Identifying the anomalous behavior of large-value payment system participants with artificial neural networks. [Colombia use case]

Course on Financial Technologies and Central Banking
CEMLA & UCL
México City | México | November 12-14, 2019

Oscar Acero
Engineer
Banco de la República
oaceroac@benrep.gov.co

Carlos León
Senior Researcher
Banco de la República & Tilburg University
cleonrin@banrep.gov.co / carlosleonr@hotmail.com

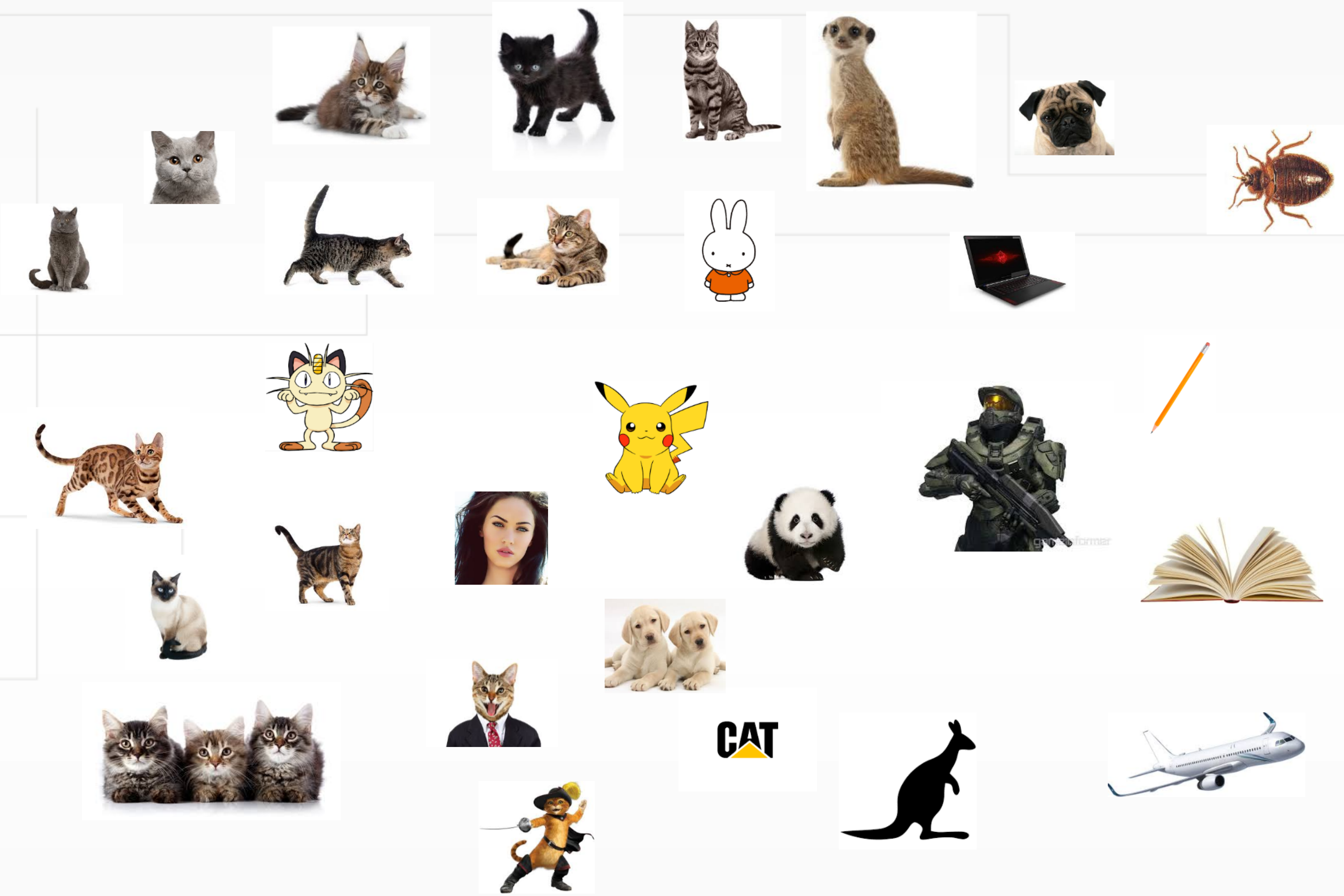
Fabio Ortega
Banco de la República
Monitoring Leader
fortegca@banrep.gov.co

