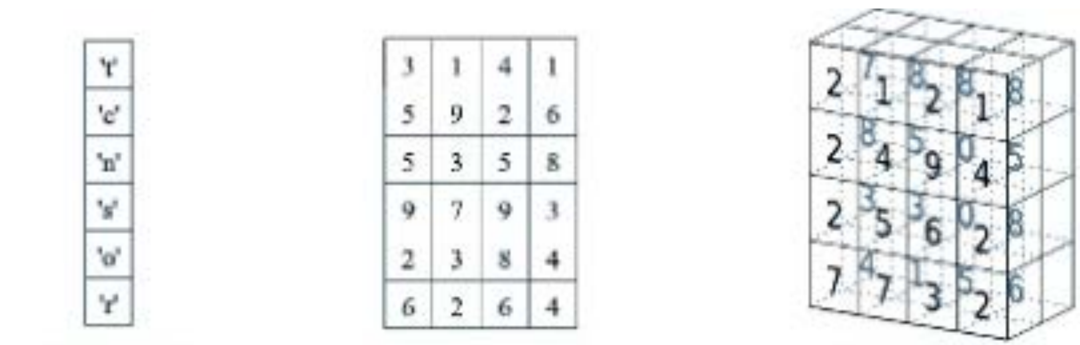
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
4. Preparing your data

1- Data vectorization:

- All inputs in a NN must be tensors (floating points!)
- No matter your raw data is (images, sounds, text)….turn it into tensors.

- Wait!!! What are tensors?
    - Multidimensional arrays.



Tensor of dimension[1] — Tensor of dimensions[2] — Tensor of dimensions[3]

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

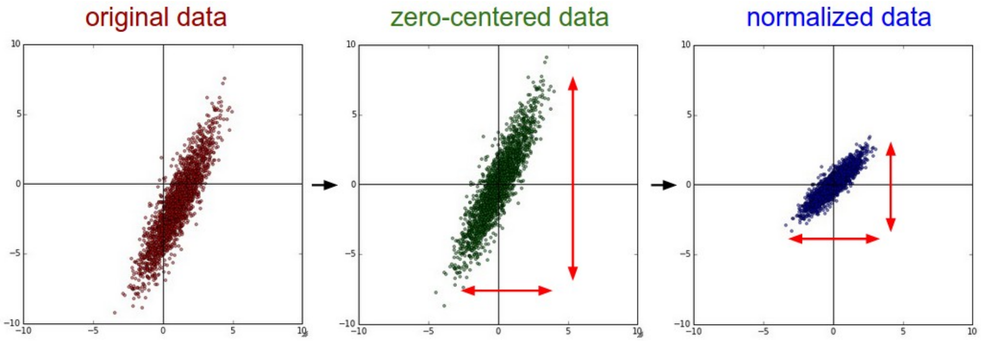
Universal workflow of ML:
4. Preparing your data

1- Data vectorization.
2- Value normalization:

Is not safe to feed into a NN data that takes large values o heterogeneous data.

Doing that can trigger large gradient updates → Prevents the NN form converging.
To make learning easier for your network:

1. Take small values → Usually in the 0 to 1 range (or -1 to 1 range).
2. Be homogeneous → All features should take values in the same range.



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

INSTITUTO UNIVERSITARIO
SIANI
INGENIERIA COMPUTACIONAL

Universal workflow of ML:
5. Baseline model

Goal → Achieve statistical power = develop a model capable of beating a dumb classifier.

Let's consider a throwing a coin problem.

2 classes…what is the baseline case?

Let's consider the MNIST problem.

10 classes…what is the baseline case?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Universal workflow of ML:
5. Baseline model

Goal → Achieve statistical power = develop a model capable of beating a dumb classifier.

Let's consider a throwing a coin problem.

2 classes…what is the baseline case?

Anything with an accuracy higher than 0.5 can be said to have statistical power

Let's consider the MNIST problem.

10 classes…what is the baseline case?

Anything with an accuracy higher than 0.1 can be said to have statistical power

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

Goal → Achieve statistical power = develop a model capable of beating a dumb classifier.

It is not always possible to achieve statistical power.

If you can't beat a random baseline after trying multiple architectures…
Then may be that the answer to your question is not present in the input data…

Remember you made two hypotheses:
1. You hypothesize that your outputs can be predicted (given your inputs).
2. You hypothesize that you have enough available data.

