An Introduction to Machine Learning

Fabio A.

González Ph.D.

Patterns and Generalization

r robieriis

Techniques

Main Questions

An Introduction to Machine Learning

Fabio A. González Ph.D.

Depto. de Ing. de Sistemas e Industrial Universidad Nacional de Colombia, Bogotá

August 10, 2011

Learning

Main

Question

Content

Patterns and Generalization
 Generalizing from patterns
 Overfitting/ Overlearning

2 Learning Problems

Supervised Non-supervised Active On-line

- **3** Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution?

Overfitting/ Overlearning

Learning Problems

Learning Techniques

Main Questions

Outline

1 Patterns and Generalization

Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised
Non-supervise

Non-supervised

Active

On-line

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem?

How to Solve the Learning Problem?

How to Measure the Quality of a Solution?

Generalizing from

Overfitting, Overlearnin

Learning Problems

Learning Techniques

Main Questions

Outline

4□ → 4□ → 4 □ → 1 □ → 9 Q (~)

1 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

- 2 Learning Problems
 Supervised
 - Non-supervised
 - Active
 - On-line
- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution?

Generalizing from patterns

Overfitting/

Learning

Learning Techniques

Main

What is a pattern?

Data regularities

Generalizing from patterns

Overfitting, Overlearnin

Learning

Learning Techniques

Main

What is a pattern?

- Data regularities
- Data relationships

Generalizing from patterns

Overlearnin

Learning

Learning Techniques

Main

What is a pattern?

- Data regularities
- Data relationships
- Redundancy

Generalizing from patterns

Overfitting/ Overlearning

Learning Problems

Learning Techniques

Main

What is a pattern?

- Data regularities
- Data relationships
- Redundancy
- Generative model

Generalizing from patterns

Overlearni

Learning

Learning Techniques

Main

Learning a Boolean function

x_1	x_2	f_1	f_2	 f_{16}
0	0	0	0	 1
0	1	0	0	 1
1	0	0	0	 1
1	1	0	1	 1

• How many Boolean functions of n variables are?

Patterns and Generalizatio

Generalizing from patterns

Overfitting

Learning

Learning Techniques

Techniques

Main Questions

Learning a Boolean function

x_1	x_2	f_1	f_2	 f_{16}
0	0	0	0	 1
0	1	0	0	 1
1	0	0	0	 1
1	1	0	1	 1

- How many Boolean functions of n variables are?
- How many candidate functions are removed by a sample?

Learning

Learning Techniques

Main

Learning a Boolean function

x_1	x_2	f_1	f_2	 f_{16}
0	0	0	0	 1
0	1	0	0	 1
1	0	0	0	 1
1	1	0	1	 1

- How many Boolean functions of n variables are?
- How many candidate functions are removed by a sample?
- Is it possible to generalize?

Generalizing from patterns
Overfitting/

Learning

Learning

Techniques

Main

Inductive bias

 In general, the learning problem is ill-posed (more than one possible solution for the same particular problem, solutions are sensitive to small changes on the problem) Learning

Learning

Techniques

Main

Inductive bias

- In general, the learning problem is ill-posed (more than one possible solution for the same particular problem, solutions are sensitive to small changes on the problem)
- It is necessary to make additional assumptions about the kind of pattern that we want to learn

Main

Inductive bias

- In general, the learning problem is ill-posed (more than one possible solution for the same particular problem, solutions are sensitive to small changes on the problem)
- It is necessary to make additional assumptions about the kind of pattern that we want to learn
- **Hypothesis space**: set of valid patterns that can be learnt by the algorithm

Learning Problems

Learning Techniques

Main

Outline

1 Patterns and Generalization

Generalizing from patterns

Overfitting/ Overlearning

2 Learning Problems

Supervised

Non-supervised

Active

On-line

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem?

How to Solve the Learning Problem?

How to Measure the Quality of a Solution?

Patterns and Generalizatio

Generalizing from

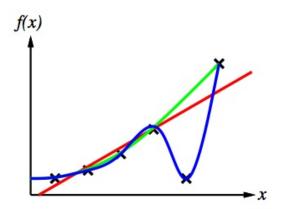
Overfitting/ Overlearning

Learning

Learning

Main

What is a good pattern?



