

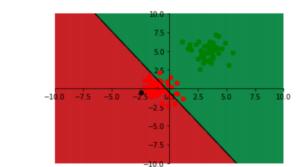
Deep Learning Perceptron

Tiago Vieira February 5, 2023

Motivation – ML w/ Perceptron



Motivation



²Simple perceptron example using NumPy

Binary Classification and Linear Regression Problems

- \blacktriangleright In the binary classification problem, each training pair (\overline{X},y) contains feature variables $\overline{X}=(x_1,\dots x_d)$, and label y drawn from $\{-1,+1\}$.

 Example: Feature variables might be frequencies of words in an email, and the class variable might be an indicator of spam. Given labeled emails, recognize incoming spam. ▶ In linear regression, the *dependent* variable *y* is real-valued.
- Feature variables are frequencies of words in a Web page, and the dependent variable is a prediction of the number of accesses in a fixed period. ▶ Perceptron is designed for the binary setting.

The Perceptron: Earliest Historical Architecture

 $\begin{array}{c} x_3 \longrightarrow \bigcirc \begin{array}{c} w_3 \\ \\ w_4 \end{array} \longrightarrow \begin{array}{c} \\ \\ w_5 \end{array}$

▶ The d nodes in the input layer only transmit the d features $\overline{X} = [x_1 \dots x_d]$ without performing any computation. lackbox Output node multiplies input with weights $\overline{W} = [w_1 \dots w_d]$ on incoming edges, aggregates them, and applies sign activation: $\hat{y} = \operatorname{sign}\{\overline{W} \cdot \overline{X}\} = \operatorname{sign}\{\sum^u w_j x_j\}$

What is the Perceptron Doing?

 \blacktriangleright Tries to find a $\it linear\ separator\ \overline{W}\cdot \overline{X}=0$ between the two classes. \blacktriangleright Ideally, all positive instances (y=1) should be on the side of the separator satisfying $\overline{W} \cdot \overline{X} > 0$.

as input nows opens transmitor a compete a post tupter a separa as weights a ogrego, ophiconso a righ other y= rign (W.X) = rign(\$ "; ")

I perception voi touto separes os intéries positives des reget res por meiro de uma lonha de reportado

lacktriangle All negative instances (y=-1) should be on the side of the separator satisfying

Poma por todor or poror (X14) e rediza

roma l'ovoliande De & predito é aquel

dotoret de trains

▶ In many settings (e.g., skewed class distribution) we need an invariant part of the $\hat{y} = \operatorname{sign}\{\overline{W} \cdot \overline{X} + b\} = \operatorname{sign}\{\sum_{j=1}^d w_j x_j + b\} = \operatorname{sign}\{\sum_{i=1}^{d+1} w_j x_j\}$

lackbox On setting $w_{d+1}=b$ and x_{d+1} as the input from the bias neuron, it makes little difference to learning procedures \Rightarrow Often implicit in architectural diagrams

Bias Neurons

Training a Perceptron lacktriangle Go through the input-output pairs (\overline{X},y) one by one and make updates, if predicted value \hat{y} is different from observed value $y\Rightarrow$ Biological readjustment of

synaptic weights. $\overline{W} \Leftarrow \overline{W} + \alpha \underbrace{(y - \hat{y})}_{\overline{Z}} \overline{X}$ $\overline{W} \Leftarrow \overline{W} + (2\alpha)y\overline{X}$ [For misclassified instances $y - \hat{y} = 2y$]

 \blacktriangleright Parameter α is the learning rate \Rightarrow Turns out to be irrelevant in the special case of the perceptron lacktriangle One cycle through the entire training data set is referred to as an $epoch \Rightarrow$ Multiple epochs required ► How did we derive these updates?

What Objective Function is the Perceptron Optimizing?

▶ At the time, the perceptron was proposed, the notion of loss function was not popular \Rightarrow Updates were heuristic ▶ Perceptron optimizes the perceptron criterion for *i*th training instance:

 $L_i = \max\{-y_i(\overline{W}\cdot\overline{X_i}), 0\}$ - Loss function tells us how far we are from a desired solution \Rightarrow Perceptron criterion is 0 when $\overline{W}\cdot\overline{X_i}$ has same sign as y_i .

