INSTITUTO UNIVERSITARIO
SIANI
INGENIERIA COMPUTACIONAL

Regularization
CASA course
(26/10/2018)

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 TS3

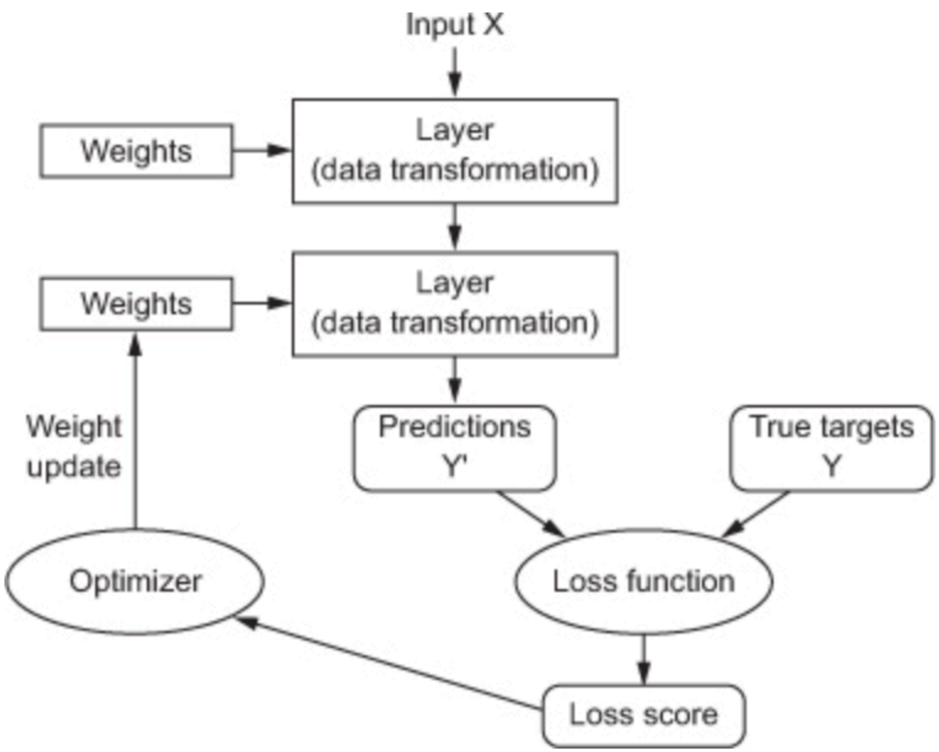
- Overfitting & Underfitting
- How to detect overfitting
- How to deal with overfitting
- How to solve overfitting: L1/L2-norms
- How to solve overfitting: Dropout

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

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

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Overfitting & Underfitting

Let's say we want to predict if a paper is likely to be published based on their abstract.

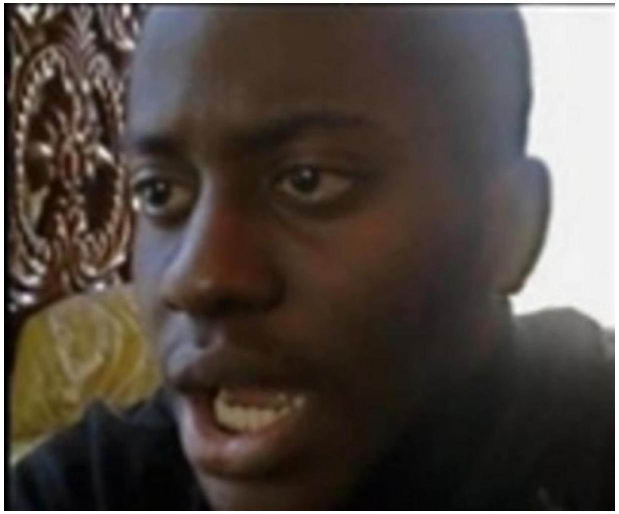
Now, assume we train a model from a dataset of 10,000 abstracst and their outcomes.

