Artificial Neural Networks

Alvaro J. Riascos Villegas
University of los Andes and Quantil

July 6 2018

Contenido

Introduction

- ANN is a highly nonlinear and tractable machine learning algorithm.
- It is an universal approximator.
- Advances in calibration methodologies in terms of speed and enormous success for solving pattern recognition problems such as image recognition, voice translation, etc.
- The success solving these tasks has put ANN and Deep Neural ANN at the center stage of research and industry applications.

Introduction

- ANN is a highly nonlinear and tractable machine learning algorithm.
- It is an universal approximator.
- Advances in calibration methodologies in terms of speed and enormous success for solving pattern recognition problems such as image recognition, voice translation, etc.
- The success solving these tasks has put ANN and Deep Neural ANN at the center stage of research and industry applications.

Introduction

- ANN is a highly nonlinear and tractable machine learning algorithm.
- It is an universal approximator.
- Advances in calibration methodologies in terms of speed and enormous success for solving pattern recognition problems such as image recognition, voice translation, etc.
- The success solving these tasks has put ANN and Deep Neural ANN at the center stage of research and industry applications.

Basic ANN: Feedforward Neural Net

- The most basic ANN is called Feedforward Neural Net or Multilayer Perceptor.
- The logistic model is a special case, you already know the simplest ANN!
- These networks are difficult to optimize globally.
- A key idea to carry a computationally efficient optimization is the idea of backpropagation.

Basic ANN: Feedforward Neural Net

- The most basic ANN is called Feedforward Neural Net or Multilayer Perceptor.
- The logistic model is a special case, you already know the simplest ANN!
- These networks are difficult to optimize globally.
- A key idea to carry a computationally efficient optimization is the idea of backpropagation.

Basic ANN: Feedforward Neural Net

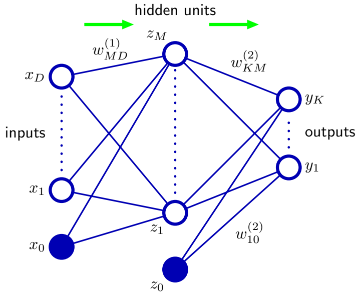
- The most basic ANN is called Feedforward Neural Net or Multilayer Perceptor.
- The logistic model is a special case, you already know the simplest ANN!
- These networks are difficult to optimize globally.
- A key idea to carry a computationally efficient optimization is the idea of backpropagation.

Basic ANN: Feedforward Neural Net

- The most basic ANN is called Feedforward Neural Net or Multilayer Perceptor.
- The logistic model is a special case, you already know the simplest ANN!
- These networks are difficult to optimize globally.
- A key idea to carry a computationally efficient optimization is the idea of backpropagation.

Basic ANN: Two layers

- A two layer ANN can be represented by the following graph. There is one hidden layer and one output layer. Each layer may have an arbitrary number of units (i.e., neurons)



Basic ANN: Two layers

- Assume you have D input variables: $\{x_1, ..., x_D\}$ and M output variables $\{y_1, ..., y_M\}$.
- Let h_1, h_2 be activations functions (h_2 is the activation function of the output layer).
- The following equation describe a two layer feed-forward ANN:

Basic ANN: Two layers

$$a_j^{(1)} = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}$$

$$z_j^{(1)} = h_1(a_j^{(1)})$$

$$a_j^{(2)} = \sum_{i=1}^{M} w_{ji}^{(2)} z_i^{(1)} + w_{j0}^{(2)}$$

$$z_j^{(2)} = h_2(a_j^{(2)})$$

where $w_{j0}^{(1)}, w_{j0}^{(2)}$ represent the bias (constant) in each layer.

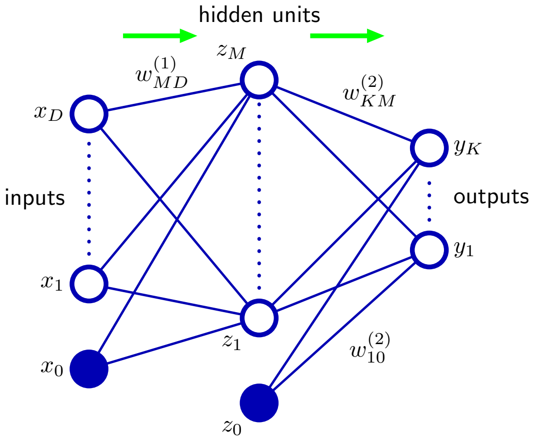
Basic ANN: Two layers

- Let $z_i^{(0)} = x_i, y_j = z_j^{(2)}$
- If we define the additional variables $x_0 = 1, z_0^{(1)} = 1$ and $z_0^{(2)} = 1$, we can rewrite the equations describing the ANN as :

$$y_k(x, w) = h_2(\sum_{j=0}^{M} w_{kj}^{(2)} h_1(\sum_{i=0}^{D} w_{ji}^{(1)} x_i))$$