They can be false!

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

Universal workflow of ML:
5. Baseline model

Goal → Achieve statistical power = develop a model capable of beating a dumb classifier.

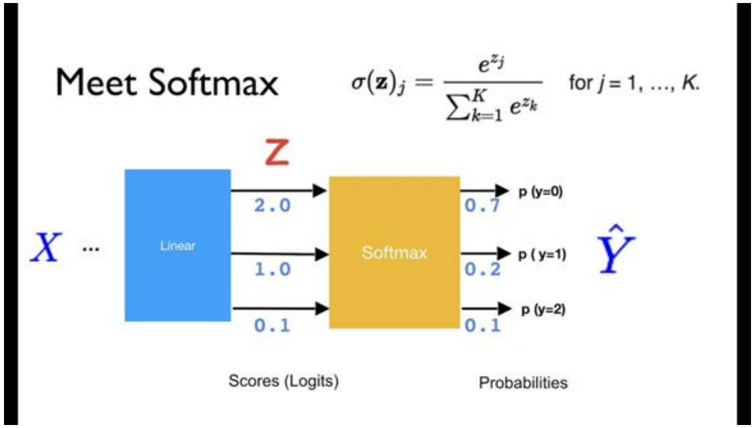
Let's assume things go well…three key choices to build our first model:

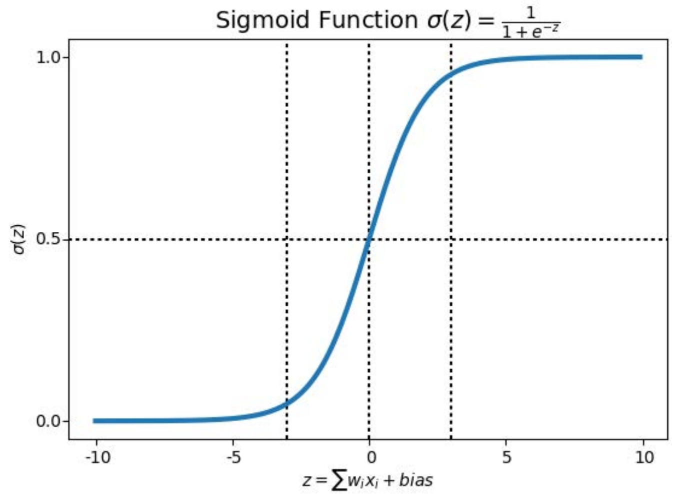
- Last-layer activation.
- Loss function.
- Optimization configuration.

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- **Last-layer activation**:
  - Establishes a useful constraints on the NN output.





Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- **Last-layer activation**:
    - Establishes a useful constraints on the NN output.

    - Do you want to know more about activation functions?:
      https://www.youtube.com/watch?v=tf9p1xQbWNM

This list is brilliant (Hugo Larochelle):
https://www.youtube.com/playlist?list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- Last-layer activation.
- **Loss function**:
  - Should match the type of problem you are solving.
  - Used to compare the networks predicted output with the real output.
  - It tells the model how the weights should be updated.
  - Common loss functions:
    - Mean squared error
    - Cross entropy (binary/categorical)
    - Etc.

Remember:

model.compile(optimizer='rmsprop', **loss='categorical_crossentropy'**, metrics=['acc'])

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

How to combine them with the type of problem?

- Last-layer activation.
- Loss function.

Table 4.1 **Choosing the right last-layer activation and loss function for your model**

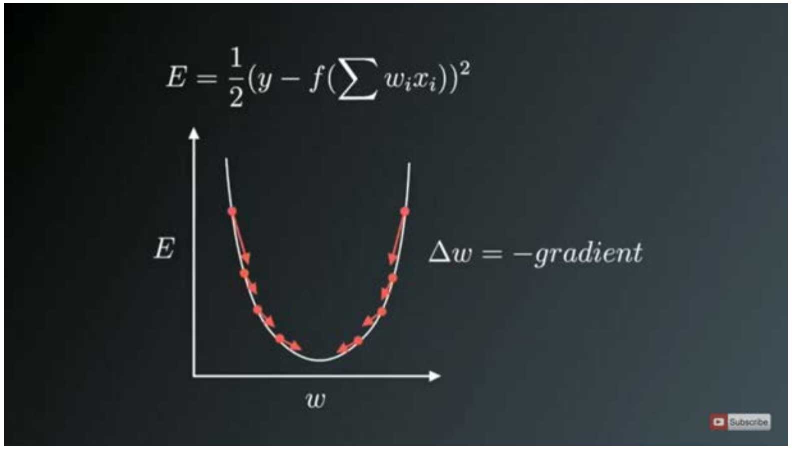
Problem type	Last-layer activation	Loss function
Binary classification	`sigmoid`	`binary_crossentropy`
Multiclass, single-label classification	`softmax`	`categorical_crossentropy`
Multiclass, multilabel classification	`sigmoid`	`binary_crossentropy`
Regression to arbitrary values	None	`mse`
Regression to values between 0 and 1	`sigmoid`	`mse` or `binary_crossentropy`