Generalizing from patterns

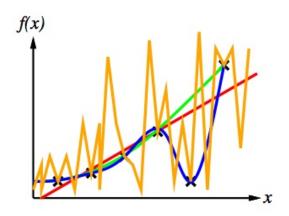
Overfitting/ Overlearning

Learning

Learning Techniques

Main Questions

What is a good pattern?



Overlearning Learning

Learning Techniques

Technique

Occam's Razor

from Wikipedia:

Occam's razor (also spelled Ockham's razor) is a principle attributed to the 14th-century English logician and Franciscan friar William of Ockham. The principle states that the explanation of any phenomenon should make as few assumptions as possible, eliminating, or "shaving off", those that make no difference in the observable predictions of the explanatory hypothesis or theory. The principle is often expressed in Latin as the lex parsimoniae (law of succinctness or parsimony).

"All things being equal, the simplest solution tends to be the best one." Fabio A.

González Ph.D.

Patterns and Generalizatio

Learning Problems

Non-supervise
Active

Learning Techniques

Main

Outline

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised Non-supervised Active

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution?

Generalization

Learning Problems

Supervised
Non-supervised

Active On-line

Techniques

Main Questions

- Supervised learning
- Non-supervised learning
- Semi-supervised learning
- Active learning
- On-line learning

Supervised

Non-superv Active

Learning Techniques

Main

Outline

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems Supervised

Non-supervised Active

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution?

Patterns an Generalizati

Learning

Supervised

Non-supervis

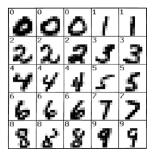
Active

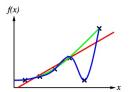
Learning

Techniques

Supervised learning

 Fundamental problem: to find a function that relates a set of inputs with a set of outputs





Patterns and Generalization

Learning

Supervised

Non-supervis

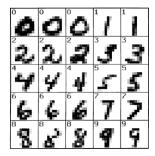
On-line

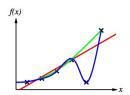
Learning Technique

Main Question

Supervised learning

- Fundamental problem: to find a function that relates a set of inputs with a set of outputs
- Typical problems:





Patterns and Generalization

Learning

Supervised

Non-supervis

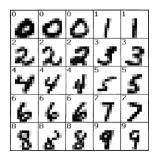
On-line

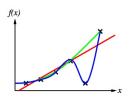
Learning Techniques

Main Questions

Supervised learning

- Fundamental problem: to find a function that relates a set of inputs with a set of outputs
- Typical problems:
 - Classification





Patterns and Generalization

Learning

Supervised

Non-supervis

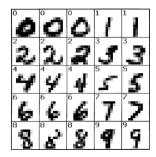
On-line

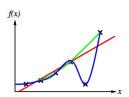
Learning Techniques

Main

Supervised learning

- Fundamental problem: to find a function that relates a set of inputs with a set of outputs
- Typical problems:
 - Classification
 - Regression





Fabio A.

Problems

Non-supervised

Non-supervi

On-line

Learning Techniques

l echniques

A Datterna and Consultation

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised

Non-supervised

Active

On-line

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution?

Outline

Supervised Non-supervised

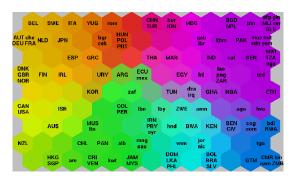
Active

Learning Techniques

Main Questions

Non-supervised learning

• There are not labels for the training samples



Patterns and Generalization

Learning Problems

Supervised Non-supervised

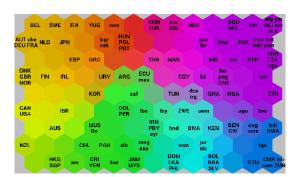
Active On-line

Learning

Main

Non-supervised learning

- There are not labels for the training samples
- Fundamental problem: to find the subjacent structure of a training data set