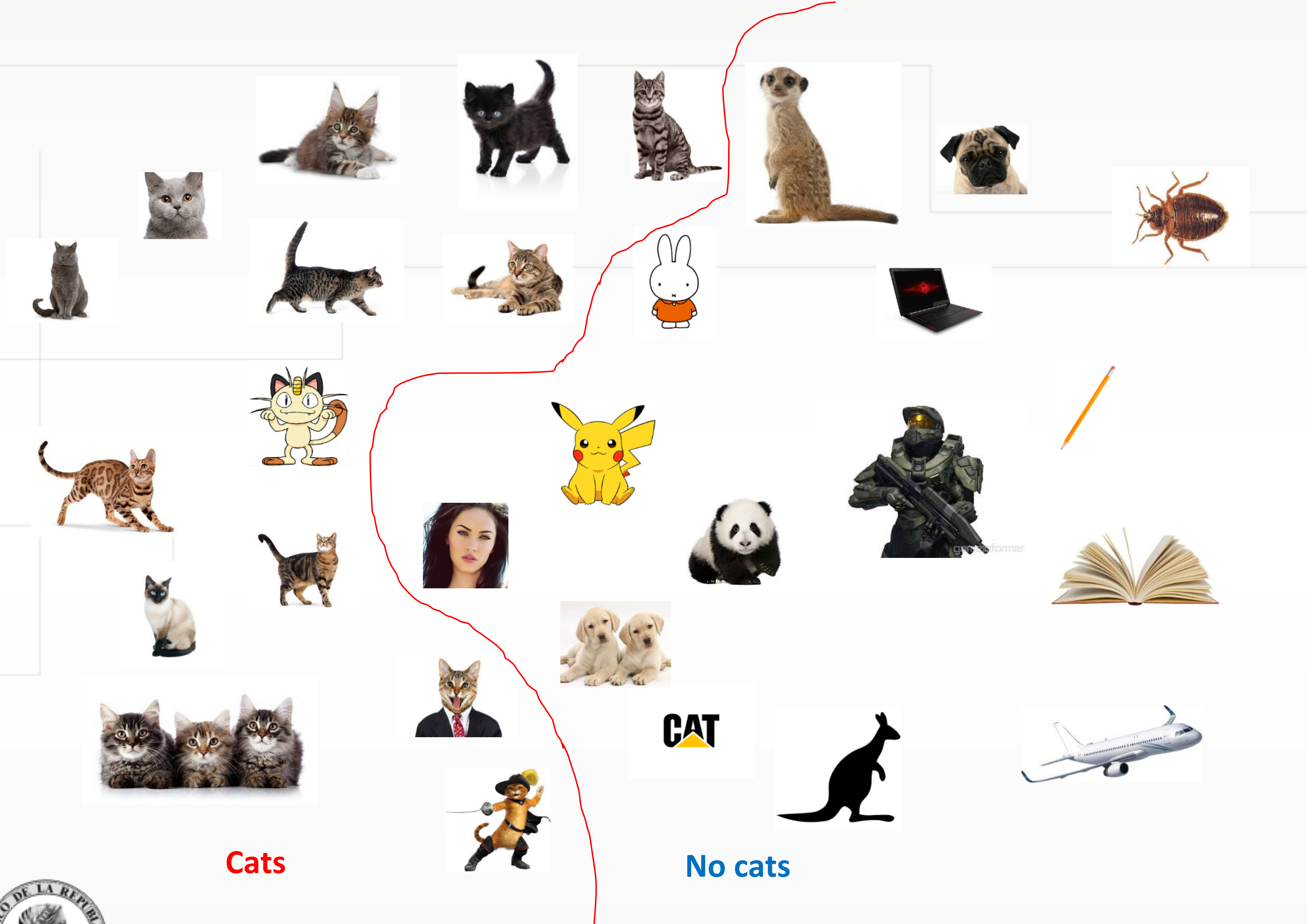
Disclaimer

The opinions and statements are the sole responsibility of the authors and do not represent neither those of Banco de la República nor of its Board of Directors. Comments and suggestions from Hernando Vargas, Clara Machado, Freddy Cepeda, José Fernado Moreno, Jorge Cely, and other members of the technical staff of Banco de la República to previous and current research work are appreciated. Any remaining errors are the authors' own.