Then, we try the model out on the original dataset, and it predicts outcomes with **99% acc**.

Overfitting & Underfitting

Then, we try the model out on the original dataset, and it predicts outcomes with 99% acc.

But when we run the model on a new (“unseen”) dataset of abstracts: 50% acc.!!



Our model doesn’t generalize well from our training data to unseen data.

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

Underfitting occurs when a model is too simple

- informed by too few features
- regularized too much

Makes it inflexible in learning from the dataset.

Simple learners tend to have:

- **Less variance** in their predictions
- **More bias** towards wrong outcomes

Overfitting & Underfitting

Complex learners tend to have:

- **More variance** in their predictions
- **Less bias** towards wrong outcomes

Bias and variance???

Both bias and variance are forms of prediction error in machine learning

Overfitting & Underfitting

Both bias and variance are forms of prediction error in machine learning

Typically:

We can reduce error from bias but might increase error from variance as a result, or vice versa.

This trade-off between too simple (high bias) vs. too complex (high variance) is a key concept in statistics and machine learning.

It affects all supervised learning algorithms.

Must read:

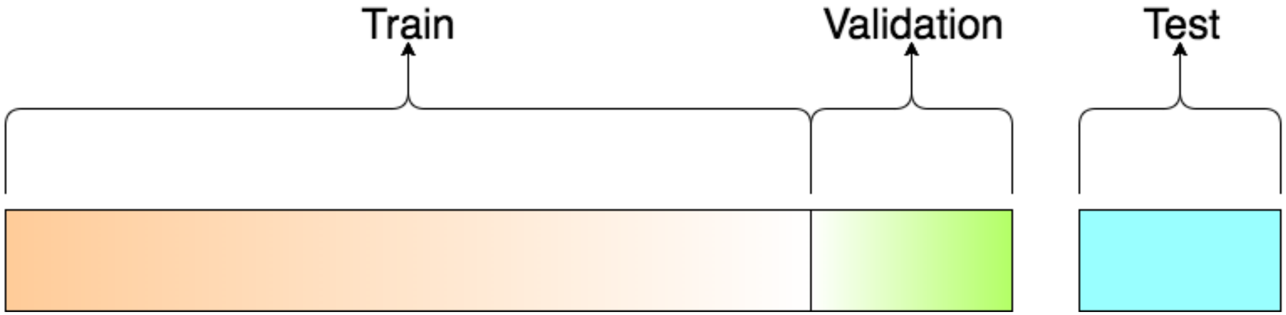
https://elitedatascience.com/bias-variance-tradeoff

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

How to detect overfitting

To address this, we can split our initial dataset into separate:

- Training set → Training set & Validation set
- Test set → Don't touch this until the very end.



This method can approximate of how well our model will perform on new data.

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

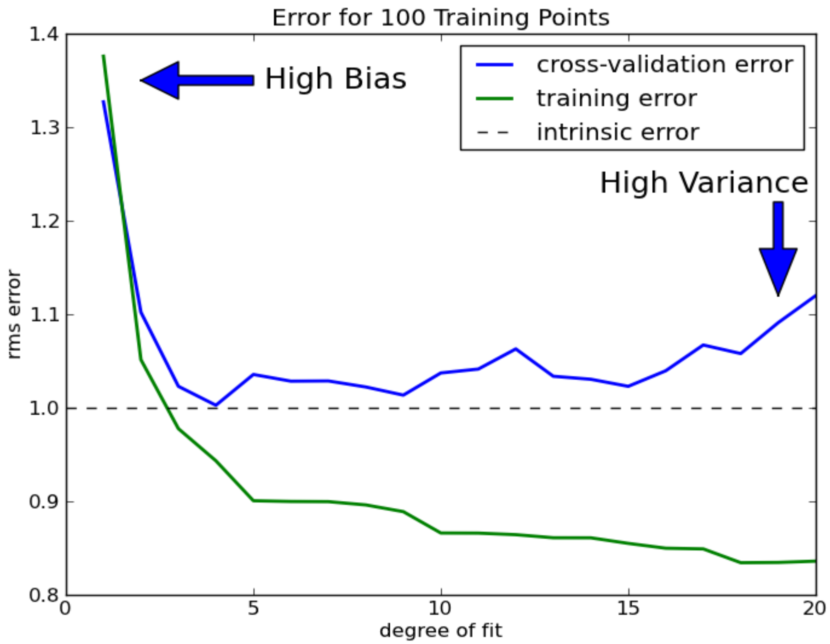
INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