Generalization

Learning Problems

Non-supervised Active

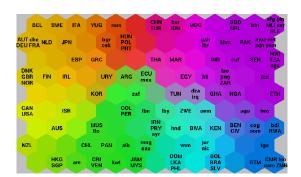
On-line

Learning Techniques

Main

Non-supervised learning

- There are not labels for the training samples
- Fundamental problem: to find the subjacent structure of a training data set
- Typical problems: clustering, data compression



Generalizatio

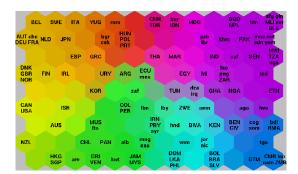
Problems
Supervised
Non-supervised

Learning

Main

Non-supervised learning

- There are not labels for the training samples
- Fundamental problem: to find the subjacent structure of a training data set
- Typical problems: clustering, data compression
- Some samples may have labels, in that case it is called semi-supervised learning



Fabio A.

Non-supervise

Active

Active On-line

Learning Techniques

Main Questions

Outline

Patterns and Generalization
 Generalizing from patterns
 Overfitting/ Overlearning

2 Learning Problems

Supervised Non-supervised

Active

On-line

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution? Generalizatio

Supervised

Non-supervis Active

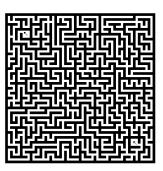
On-line

Learning Technique

Main

Active/reinforcing learning

 Generally, it happens in the context of an agent acting in an environment



Problems
Supervised
Non-supervise

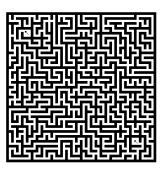
Active

L earning

Technique

Active/reinforcing learning

- Generally, it happens in the context of an agent acting in an environment
- The agent is not told whether it has make the right decision or not



Supervised
Non-supervise

Active

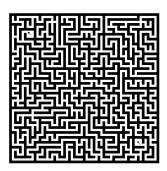
Learning

Techniques

Main Question

Active/reinforcing learning

- Generally, it happens in the context of an agent acting in an environment
- The agent is not told whether it has make the right decision or not
- The agent is punished or rewarded (not necessarily in an immediate way)



Problems
Supervised
Non-supervise

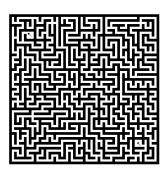
Active On-line

Learning Technique

Main Question

Active/reinforcing learning

- Generally, it happens in the context of an agent acting in an environment
- The agent is not told whether it has make the right decision or not
- The agent is punished or rewarded (not necessarily in an immediate way)
- Fundamental problem: to define a policy that allows to maximize the positive stimulus (reward)



Fabio A.

González Ph.D.

Patterns and Generalizatio

Problems
Supervised

Non-supervised Active

On-line

Techniques

Main Questions

Outline

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised
Non-supervised
Active

On-line

- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution? Fabio A. González Ph.D.

Patterns and Generalization

Learning

Supervised Non-supervised

Active

On-line

Techniques

Main Questions

On-line learning

• Only one pass through the data

Learning

Supervised Non-supervised

On-line

Learning

Main

- Only one pass through the data
 - big data volume

Learning

Supervised Non-supervised

Active

On-line

Learning Techniques

Main Questions

- Only one pass through the data
 - big data volume
 - real time

Learning

Supervised Non-supervised

Active

On-line

Techniques

Main Questions

- Only one pass through the data
 - big data volume
 - real time
- It may be supervised or unsupervised

On-line

Learning Technique

Main Questions

- Only one pass through the data
 - big data volume
 - real time
- It may be supervised or unsupervised
- Fundamental problem: to extract the maximum information from data with minimum number of passes

Learning

Techniques

Question

Outline

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised
Non-supervised
Active

- **3** Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution? Learning

Techniques

Main Questions

Representative techniques

- Computational
 - Decision trees
 - Nearest-neighbor classification
 - Graph-based clustering
 - Association rules
- Statistical
 - Multivariate regression
 - Linear discriminant analysis
 - Bayesian decision theory
 - Bayesian networks
 - K-means

- Computational-Statistical
 - SVM
 - AdaBoost
- Bio-inspired
 - Neural networks
 - Genetic algorithms
 - Artificial immune systems

Problems

Learning Technique

Main

Questions

How to State the Learning Problem? How to Solve the Learning Problem? Outline

 Patterns and Generalization Generalizing from patterns Overfitting/ Overlearning

2 Learning Problems

Supervised Non-supervised Active

3 Learning Techniques

4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem? How to Measure the Quality of a Solution? r robieriis

Techniqu

Ouestion

How to State the Learning Problem?