Contents

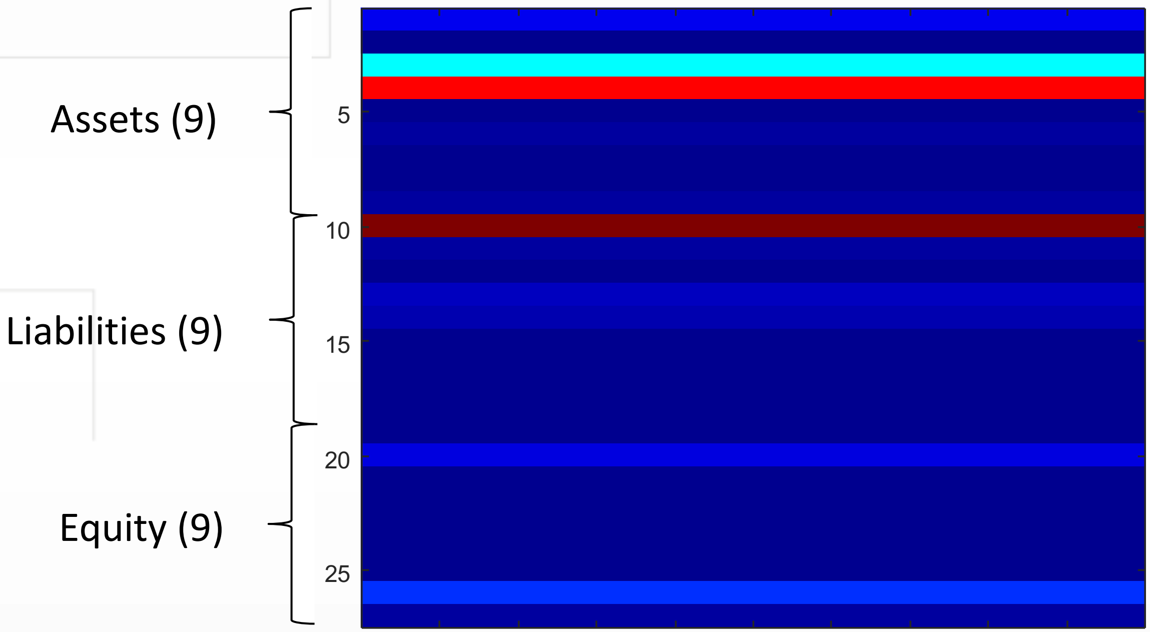
Contents

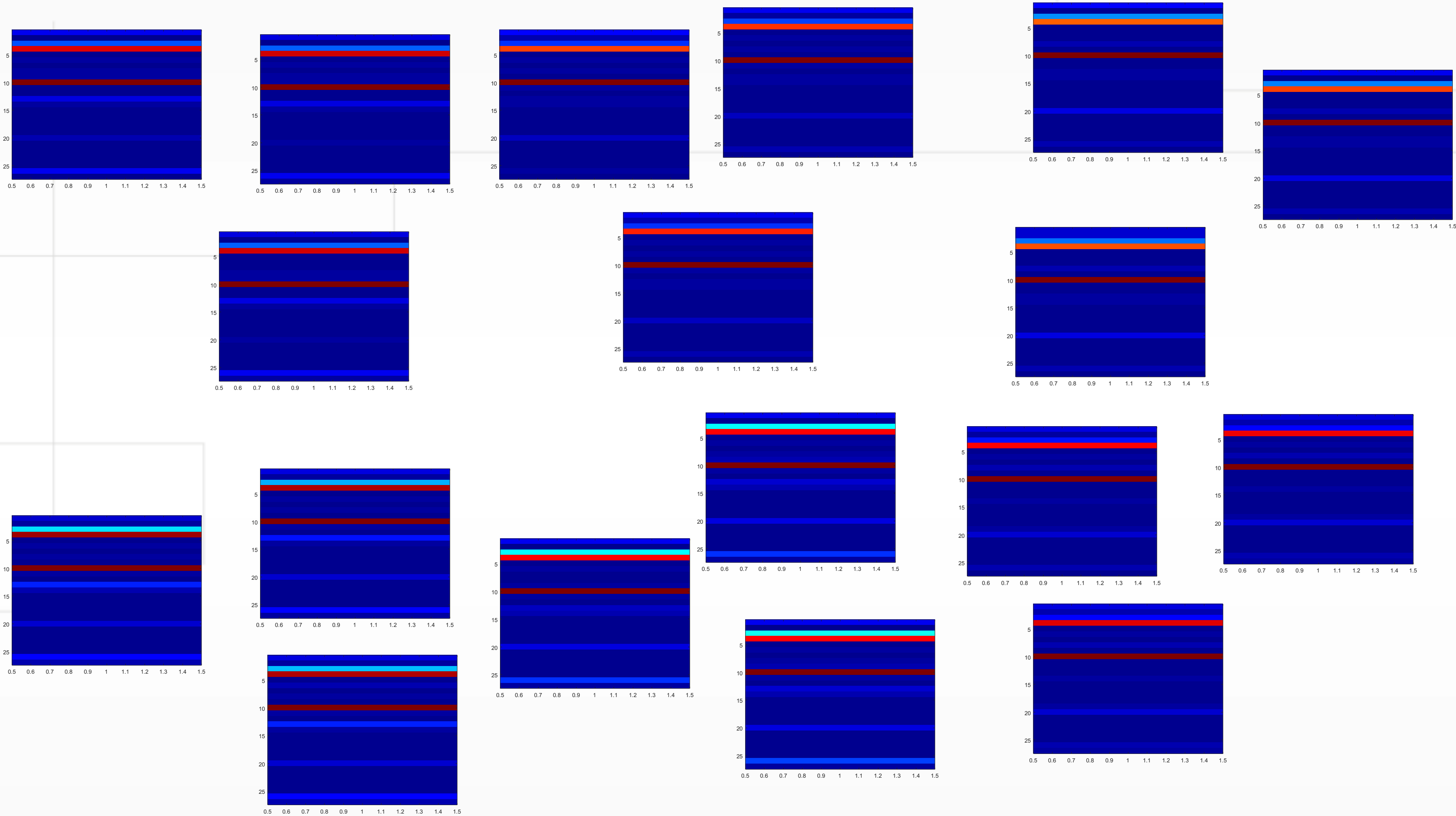

CAT