How to detect overfitting

If our model does much better on the training set than on the test set, then we're likely overfitting.

For example:

- our model saw 99% accuracy on the training set,
- but only 55% accuracy on the test set.



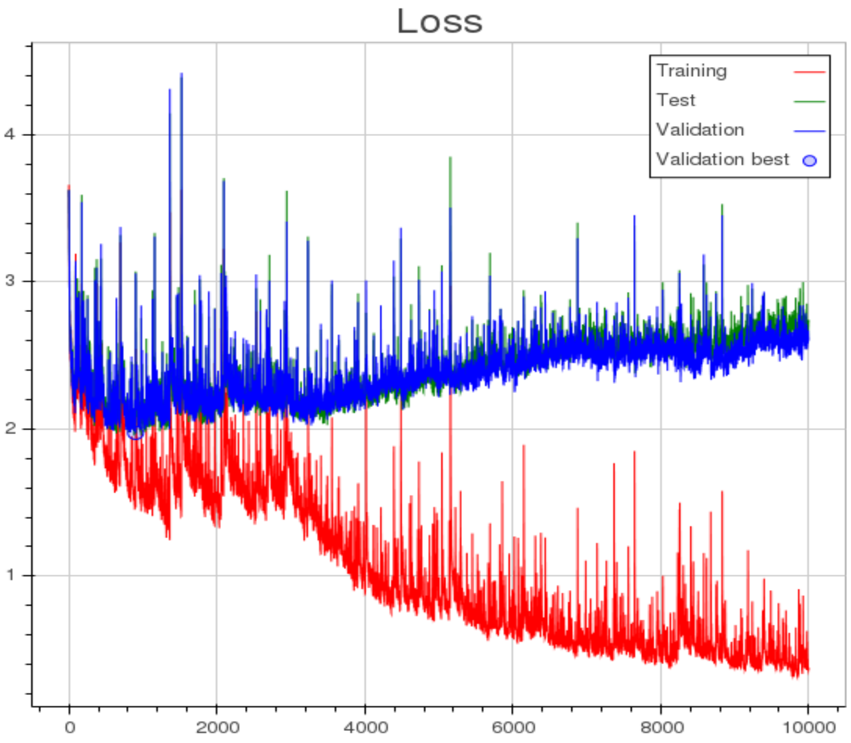
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

How to detect overfitting

If our model does much better on the training set than on the test set, then we're likely overfitting.

Also watch out the loss function:



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

How to deal with overfitting:
1- Reducing the network's size

The simplest way to prevent overfitting is to reduce the size of the model.

Learnable parameters = also kown as model's capacity:

- A model with more parameters = more memorization capacity (mapping)

DL models tend to be good at fitting to the training data, but real challenge is generalization, not fitting.

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

How to deal with overfitting:
1- Reducing the network's size

On the othe hand:

- A model with less parameters can fall into the underfitting issue.

You must find a compromise between too much capacity and not enough capacity.

How can we do this? What is the equation to achieve this?

