Introduction to neural networks

Training aspects

Juan Irving Vásquez
jivg.org

Centro de Innovación y Desarrollo Tecnológico en Cómputo
Instituto Politécnico Nacional

October 9, 2023©

Contents

Roadmap

1. Performance measurement
2. Dataset split
3. Performance metrics
4. Overfitting

Performance measurement

- The central challenge in machine learning is that we must perform well on new, **previously unseen** inputs.
- The ability to perform well on previously unobserved inputs is called **generalization**.
    - Dataset split
    - Metric
    - Significance

Performance measurement

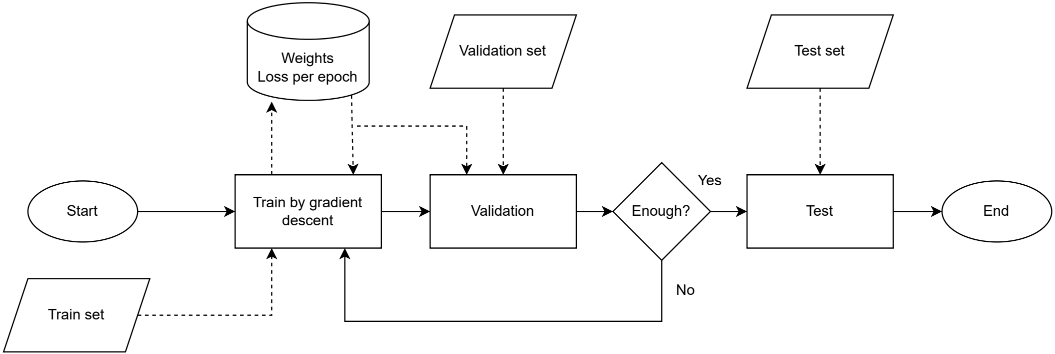


Figure: Train and test workflow

Roadmap

Dataset split

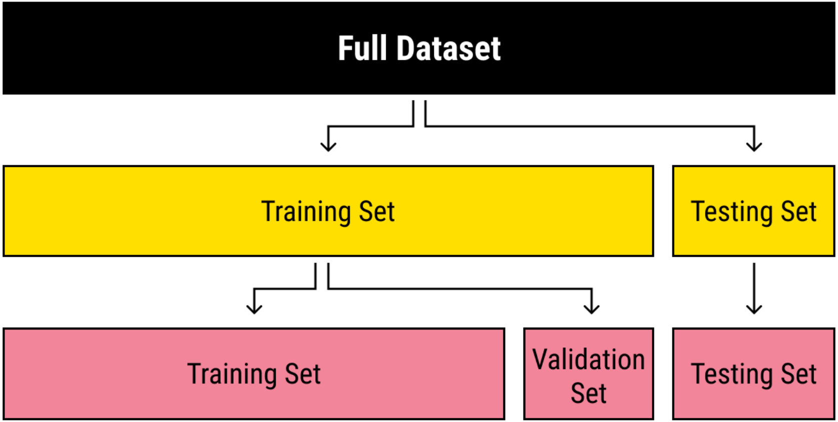


Figure: Dataset split [HJC]

Dataset split

- Lets suppose a dataset, $X_\Omega = \{x_1, \ldots, x_m\}$, with positive examples
- We could use a subset of our set of observations as an estimate of the performance

$$X = x_1, \ldots, x_n$$

$$X' = x_{n+1}, \ldots, x_m$$

  , X is the **training set** and X' is usually called the **validation set** [Smola].
- Typically, one uses about 80% of the training data for training and 20% for validation [GoodFellow]

Roadmap

Performance metrics

Common measurements for categorical outputs

- Accuracy
- Precision
- Recall
- F1 score

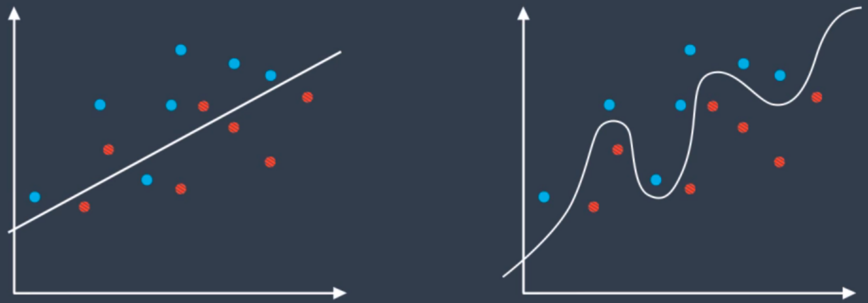
Significance

Significance

Roadmap

Overfitting problem

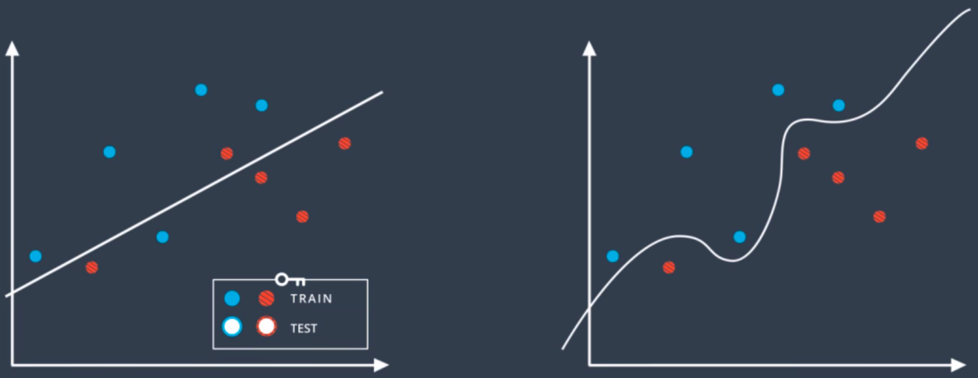
Which model is better?