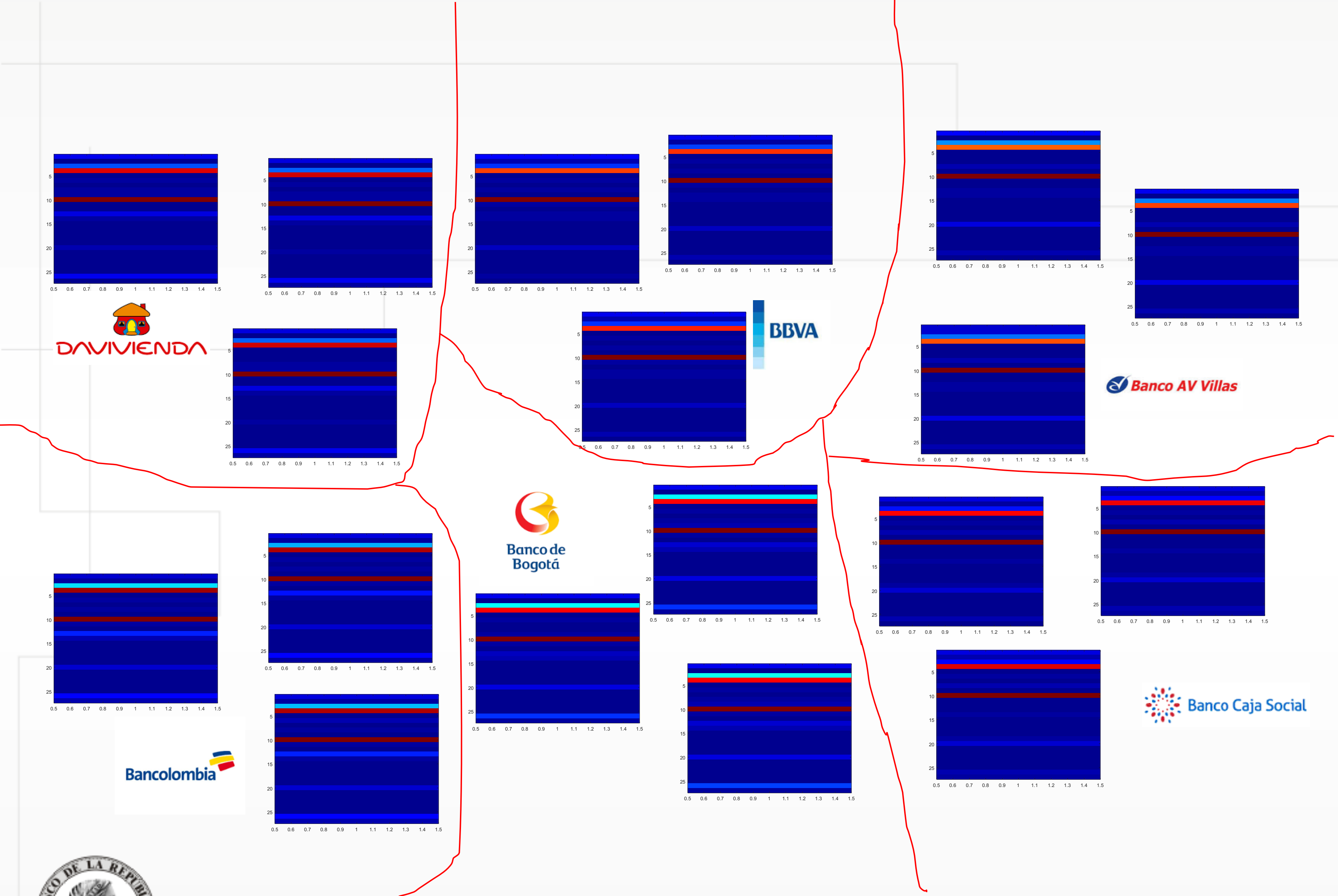

CAT
Cats
No cats

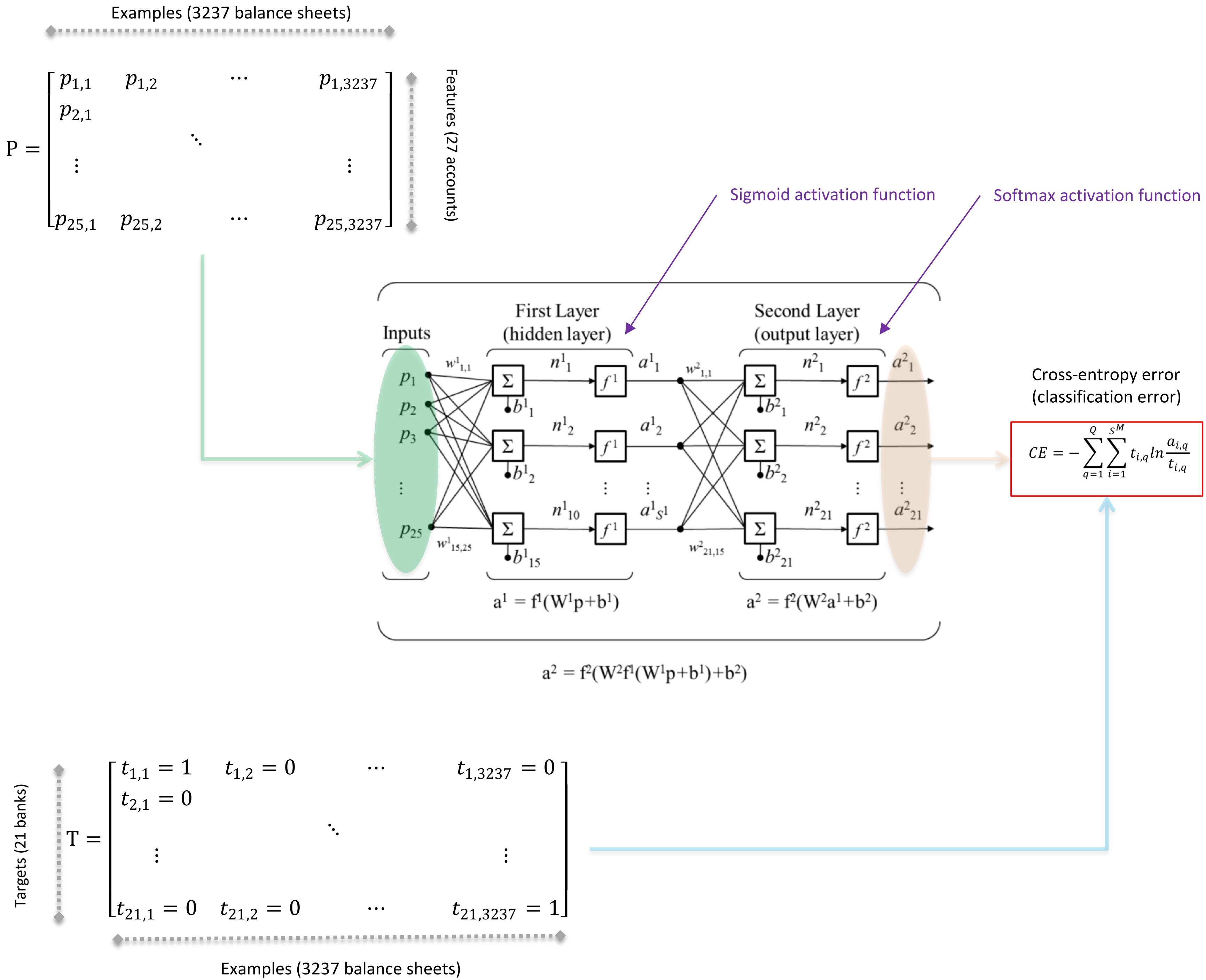
Where did we start?