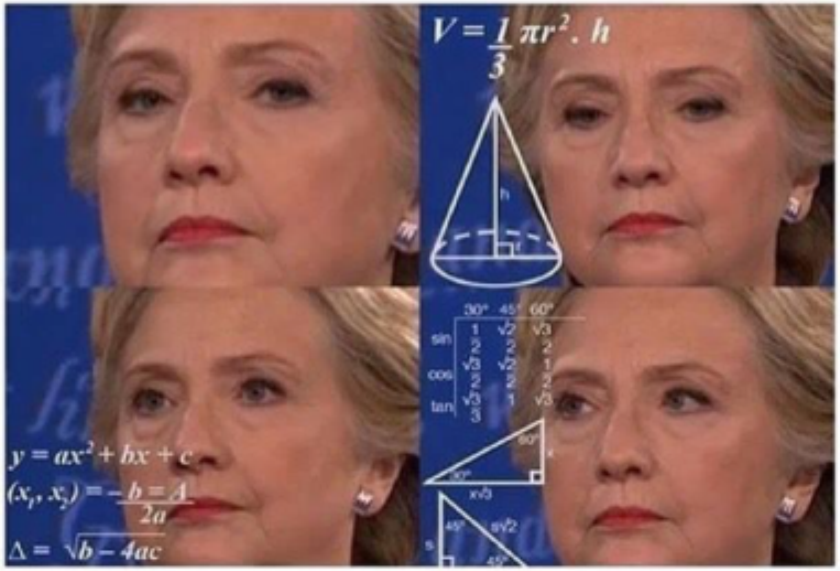
How to deal with overfitting:
1- Reducing the network's size

There is **not** magical formula to determine the:

- Right number of layers.
- Right number of units for each layer.

You must evaluate an array of different architectures in order to find the correct model size for your data.

USE THE VALIDATION SET!



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

How to deal with overfitting: 1- Reducing the network's size

Let's go to the code!

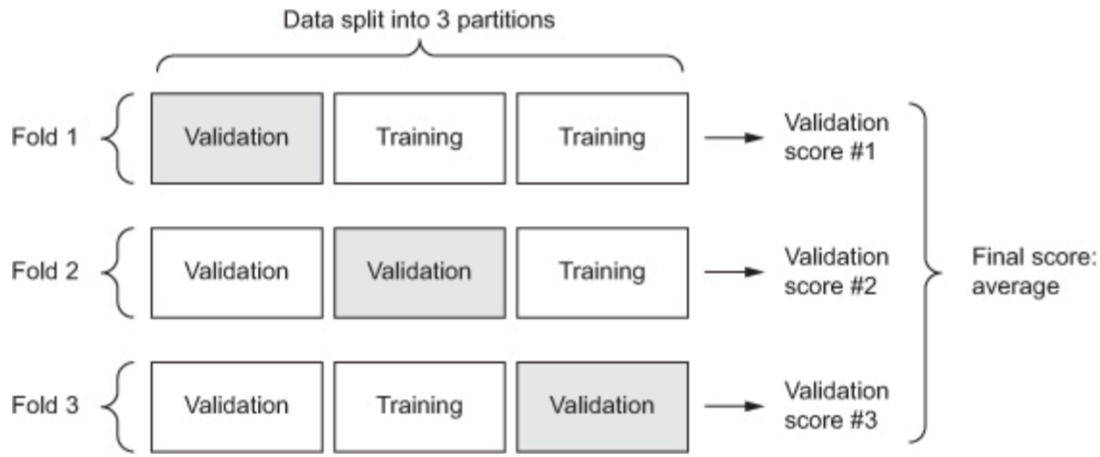
How to deal with overfitting: 2- Cross validation

Cross-validation is a powerful preventative measure against overfitting.

The idea:

1- Use your initial training data to generate multiple mini train-validation splits.

2- Use these splits to tune your model.



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

How to deal with overfitting:
2- Cross validation

In standard k-fold cross-validation:

1. We partition the data into k subsets, called folds.
2. Then, we iteratively train the algorithm on k-1 folds while using the remaining fold as the validation set (called the “holdout fold”).

Advantages:

- Allows us to tune hyperparameters with only the original training set.
- The test set is kept as a truly unseen dataset for selecting the final model.