Basic ANN: Two layers

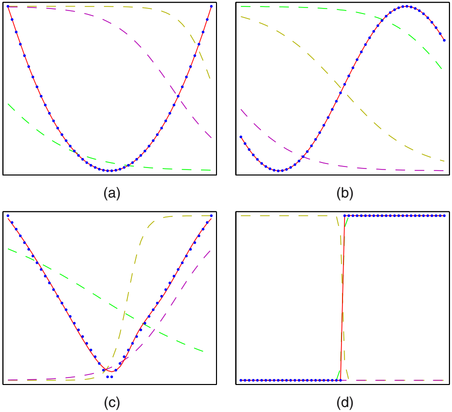
- This two layer terminology reflects the fact that we have to estimate two set of parameters (the linear weights at each layer).



Basic ANN: Universal Approximation Property

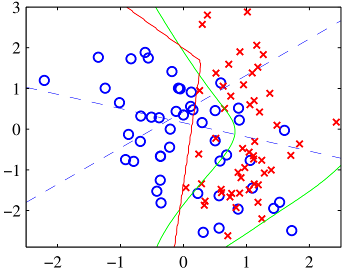
- An ANN with two layers and linear output activation functions can approximate any continuos function on a bounded domain (more precisely, on a compact domain) with enough neurons.
- This is true for many activation functions in the hidden layer (though not for polynomials).

Approximation properties examples (2 layers, 3 neurons)



(a) (b) (c) (d)

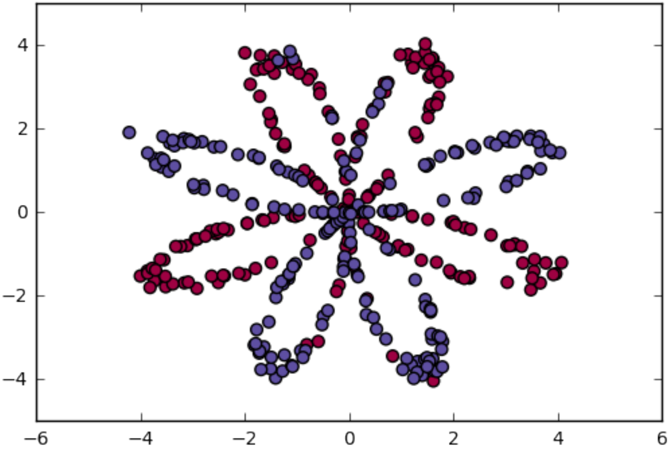
- Simulated data 50 blue dots. ANN with 2 layers, 3 neurons, tanh activation in the hidden layer and linear activation in output layer. Dotted lines show the results of the 3 neurons.



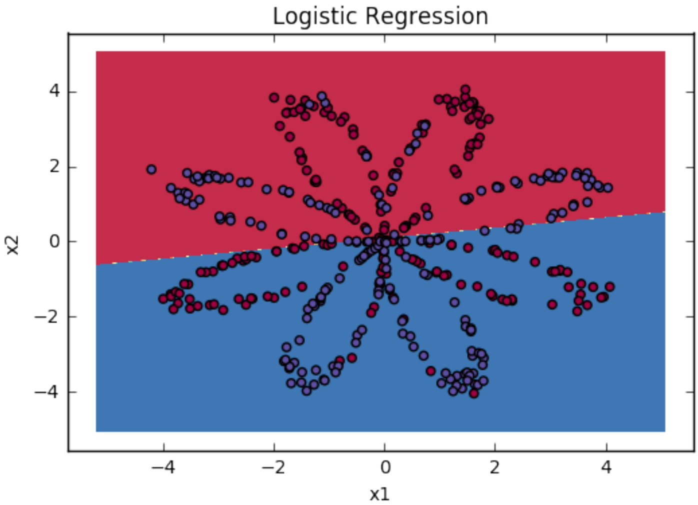
- ANN binary classification problem with two layers and two neurons. Dotted lines are the classification hypersurfaces of each neuron.
- Red line de ANN classification result and green line, Bayesian classifier.