A bank's balance sheet as a 27x1 pixels image






DAVIVIENDA
BBVA
Banco AV Villas
Banco de Bogotá
Bancolombia
Banco Caja Social



Examples (3237 balance sheets)
P =
Features (27 accounts)
Sigmoid activation function
Softmax activation function
First Layer
(hidden layer)
Second Layer
(output layer)
Inputs
a¹ = f¹(W¹p+b¹)
a² = f²(W²a¹+b²)
a² = f²(W²f¹(W¹p+b¹)+b²)
Cross-entropy error
(classification error)
CE = −∑∑ t_{i,q} ln (a_{i,q} / t_{i,q})
Targets (21 banks)
T =
Examples (3237 balance sheets)

Where did we start?

Set	Misclassification (Average and standard deviation, %) 5 neurons	10 neurons	**15 neurons**	20 neurons	25 neurons
Training In-sample	19.75% [15.37%]	3.41% [9.84%]	**0.61% [0.43%]**	0.15% [0.29%]	0.10% [0.23%]
Validation	20.99% [15.23%]	4.86% [9.87%]	**1.64% [0.81%]**	1.00% [0.70%]	0.91% [0.72%]
Test	21.53% [15.44%]	5.19% [9.86%]	**1.72% [0.80%]**	1.23% [0.66%]	0.94% [0.63%]

Table 2. Overall average results of the artificial neural network after training with cross-validation early-stopping. The average and standard deviation (in brackets) is estimated on 100 independent training processes.

Where did we start?

Whose Balance Sheet Is This? Neural Networks for Banks' Pattern Recognition

Carlos León
Banco de la República and Tilburg University; e-mail: cleonrin@banrep.gov.co
José Fernando Moreno
Banco de la República
Jorge Cely
Banco de la República

Abstract
The balance sheet is a snapshot that portrays the financial position of a firm at a specific point in time. Under the reasonable assumption that the financial position of a firm is unique and representative, we use a basic artificial neural network pattern recognition method on Colombian banks' 2000–2014 monthly 25-account balance sheet data to test whether it is possible to classify them with fair accuracy. The results demonstrate that the chosen method is able to classify out-of-sample banks by learning the main features of their balance sheets, and with great accuracy. The results confirm that balance sheets are unique and representative for each bank, and that an artificial neural network is capable of recognizing a bank by its financial accounts. Further developments fostered by our findings may contribute to enhancing financial authorities' supervision and oversight duties, especially in designing early-warning systems.

Keywords
supervised learning, machine learning, artificial neural networks, classification

1 Introduction

The balance sheet shows the financial position of a firm at a particular moment in time; it is a valuable source of information about the past performance of a firm, and a starting point for forecasts of future performance (Chisholm, 2002). Investors, creditors, and other decision makers use balance sheets to assess the overall composition of resources, the constriction of external obligations, and the firm's flexibility and ability to change to meet new requirements (Kaliski, 2001).

In the banking industry, according to the *Core Principles for Effective Banking Supervision* (Basel Committee on Banking Supervision, 1997, 1998), balance sheets are among the minimum periodic reports that banks should provide to supervisors to conduct effective supervision and to evaluate the condition of the local banking market. As highlighted by Mishkin (2004), traditional supervisory examination has focused on the assessment of banks' balance sheets, and has been related to economic activity and the advent of financial crisis.