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- Last-layer activation
- Loss function
- **Optimization configuration**:
  - What optimizer you will use?
  - What learning rate?



Remember:
model.compile(**optimizer='rmsprop'**, loss='categorical_crossentropy', metrics=['acc'])

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- Last-layer activation
- Loss function
- **Optimization configuration**:
  - What optimizer you will use?
  - What learning rate?

https://cdn-images-1.medium.com/max/1600/1*XVFmo9NxLnwDr3SxzKy-rA.gif

Read: https://towardsdatascience.com/types-of-optimization-algorithms-used-in-neural-networks-and-ways-to-optimize-gradient-95ae5d39529f

Universal workflow of ML:
5. Baseline model

Let's assume things go well…three key choices to build our first model:

- Last-layer activation
- Loss function
- **Optimization configuration**:

  Stochastic Gradient Descent(SGD):

  Updates weights every training sample…takes too long!

  Loss function fluctuates to different intensities

  Mini-batch Gradient Descent:

  Updates weights every batch size of training samples

  Leads us to stable convergence.

  Range 64 to 256 works fine.

https://cdn-images-1.medium.com/max/1600/1*SjtKOauOXFVjWRR7iCtHiA.gif

HOW CAN WE CODE THIS? ANY IDEA?

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Baseline model

Goal → Achieve statistical power = develop a model capable of beating a dumb classifier.

Let's assume things go well…three key choices to build our first model:

- Last-layer activation
- Loss function
- **Optimization configuration**:

  Stochastic Gradient Descent(SGD):

  model.fit(x_train, y_train, epochs=5, **batch_size=1**)

  Mini-batch Gradient Descent:

  model.fit(x_train, y_train, epochs=5, **batch_size=128**)

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
5. Scaling up…create a model that overfits

Is your model enough poweful? Does it have enough layers? Is it the best model?

Remember:

- A NN with a single layer and 2 units have statistical power on MNIST.
- But it doesn't solve the problem well.

The true is: If your model doesn't overfit, probably is not enough.

To figure out how big a model you'll need, you must develop a model that overfits

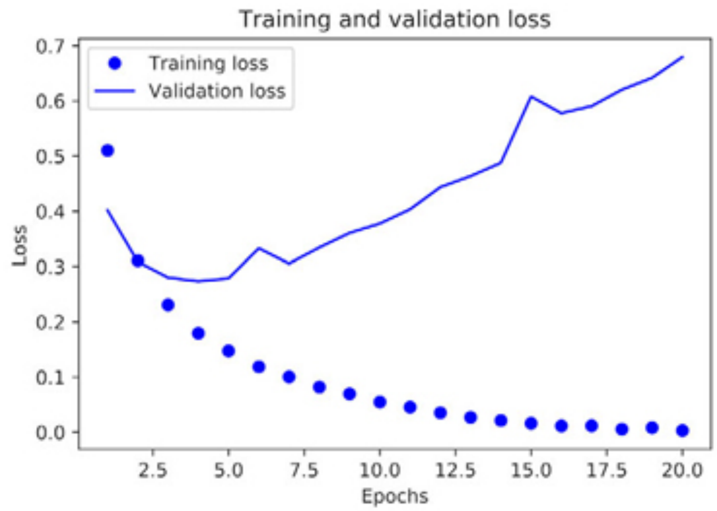
How? Any ideas?

Universal workflow of ML:
5. Scaling up…create a model that overfits

To figure out how big a model you'll need, you must develop a model that overfits:

1. Add Layers.
2. Make the layer bigger.
3. Train for more epochs.

Always monitor training loss and validation loss.



Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

Universal workflow of ML:
6. Regularize your model

Next week!

Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es