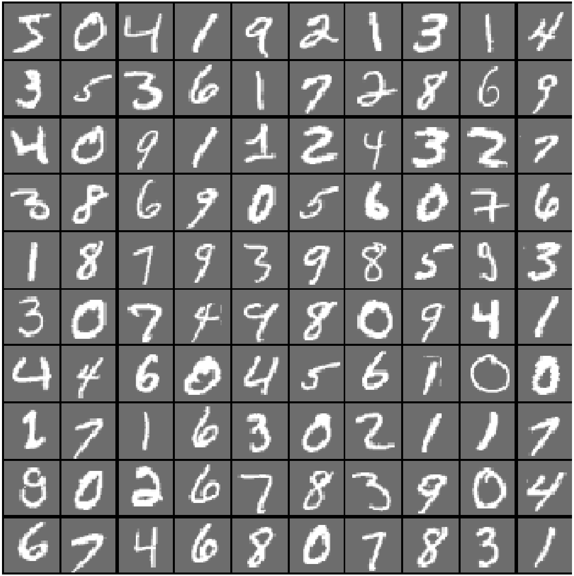
Machine Learning Applications

Otaviano da Cruz Neto

Universidade Federal Fluminense - ICEX VR

16 de Novembro de 2018

Recognize Handwritten Digits



Introdução

- **Individual Data ($X^{(i)}$,$Y^{(i)}$)**

$$X^{(i)} = [x_1, x_2, \cdots, x_m] \tag{1}$$

$$Y^{(i)} = y^{(i)} \tag{2}$$

- **Data Sample(X e Y)**

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_m^{(1)} \\ 1 & x_1^{(2)} & \cdots & x_m^{(2)} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_1^{(N)} & \cdots & x_m^{(N)} \end{bmatrix} \tag{3}$$

Introduction

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} \tag{4}$$

- **Weight Vector**

$$W = [w_0, w_1, w_2, \cdots, w_m] \tag{5}$$

- **Linear Hypotheses**

$$h(w) = XW^T \tag{6}$$

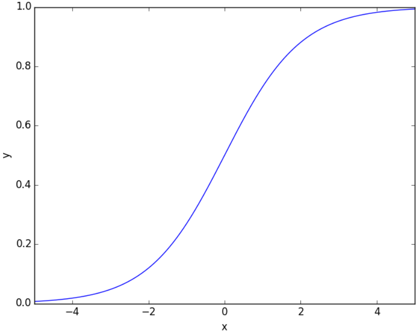
Logistic Regression

- **Classification**
  The classification is a regression problem which associates a discrete values for each data. In this case we will analyze for binary classification.
- **Logistic Regression** The logistic regression uses probability of each data sample be classified by two discrete values. The logistic regression hypotheses is described by Sigmoid Function,

$$h_w(X^{(i)}) = \frac{1}{1 + e^{-X^{(i)}W^T}} \tag{7}$$

Logistic Regression

- **Sigmoid Function**



Logistic Regression

- **Decent Gradient**
  To use this technique is necessary calculate the derivative of Sigmoid Function,

$$\frac{d}{dW} h_W(X^{(i)}) = \frac{1}{1 + e^{-X^{(i)}W^T}} \left(1 - \frac{1}{1 + e^{X^{(i)}W^T}}\right) \tag{8}$$
$$= h_W(X^{(i)})(1 - h_W(X^{(i)}))$$

Logistic Regression

For each classification is valid,

$$P(Y^{(i)} = 1|X^{(i)}, W) = h_W(X^{(i)})$$
$$P(Y^{(i)} = 0|X^{(i)}, W) = 1 - h_W(X^{(i)})$$

This conditional probability have a functional expression :

$$p(Y^{(i)}|X^{(i)}, W) = (h_W(X^{(i)}))^{Y^{(i)}}(1 - h_W(X^{(i)}))^{1-Y^{(i)}} \tag{9}$$

Logistic Regression

The probability for N numbers of IID (Independent and Identical Distribution) sample is defined by,

$$L(W) = \prod_{i=1}^{N} (h_W(X^{(i)}))^{Y^{(i)}} (1 - h_W(X^{(i)}))^{1-Y^{(i)}} \tag{10}$$