In this sense, a bank's balance sheet may be regarded as a representative source of information. Each bank's past decisions and performance, its business model, and its views about the future are condensed in its balance sheet. Consequently, it is reasonable to assume that the balance sheet may be considered a snapshot of a bank; a unique and characteristic combination of financial accounts (i.e. the elements of financial statements) that not only allows for assessing a bank's financial stance, but that also differentiates it from its peers.

Under the reasonable assumption that a bank's balance sheet is unique and representative, we use a basic artificial neural network pattern recognition method on Colombian banks' 2000–2014 monthly 25-account balance sheet data to test whether it is possible to classify them with fair accuracy. Analogous to widespread facial recognition on individuals' photographs or fingerprint scanning, we aim to classify banks by examining their accounting snapshots.

Based on the well-documented effectiveness of artificial neural networks as classifiers (see Wu, 1997; Zhang et al., 1999; McNelis, 2005; Han and Kamber, 2006), and on the presumed informational content of balance sheets, we expect to find a model able to classify out-of-sample balance sheets to their corresponding bank with

The opinions and statements in this paper are the sole responsibility of the authors and do not represent either those of Banco de la República or of its Board of Directors. J.F. Moreno co-authored the original version of this paper while affiliated to Banco de la República. Comments and suggestions from Hernando Vargas, Clara Machado, Freddy Cepeda, Fabio Ortega, and other members of the technical staff of Banco de la República are appreciated. Any remaining errors are the authors' own.

(*) In 2017

Contents

The new research & development project

- Classifying balance sheet data is interesting. But …
  - It is low-frequency and lagged data. *Decision making and warning signals with monthly statements that are available with a 2-month lag?*
  - Limited granularity. *What about linkages among financial institutions?*
  - Accounting practices. *How truthful and precise are the data?*
  - Informational content. *Is it meaningful about financial institutions' market activity?*
- Classifying payments data is even more interesting because…
  - It is high frequency and timely. *Intraday, at the end of the day (may be in real time).*
  - Granular. *By type of transaction, by counterparty*.
  - Accounting practices-free. *Most transactions are reported by financial infrastructures.*
  - Footprint of market activity. *Payments are the outcome of most market activities.*

The new research & development project

- Then, new R&D project is “Identifying the anomalous behavior of large-value payment system participants with artificial neural networks”.
- Expected outcome:
  - Research paper: identifying anomalies in financial institutions’ behavior in the large-value payment system
  - Development of oversight tool: using out-of-sample anomalies as flags for oversight purposes (i.e. prioritizing oversight efforts).
- Methodology: basic feed-forward, backpropagation ANN.
- Data source: large-value payments system dataset.
- Software:
  - Matlab and Matlab’s Machine Learning Apps for design, research and prototyping.
  - In the future, operation will run in a server with Teradata-Python.
- Human resources: monitoring leader, engineer, **senior researcher**.

Contents

The dataset

- Large-value payments system data –in Teradata data warehouse.
  - Extracted manually in an Excel file (~20,2 Mb), to process in Matlab.
  - In the future, operation will be automatic in an external server, to process in Python.
- 113 features for each financial institution, for each day, comprising…
  - Net and gross payments, overall and by type of payment (top-20 types).
  - Liquidity sources (e.g. CB repos, multilateral liquidity savings, incoming payments).
  - Concentration in/out payments by type, by counterparty (Herfindahl-Hirschman Index).
  - Network centrality (e.g. in/out degree & strength, authority, hub, PCA-overall centrality).
  - Simulated impact on overall LVPS' liquidity (decrease in overall payments due to failure to make discretionary payments)
- Features normalized [0,1], for each financial institution, for each date.
- ~200 types of non-top-20 transactions which are collapsed into "other"

The dataset

- **112 financial institutions (targets)**
  - Banks (25)
  - Other banking (24)
  - Securities broker-dealer firms (19)
  - Insurance firms (14)
  - Private pension funds (4)
  - Investment funds (26)
- **243 days** (i.e. 1 year)
- **That is, about 24,400 rows and 115 columns** (20,2Mb .xlsx, 14.5Mb .csv)

Contents

The model

- How do we address our R&D project? Artificial Neural Networks (ANN)
  - Effective classifiers, better than classical statistical methods (Wu (1997), Zhang et al. (1999), McNelis (2005), and Han & Kamber (2006))
  - No assumptions about the statistical porperties of the data (Zhang et al. (1999), McNelis (2005), Demyanyk & Hasan (2009), and Nazari & Alidadi (2013))
  - Able to deal with non-linear relationships between factors in the data (Bishop (1995), Han & Kamber (2006), Demyanyk & Hasan (2009), Eletter et al. (2010), and Hagan et al. (2014))
- But... ANN have been criticized because results are opaque and they lack interpretability –*black box criticism* (Han & Kamber (2006), Angelini et al. (2008), and Witten et al. (2011)) ... do we care?