Read: https://elitedatascience.com/machine-learning-iteration

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

How to deal with overfitting:
3- Train with more data

It won’t work everytime → more data can help algorithms detect the signal better.

If we just add more noisy data, this technique won’t help.

That’s why you should always ensure your data is clean and relevant.

Read: https://elitedatascience.com/data-cleaning

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

How to deal with overfitting: 3- Train with more data

It won't work everytime → more data can help algorithms detect the signal better.

If we just add more noisy data, this technique won't help.

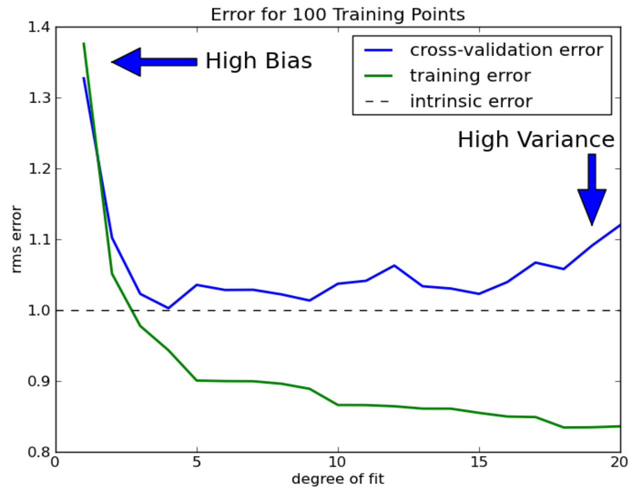
That's why you should always ensure your data is clean and relevant.

Read: https://elitedatascience.com/data-cleaning

How to deal with overfitting:
4- Early stopping

When we train a machine iteratively, we can measure how well each iteration of the model performs:

1. Up until a certain number of iterations, new iterations improve the model.
2. After that point, the model begins to overfit the training data.

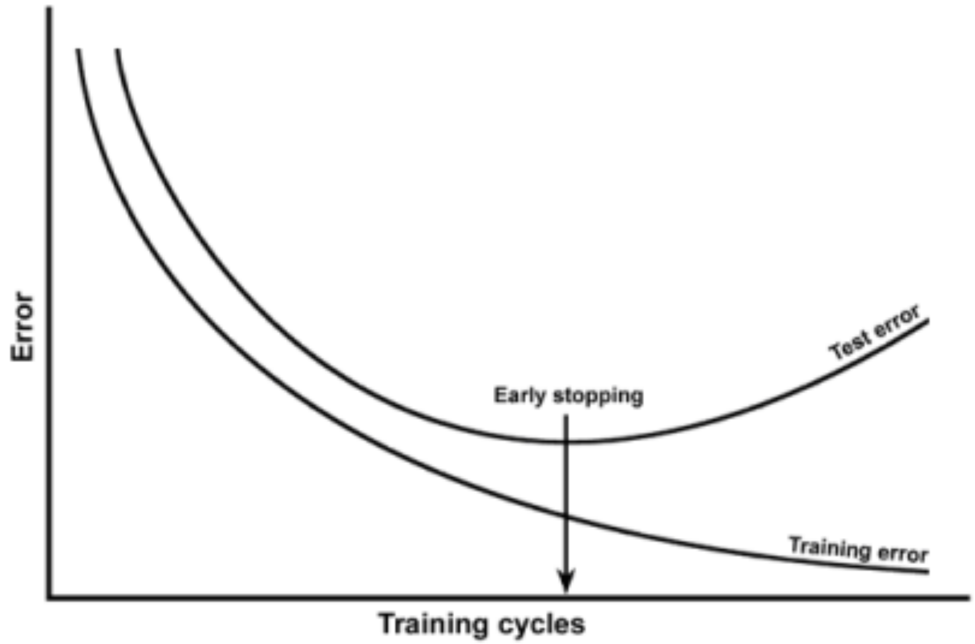


Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

How to deal with overfitting:
4- Early stopping

Early stopping→ stopping the training process before the learner passes that point.