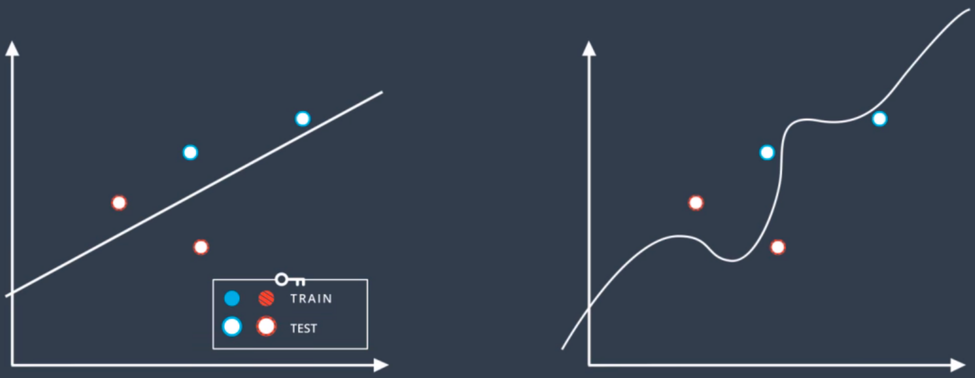
TRAIN
TEST

Overfitting problem

The importance of testing



The importance of testing



Overfitting and generalization

- **Underfitting** occurs when the model is not able to obtain a sufficiently low error value on the training set.
- **Overfitting** occurs when the gap between the training error and test error is too large.
- A model's **capacity** is its ability to fit a wide variety of functions.

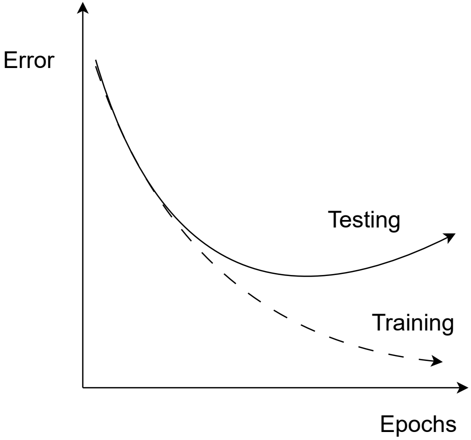


Figure: Loss as epochs advance.

Regularization

- **Regularization** is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error.
    - Early termination
    - L1, L2
    - Dropout

Regularization
Early termination

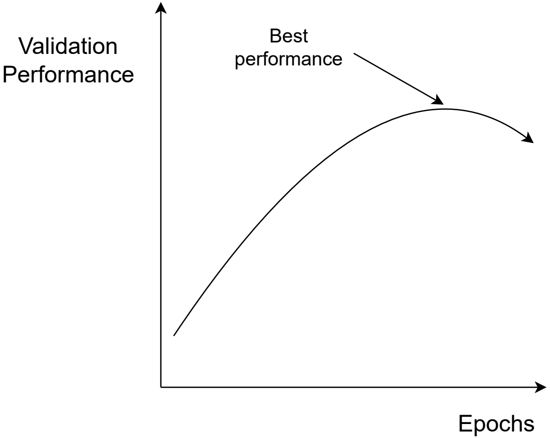
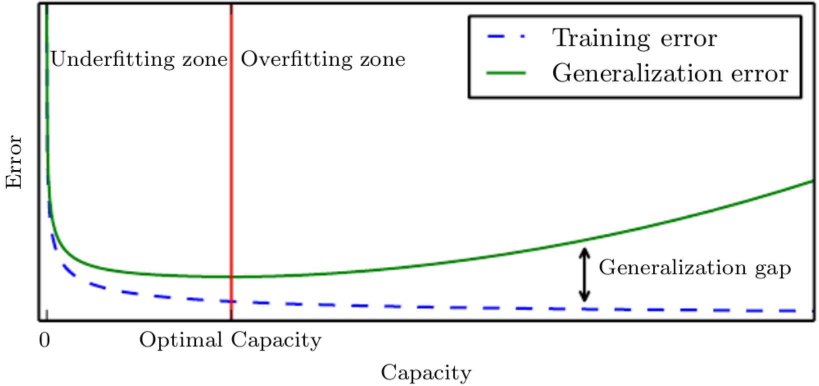


Figure: Performance as epochs advance.

Model Capacity Selection

- Usually select the simplest model
- The optimal model is the one that has the same capacity than the problem.



Lectures

- Chapter 5 Machine Learning Basics. Ian Goodfellow, et al. Deep Learning. MIT
- Chapter 2 Training feed-forward neural networks. Buduma. Fundamentals of deep learning.

References

dfellow Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

ucar15 Sucar, Luis Enrique. Probabilistic Graphical Models: Principles and Applications. Springer, 2015.

Smola Alex Smola and S.V.N. Vishwanathan, Introduction to Machine Learning, Cambirdge University Press.

Udacity Self Driving Car Nanodegree

vastava Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.

HJC https://houstonjobconnection.com/article/what-is-a-dataset-in-machine-learning