The model

- *Black box criticism* comes from a desire to tie down empirical estimation with an underlying economic theory (McNelis, 2005)
- We do not care about the *black box criticism* because we have no underlying economic theory to test
- This is *predictive modeling* –not *explanatory modeling* (see Shmueli, 2010)

Explanatory Modeling	Predictive Modeling
• The aim is to test a causal theory (traditional econometrics) • Requires building an underlying causal structure (a theoretical prior) • Need to work on expected role of variables	• The aim is to predict or classify successfully • No need to build an underlying causal structure (a theoretical prior) • No need to delve into the expected role of the variables
Econometrics	**Machine Learning**

The model

- ANNs are networks of interconnected artificial neurons, with the weights of those connections resulting from a learning process that attempts to minimize the prediction/classification error of the input-output function
- The central idea of ANNs is to extract linear combinations of the inputs as derived features, and then model the output (i.e. the target) as a nonlinear function of these features. (Hastie et al., 2013)
- The simplest case is the *feed-forward* ANN (our choice for what follows).
- Other ANNs cases are more complex but may open new ways to solve more complex problems (e.g. recurrent ANNs, convolutional ANNs, reinforcement ANNs). We do not consider them.

Examples (**24,432 payment samples**)

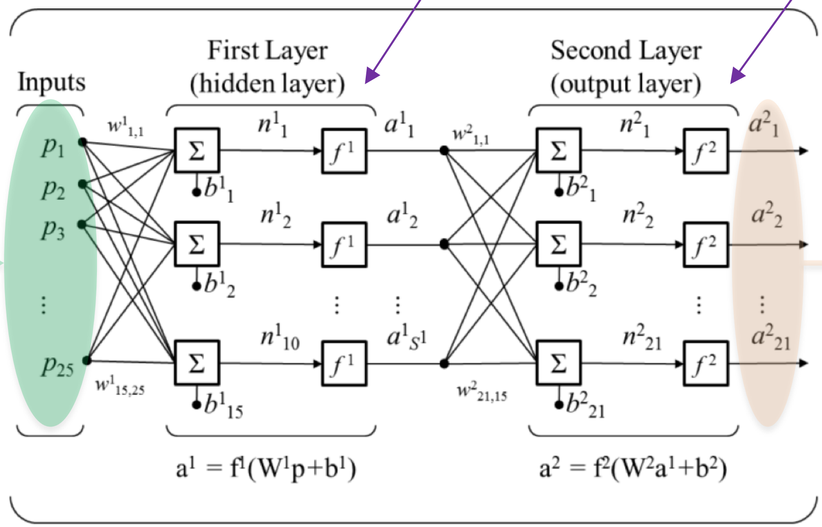
$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,3237} \\ p_{2,1} & & & \\ \vdots & & \ddots & \vdots \\ p_{25,1} & p_{25,2} & \cdots & p_{25,3237} \end{bmatrix}$$

Features (113)

The model

Sigmoid activation function

Softmax activation function



$$a^2 = f^2(W^2 f^1(W^1 p + b^1) + b^2)$$

Cross-entropy error
(classification error)

$$\boxed{CE = -\sum_{q=1}^{Q}\sum_{i=1}^{S^M} t_{i,q} \ln \frac{a_{i,q}}{t_{i,q}}}$$

Targets (**112 financial institutions**)

$$T = \begin{bmatrix} t_{1,1} = 1 & t_{1,2} = 0 & \cdots & t_{1,3237} = 0 \\ t_{2,1} = 0 & & & \\ \vdots & & \ddots & \vdots \\ t_{21,1} = 0 & t_{21,2} = 0 & \cdots & t_{21,3237} = 1 \end{bmatrix}$$

Examples (**24,432 payment samples**)

The model

- Training: *adjusting W and b to attain an input-output relationship target under the chosen transfer functions for a set of examples.*
- How do we train? *Backpropagation: W and b are modified in backwards direction, from the output layer.*
- How do we avoid overfitting? *Early stopping with cross-validation: Halt the minimization process before the complexity of the solution inhibits its generalization capability.*

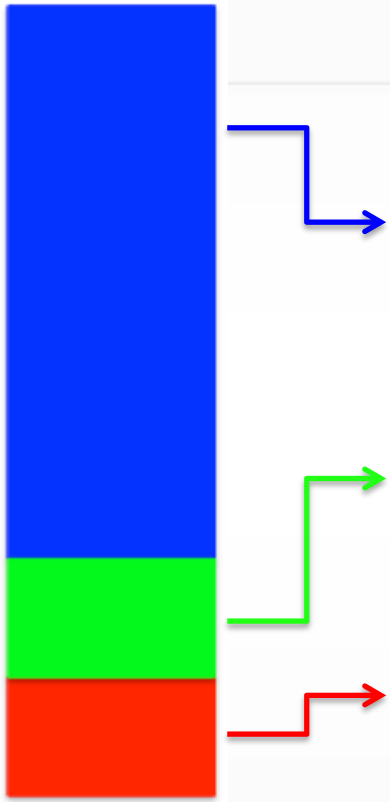
The goal is not to memorize the training data, but to model the underlying generator of the data (Bishop, 1995)

The model

Training dataset
(70%)

Validation dataset
(15%)

Test dataset
(15%)



The training set is used to minimize the error between the prediction and the actual target value

The validation dataset is used simultaneously (as the neural network is trained) to check how the estimated parameters fit out-of-sample data. When validation error starts to increase (i.e. overfitting starts), the training stops.

The error obtained on the test dataset is used to check the future performance of the artificial neural network on out-of-sample data, i.e. its generalization capability.

The model

- Some heuristics
  - How to decide the number of layers? Often a single hidden layer is all that is necessary (see Zhang et al. (1999), Witten et al. (2011)). So, we start with the simplest ANN.
  - How to decide the number of neurons? We try several numbers, from 20 to 150, in increments of 5. After no clear improvement in error is attained, we stop (~60 neurons).
  - As the result is dependent on initial values of parameters, we run several independent training processes. We compare average results.