Learning Problem?

How to Measure the

Outline

- Patterns and Generalization
 Generalizing from patterns
 Overfitting/ Overlearning
- 2 Learning Problems Supervised Non-supervised Active
 On line
- 3 Learning Techniques
- 4 Main Questions

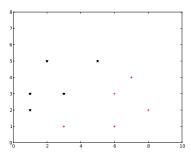
How to State the Learning Problem?

How to Solve the Learning Problem? How to Measure the Quality of a Solution?

How to State the

Learning Problem?

Two Class Classification Problem



• The idea is to build a linear classifier function, $f: \mathbb{R}^2 \to \mathbb{R}$. such that:

$$f(x,y) = \begin{cases} <0 & \text{if } (x,y) \in C_0\\ \ge 0 & \text{if } (x,y) \in C_1 \end{cases}$$

Learning

Learning

Technique

O

How to State the Learning Problem?

How to Solve the

How to Measure the Quality of a

Loss Function

- Training set: $S = \{((x_1, y_1), l_1), \dots, ((x_n, y_n), l_n)\}$
 - Example:

$$S = \{((1,2),-1),((1,3),-1),((3,1),1),\dots\}$$

Quality of a

Loss Function

- Training set: $S = \{((x_1, y_1), l_1), \dots, ((x_n, y_n), l_n)\}$
 - Example: $S = \{((1,2),-1),((1,3),-1),((3,1),1),\dots\}$
- Loss function:

$$L(f, S) = \frac{1}{2} \sum_{(x_i, y_i) \in S} (f(x_i, y_i) - l_n)^2$$

Loss Function

- Training set: $S = \{((x_1, y_1), l_1), \dots, ((x_n, y_n), l_n)\}$
 - Example: $S = \{((1,2),-1),((1,3),-1),((3,1),1),\dots\}$
- Loss function:

$$L(f, S) = \frac{1}{2} \sum_{(x_i, y_i) \in S} (f(x_i, y_i) - l_n)^2$$

Are there other alternative loss functions?

Fabio A. González Ph.D.

Patterns and Generalization

Learning

Learning Technique

Techniqu

Question

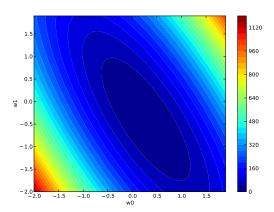
How to State the Learning Problem?

How to Solve the

How to Measure to

Square Error Loss

$$f(x,y) = w_1 x + w_0 y$$



Learning

Technique

Main

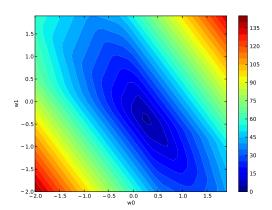
How to State the Learning Problem?

Learning Problem

How to Measure th Quality of a

L_1 Error Loss

$$f(x,y) = w_1 x + w_0 y$$



Learning

Learning

Techniqu

Question

How to State the Learning Problem?

How to Solve the

How to Measure the Quality of a

Learning as Optimization

• General optimization problem:

$$\min_{f \in H} L(f,S)$$

How to Measure the Quality of a Solution?

Learning as Optimization

• General optimization problem:

$$\min_{f \in H} L(f, S)$$

Two Class 2D Classification:

$$H = \{f : f(x, y) = w_2 x + w_1 y + w_0, \forall w_0, w_1, w_2 \in \mathbb{R}\}\$$

$$\min_{f \in H} L(f, S) = \min_{W \in \mathbb{R}^3} \frac{1}{2} \sum_{(x_i, y_i) \in S} (w_2 x_i + w_1 y_i + w_0 - l_i)^2$$

How to State the

How to Solve the

How to Measure the

Outline

- Patterns and Generalization
 Generalizing from patterns
 Overfitting/ Overlearning
- 2 Learning Problems
 Supervised
 Non-supervised
 Active
- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem?

