



# Pattern Recognition

## *Last exam*

Iván A. Trujillo A.

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

## 1 Problem 1

Choose two methodological approaches to solve a classification problem in the context of IoT to images recognition. suppose that the problem have two classes  $-1$  and  $1$ , We have the  $N = 1000$  with a  $\mathbf{X}^5$ .

In this problem we could outline two criteria to the classification task, we could uses a black and a white box model, the first could be used to gain prediction over a new data set (no fit or training data), and the second to make inferences about the production process and gain readability of the model.

The following two methodologies are proposed: **Perceptron and decision trees**, specifically proposed by twice reasons: are complementary, the first work a black box and the second is a white box model, the first is limited to be a linear classifier and the last overcome not linearity.

### 1.1 Perceptron

**Theorem** if exist a linear frontier then the perceptron converge.

The previously announced theorem, guaranteed a convergence in the model if only if the both classes could be separated linearly.

In the implementation of perceptron it is necessary define a boundary over the iterations (epochs) due that if the problem not could be reduce a linear classifier then the perceptron is a infinite loop, this problem due of the updates of parameters is according the following rule:

$$\Delta w_i = (y_i - \hat{y}_i) = 0 \quad (1)$$

where  $y_i$  is the real observed data, and  $\hat{y}_i$  is the predicted class.

when the  $y_i = -1$  and  $\hat{y}_i = 1$  then  $\Delta w = -2$ , in otherwise  $y_i = 1$  and  $\hat{y}_i = -1$  then  $\Delta w = 2$ .

Then when there are mistakes

$$\varphi(w_{i+1}^t \mathbf{x}_i) = \varphi((w_i + \Delta w_i)^t \mathbf{x}_i) = y_i \quad (2)$$

This mean that weights for the vector of features of the sample  $i$  are update to predict the correct class.

$$w_{i+1} = w_i + \eta \Delta w_i \mathbf{x}_i \quad (3)$$

The cost function used in hard sense not guaranteed and optimal solution:

$$J(w)_{hard} = \sum_{i=1}^n = \max(-y_i \hat{y}_i, 0) \quad (4)$$

$J$  only count the number of mismatches. However this function not is differentiable.

## 1.2 Decision tree

Work as a if-else system ruled based on the bias-variance problem, that tend to overfitting data, and reduce the accuracy in new data, however this could be reduced with pruning, or fixing the depth of the algorithm, given the split of data could be very deep.

Decision trees is readable due of is based in order, the uncertainty, for instance suppose that you flip a coin and maximum entropy is reached when the success have  $p = \frac{1}{2}$  due we do not have additional information about what side lands.

The event with small probability of occurrence give more information than those with a major probability.

$$I(x) = \log \frac{1}{p(x)} = -\log p(x) \quad (5)$$

the average surprisal:

$$\sum p(x)I(x) \quad (6)$$

we can rewrite as:

$$H(x) = -\sum p(x)\log p(x) \quad (7)$$

suppose that we have a variable  $x \sim N(\mu, \sigma)$  do not have more entropy than a variable  $y \sim U(a, b)$  given that each 'character' have the same probability of occur.

The main reason for the readability of the trees is that is constructed in the reduction of this entropy in all tree, that is got with each node (averaging in each class the order of information)

***Another important reason of uses these complementary methodologies, is one is for perceptron the variable scale affect the perform of the model, and the uses of categorical variable requiere preprocessing while in decision trees not.***

## 2 Problem 2

A balanced model that not affect by the size of classes.

### 2.1 Logistic Regression

Constructed over the sigmoid function, to modeling population growth in limited context, this model gain importancen in classification given three mainly reasons:

- is limited to the  $[0,1]$  interval
- The effect of a feature over the log of odds not is constant.

Logistic regression have a good interpretability, not have affect by the unbalaced number of classes as Decission tree, and dont tend to overlap the frontier decision as perceptron, also not suppose stricted assumptions about normality and equal variance as Multiple Discriminant Analysis.

$$F(X)=P(X \leq x) = \int_{-\infty}^{\infty} f(t)dt; -\infty < x < \infty$$

$$P(Z) = \frac{e^z}{1 + e^z}$$

$$\frac{P_z}{1 - P_z} = e^z$$

$$\ln \left( \frac{P_z}{1 - P_z} \right) = Z = \beta_1 + \beta_2 x_i$$