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

How to deal with overfitting: 4- Early stopping

Let's go to the code!

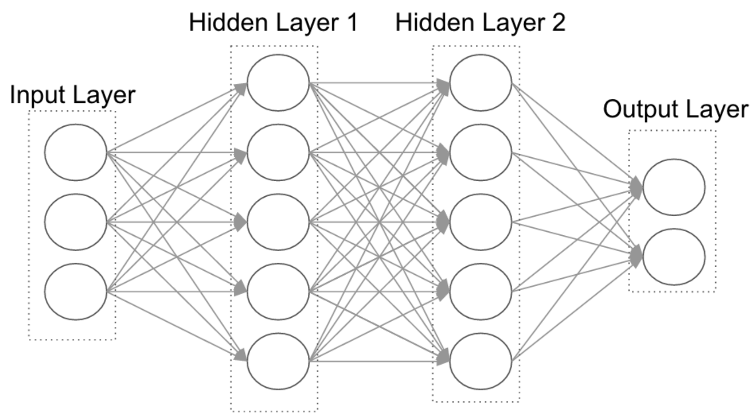
Edificio Central del Parque Tecnológico
Campus Universitario de Tafira
35017 Las Palmas de Gran Canaria
e-mail: info@siani.es · www.siani.es

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

How to solve overfitting: Regularization

Simple model → Model where the distribution of parameters values has less entropy. (less tend to overfit)

Regularization → a set of techniques for artificially forcing your model to be simpler.



Feedforward neural network with 2 hidden layers

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

How to solve overfitting: Regularization

Put constraints to the NN complexity → weight distribution more regular
How? Adding to the loss function a cost associated with having large weights

This cost can be:

- L1-norm→ cost added is proportional to the absolute value of the weights coefficients
- L2-norm→ cost added is proportional to the square of the value of the weights coefficients

Regularization penalizes large weights

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

INSTITUTO UNIVERSITARIO SIANI INGENIERIA COMPUTACIONAL

BIP LAB

How to **solve** overfitting: Regularization

L1 regularization on least squares:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2 + \boxed{\lambda \sum_{i=1}^{k} |w_i|}$$

L2 regularization on least squares:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2 + \boxed{\lambda \sum_{i=1}^{k} w_i^2}$$

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

HIPPOCRATICA CIVITAS STUDIUM SALERNI

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

How to **solve** overfitting: Regularization

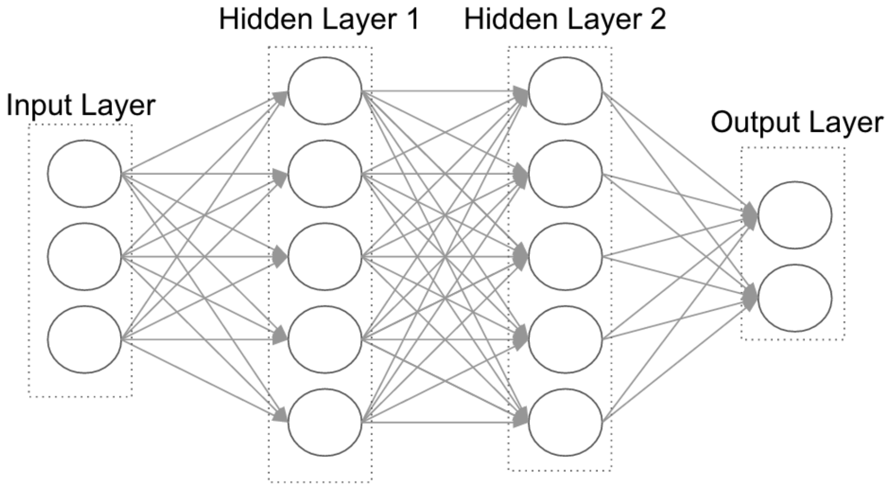
Let's go to the code!