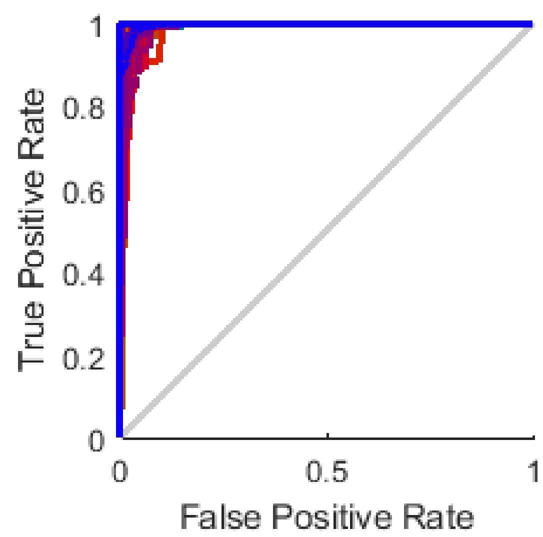
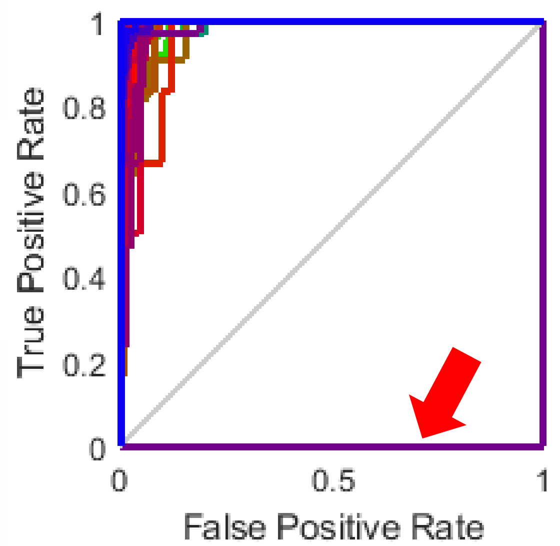
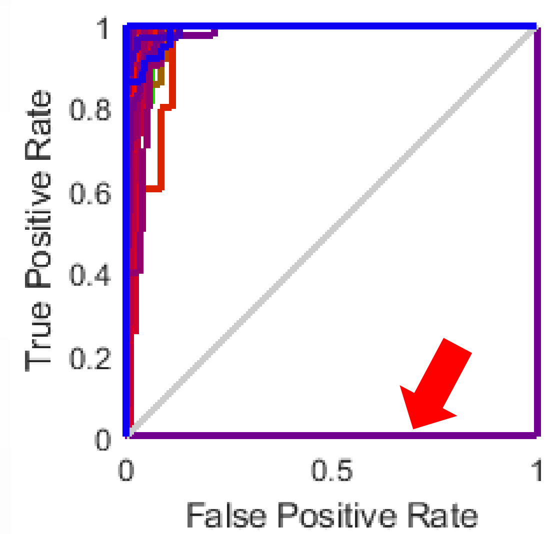
Contents

Main results

Set	Samples	Error (Misclassification)
Training	17,191 (70%)	0.09
Validation	3,684 (15%)	0.12
Test	**3,684 (15%)**	**0.11**

Misclassification error seems promising!

Main results

Training	Validation	Test
		

Performance seems promising… but it is clear the ANN struggles with some financial institutions in validation and test datasets

Main results

Group of financial institution	Contribution by number	Test Error (misclassification)	
		WRT Total	WRT Group
Private Pension Funds	0.03	0.00	0.02
Banks	0.26	0.01	0.04
Securities broker-dealer firms	0.19	0.01	0.05
Investment funds	0.24	0.02	0.08
Other banking (1)	0.03	0.00	0.11
Other banking (2)	0.12	0.02	0.18
Other banking (3)	0.03	0.01	0.21
Insurance firms	0.10	0.04	0.41
ALL FINANCIAL INSTITUTIONS	**1.00**	**0.11**	-

Good performance for the most Important (e.g. SIFIs)

But some are difficult to classify

We think we know why:

- **Those financial institutions do not participate actively in the LVPS, thus…**
- **The ANN struggles to identify a pattern in their behavior –perhaps, there is no pattern!**

Contents

Pending challenges

- Getting your feedback about…
  - Is the implementation sound and fit for our task?
  - Are results promising –as we think they are?
- Some questions regarding how to make this operational for oversight purposes:
  - What is an "optimal" time series for training? We use one year…
  - How often should we train not to miss new developments in the market but to avoid normalizing anomalous behavior? (Is there an "optimal" re-training frequency?)
  - Anomalies are those institutions that are difficult to classify with new data, and that should be examined (i.e. prioritizing oversight efforts). Then…
    - What is a good benchmark to determine an anomaly?
    - Our preliminary idea is *distance to mean test error* for that institution in the training phase.
    - We discarded a one-threshold-fits-all benchmark (e.g. all errors higher than some error)
    - What else?

Identifying the anomalous behavior of large-value payment system participants with artificial neural networks. [Colombia use case]