How to Solve the Learning Problem?

How to Measure the Quality of a Solution?

Gradient Descent

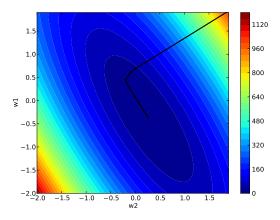
Iterative optimization of the loss function:

$$\begin{split} & \text{initialize} \quad W^0 = w_0, w_1, w_2 \\ & k \leftarrow 0 \\ & \text{repeat} \\ & \quad k \leftarrow k+1 \\ & \quad W^k \leftarrow W^{k-1} - \eta(k) \nabla L(f_{W^{k-1}}, S) \\ & \text{until} \quad |\eta(k) \nabla L(f_{W^{k-1}}, S)| < \Theta \end{split}$$

Fabio A. González Ph.D.

How to Solve the Learning Problem?

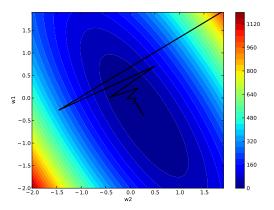
Gradient Descent Iteration Example (1)



Fabio A. González Ph.D.

How to Solve the Learning Problem?

Gradient Descent Iteration Example (2)



Generalizatio

Learning

Learning Technique

Main

How to State the Learning Problem? How to Solve the

How to Measure the Quality of a Solution?

Outline

- Patterns and Generalization
 Generalizing from patterns
 Overfitting/ Overlearning
- 2 Learning Problems Supervised Non-supervised Active
- 3 Learning Techniques
- 4 Main Questions

How to State the Learning Problem? How to Solve the Learning Problem?

How to Measure the Quality of a Solution?

Problems

Learning Fechnique

Main

Solution?

How to State the Learning Problem

Learning Problem?

How to Measure the Quality of a

Training Error vs Generalization Error

• The loss function measures the error in the training set

Loorning

Technique

reciiiiqu

Questio

How to State th Learning Problem

How to Solve the Learning Problem

How to Measure the Quality of a Solution?

Training Error vs Generalization Error

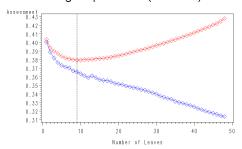
- The loss function measures the error in the training set
- Is this a good measure of the quality of the solution?

Fabio A. González Ph.D.

How to Measure the Quality of a Solution?

Training Error vs Generalization Frror

- The loss function measures the error in the training set
- Is this a good measure of the quality of the solution? Average Square Error (Gini index)



Training

Solution?

How to Solve the Learning Problem? How to Measure the Quality of a

Generalization Error

• Generalization error:

$$E[(L(f_w,S)]$$

Solution?

How to Solve the Learning Problem? How to Measure the Quality of a

Generalization Error

• Generalization error:

$$E[(L(f_w,S)]$$

How to control the generalization error during training?

How to Measure the Quality of a Solution?

Generalization Error

Generalization error:

$$E[(L(f_w,S)]$$

- How to control the generalization error during training?
 - Cross validation

Learning

Technique

Main

Questio

Solution?

How to State the Learning Problem?

Learning Problem?

How to Measure the Quality of a

Generalization Error

• Generalization error:

$$E[(L(f_w,S)]$$

- How to control the generalization error during training?
 - Cross validation
 - Regularization

Learning Problem

How to Measure the Quality of a Solution?

Regularization

Vapnik, 1995:

$$R(\alpha) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, y)$$

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(\mathbf{x}_i, \alpha)|.$$

$$R(\alpha) \le R_{emp}(\alpha) + \sqrt{\left(\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}\right)}$$

An Introduction to Machine Learning

Fabio A. González Ph.D.

Patterns and Generalization

Problems

Learning Techniques

Main

How to State the Learning Problem? How to Solve the Learning Problem?

How to Measure the Quality of a Solution? Alpaydin, E. 2004 Introduction to Machine Learning (Adaptive Computation and Machine Learning). The MIT Press. (Cap 1,2)