Contenido

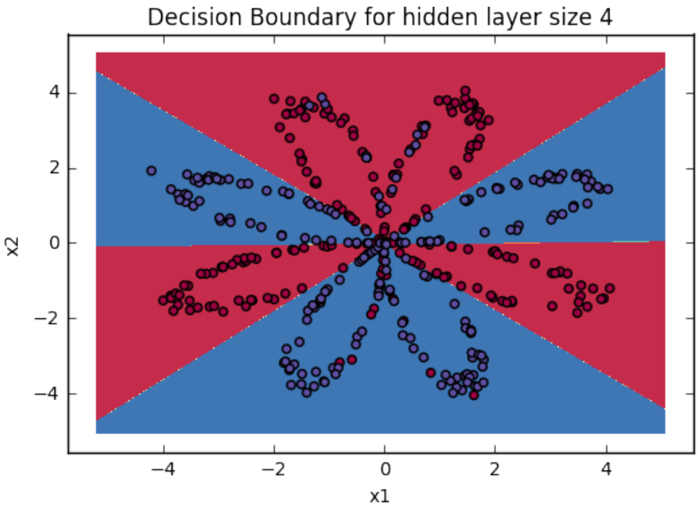
ANN in Action: Data



ANN in Action: Logistic (One layer with sigmoid activation)



ANN in Action: Two layers



Contenido

Final Observations

- ANN can be trained using standard techniques (gradient descent, etc.). The key idea is how to calculate derivatives of the loss function with respect to parameters: use the structure of the net and chain rule (i.e., backpropagation).
- Deep ANN are ANN with many layers and probably, many neurons per layer.
- Adding layers allows for a simpler representation of any continuous representation.
- **Learning data representations** (features) is possible by extracting intermediate outputs from hidden layers.

Contenido

Application: Nightlight

- Combining satellite imagery and ML to predict poverty. Jean et.al. Science, 2016.

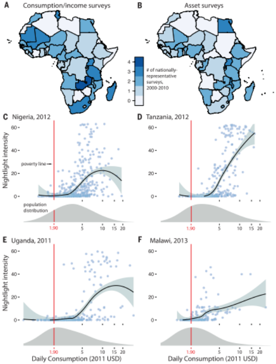


Fig. 1. Poverty data gaps. **(A)** Number of nationally representative consumption surveys occurring in each African country between 2000 and 2010. **(B)** Same as (A), for DHS surveys measuring assets. **(C** to **F)** Relationship between per capita consumption expenditure (measured in U.S. dollars) and nightlight intensity at the cluster level for four African countries, based on household surveys. Nationally representative share of households at each point in the consumption distribution is shown beneath each panel in gray. Vertical red lines show the official international extreme poverty line ($1.90 per person per day), and black lines are fits to the data with corresponding 95% confidence intervals in light blue.

Application: Learning Representations

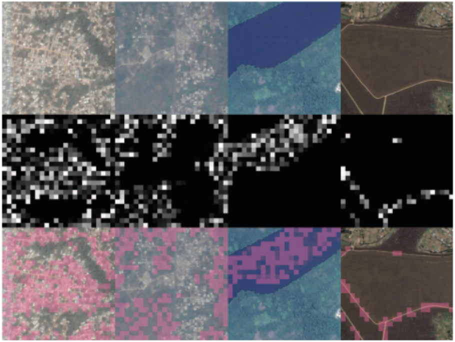


Fig. 2. Visualization of features. By column: Four different convolutional filters (which identify, from left to right, features corresponding to urban areas, nonurban areas, water, and roads) in the convolutional neural network model used for extracting features. Each filter "highlights" the parts of the image that activate it, shown in pink. By row: Original daytime satellite images from Google Static Maps, filter activation maps, and overlay of activation maps onto original images

Application: Prediction

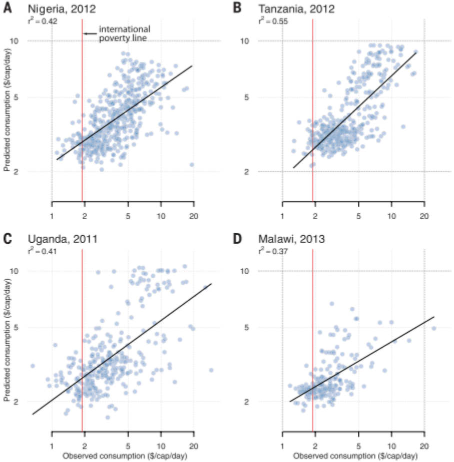


Fig. 3. Predicted cluster-level consumption from transfer learning approach (*y* axis) compared to survey-measured consumption (*x* axis). Results are shown for Nigeria (**A**), Tanzania (**B**), Uganda (**C**), and Malawi (**D**). Predictions and reported r^2 values in each panel are from fivefold cross-validation. Black line is the best fit line, and red line is international poverty line of $1.90 per person per day. Both axes are shown in logarithmic scale. Countries are ordered by population size.