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

How to solve overfitting: Regularization

DROPOUT

https://www.youtube.com/watch?v=NhZVe50QwPM



Feedforward neural network with 2 hidden layers

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

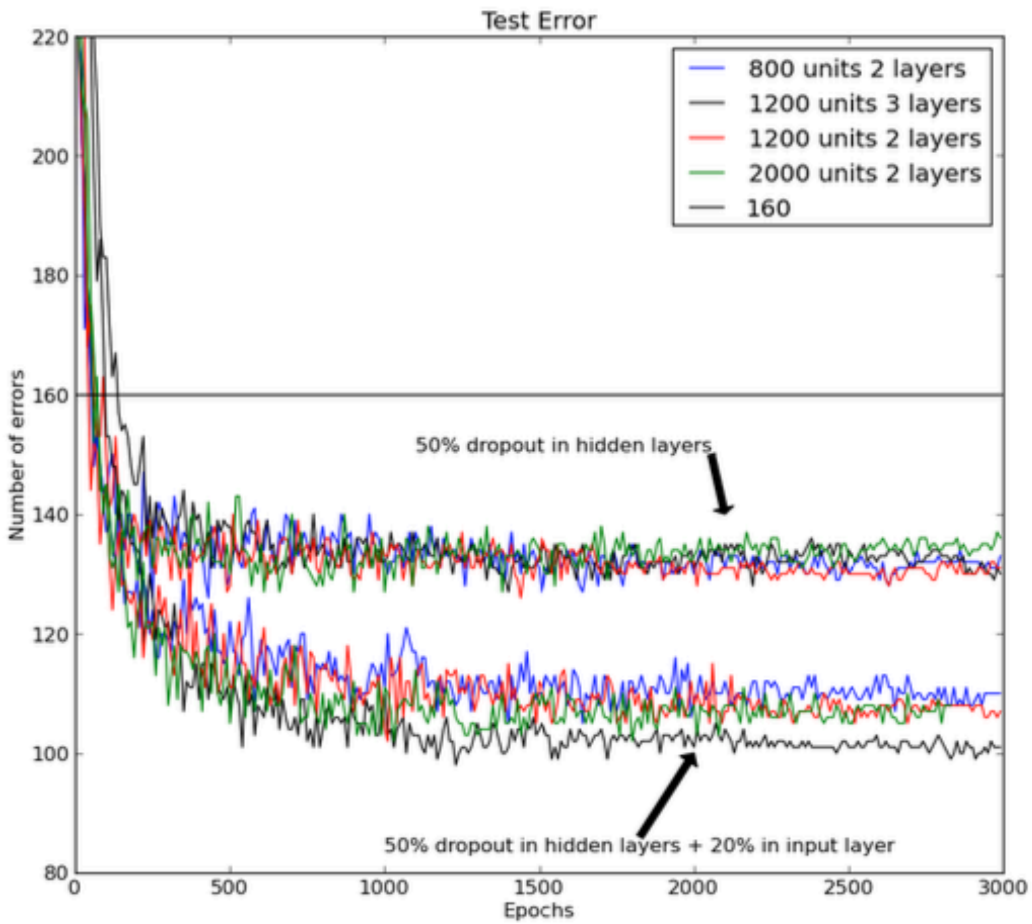
How to solve overfitting: Regularization

DROPOUT

Improving neural networks by preventing co-adaptation of feature detectors

G. E. Hinton*, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov

How to solve overfitting: Regularization



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

How to **solve** overfitting: Regularization

Some observations:

1. Dropout forces a neural network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons.

2. Dropout roughly doubles the number of iterations required to converge. However, training time for each epoch is less.

3. With H hidden units, each of which can be dropped, we have 2^H possible models. In testing phase, the entire network is considered and each activation is reduced by a factor p.

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

How to **solve** overfitting: Regularization

Let's go to the code!

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