Is necessary maximize as log(L(W)):

$$\sum_{i=1}^{N} Y^{(i)} \log h_W(X^{(i)}) + (1 - Y^{(i)}) \log(1 - h_W(X^{(i)})) \tag{11}$$

Implying in the algorithm,

$$W_{t+1} = W_t + \alpha(Y^{(i)} - h_W(X^{(i)}))X^{(i)} \tag{12}$$

Example

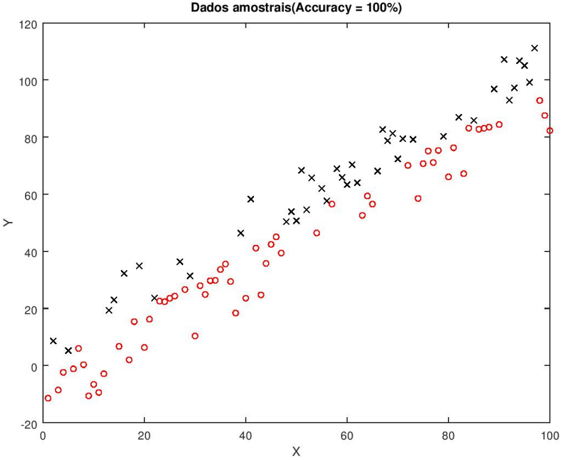


Figura 1: Binary Data Sample.

Example

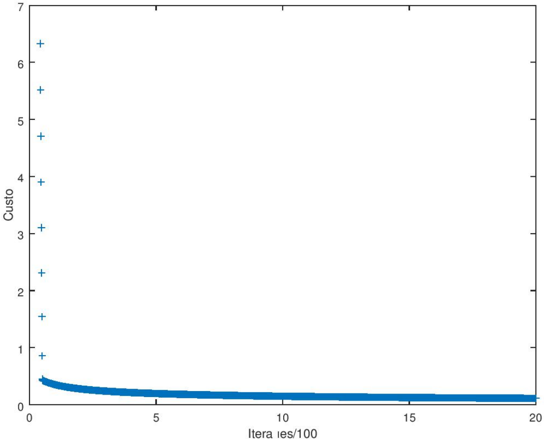


Figura 2: Cost function for number of iteration.

Example

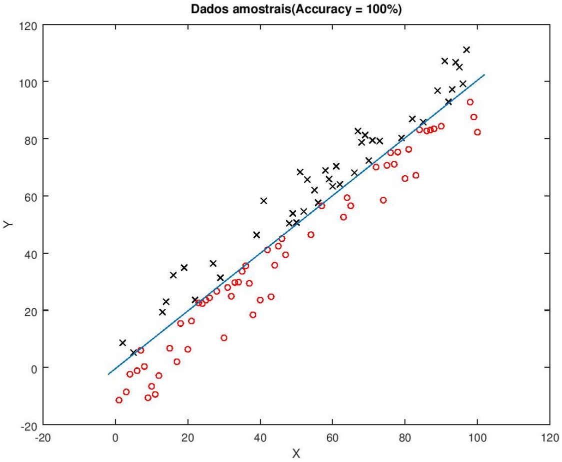


Figura 3: Binary Data Sample and Decision Boundary.

Overfitting

Overfitting is a term used for describe a statistical model whose efficiency is very good in sample but out of sample is worse due to present a random factor and measurement errors. To fix this is necessary to apply the Regularization whose purpose is to limit the sum of weights. In the Logistic Regression is applied in Decent Gradient as,λ is regularization parameter,

$$w_{t+1}^{0} = w_{t}^{0} + \alpha x_0'(Y - H_{w_0}(x_0)) \tag{13}$$

$$w_{t+1}^{(i)} = w_{t}^{(i)}(1 - \alpha\frac{\lambda}{N}) + \frac{\alpha}{N}X^{(i)\prime}(Y - H_w(X^{(i)})) \tag{14}$$

Multi-classification

- **One Vs All** :

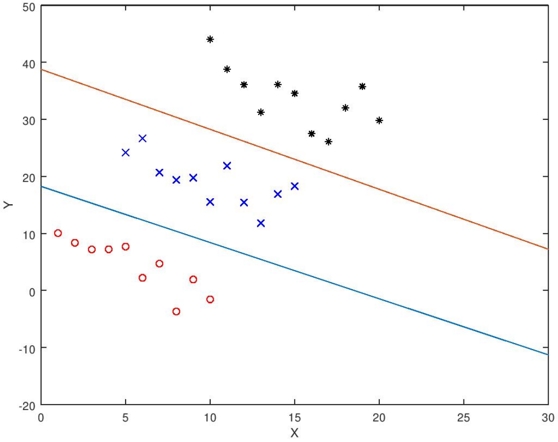
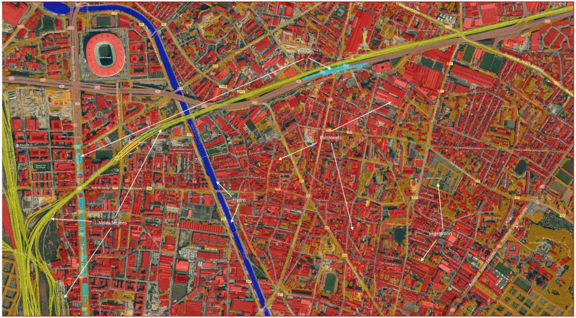


Figura 4: Data Sample and Boundary Decision.

Computerized Measures of Visual Complexity



Computerized Measures of Visual Complexity

Types of Images :

- Abstract Artistic, Abstract Non-Artistic, Representational Artistic, Representational Non-Artistic, pictures of some natural situations and human-made objects.

Separation:

- **Experiment** : Uses some tools and compression formats of images(JPEG and Fractal Compression, edge filter, HSV channel) to extract important information about image complexity, based on human experience and some information like average complexity and medium standard deviation associated.

Computerized Measures of Visual Complexity

- **Machine Learning** : Utilize machine learning techniques (Artificial Neural Network) and data (average, standard deviation) to recognize some patterns at learning image complexity with efficiency.

Features

The article utilize only 2 features (Compression Method and Edge Dispersion [Canny, Sobel]) because at high correlation.

- **Compression Method** : The compression method is a indicator of image complexity because of algorithm routine, that routines identify recurring pixels to reduce image file size. We use this to observe how predictable is the image.
- **Edge Dispersion** : The Image's Edge Dispersion is a good contribution for research area because of the high correlation.

Statistical Analyses

The correlation is defined by Pearson Correlation Coefficient(PCC).

- **PCC**: relates two variables to calculate the linear correlation between them.

$$PCC = \frac{\sum_{j=1}^{N}(f(x_j) - \bar{f(x)})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{N}(f(x_j) - \bar{f(x)})^2 \sum_{j=1}^{N}(y_j - \bar{y})^2}} \tag{15}$$

Algorithm

Utilizes how is the dependency between Edge Filters (Canny, Sobel), average, HSV Channels, Standard Deviation.

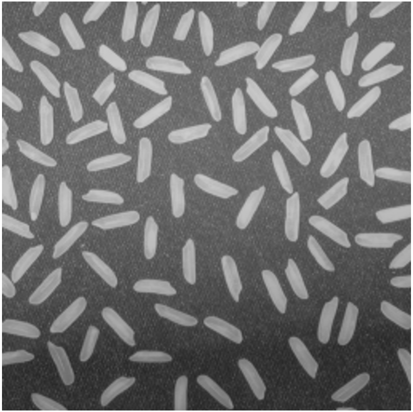


Figura 5: Image without edge filter.

Results

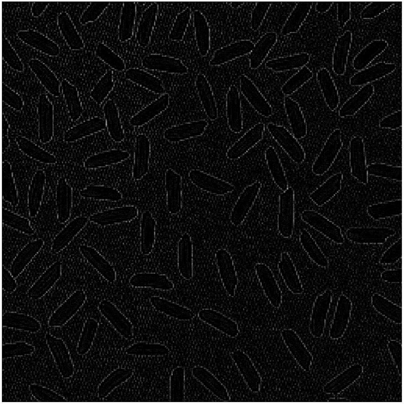


Figura 6: Image with edge filter.