Perceptron updates use $stochastic\ gradient\ descent$ to optimize the loss function and reach the desired outcome. - Updates are equivalent to $\overline{W} \Leftarrow \overline{W} - \alpha \left(\frac{\partial L_i}{\partial w_1} \dots \frac{\partial L_i}{\partial w_d} \right)$

Where does the Perceptron Fail?

► The perceptron fails at similar problems as a linear SVM Classical solution: Feature engineering with Radial Basis Function network \Rightarrow Similar to kernel SVM and good for noisy data

Historical Origins

▶ The first model of a computational unit was the perceptron (1958). Was roughly inspired by the biological model of a neuron. Was implemented using a large piece of hardware. Generated great excitement but failed to live up to inflated expectations ▶ Was not any more powerful than a simple linear model that can be implemented

Perceptron Tutorial

Perceptron tutorial (simple_perceptron.py).

Tensorflow Playground³



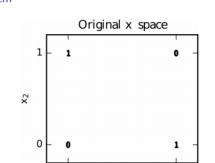
³https://playground.tensorflow.org/

The XOR Problem

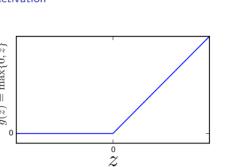
"Perceptrons" by Marvin Minsky and Seymour Papert (1969). Perceptrons cannot solve the XOR problem.

Significant decline in interest and funding of neural network research.

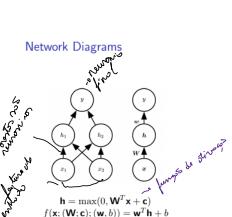
The XOR Problem



Rectified Linear Activation



Função doscontinua



 $\begin{pmatrix} h_1 \\ h_2 \\ l \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \zeta \\ w_{21} & w_{22} & \zeta_2 \\ 0 & 0 & l \end{pmatrix} \begin{pmatrix} \chi f \\ \chi g \\ \chi g \\ \chi g \end{pmatrix}$

Educado James eses ames esua) X

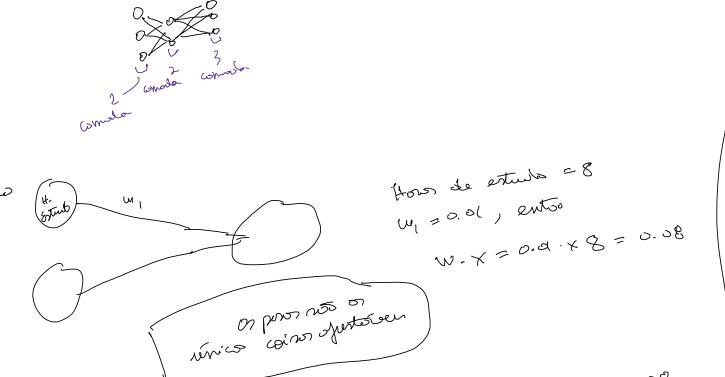
Campbo scultar offer não tomos domos

Ser mais ou monor saperire hos mudories?

Uma torar de aprondezada mais interesce

pecino mos lovo mois interesce

-ours comme pode to voio nemonio Or remarion de una nosmo comoda não ne consetom

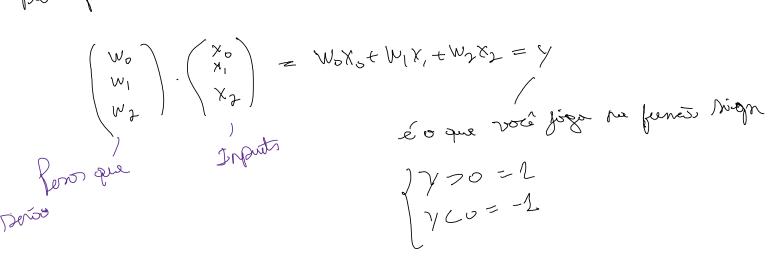


Vanor dizer que a primero mer de seperado am 1.6, entro tomos que experido en pers1. m=0.2 = 1.6 2 Ene nove volor é colcularson por una bienção de pende (los function)

Ele voi gomento o conhecimento poro se eslegues una vous entreles en vez de fozer a volapticos p opens sepuels volors 1

6 modelo voi se os peros exallistos, os promotros exallistas

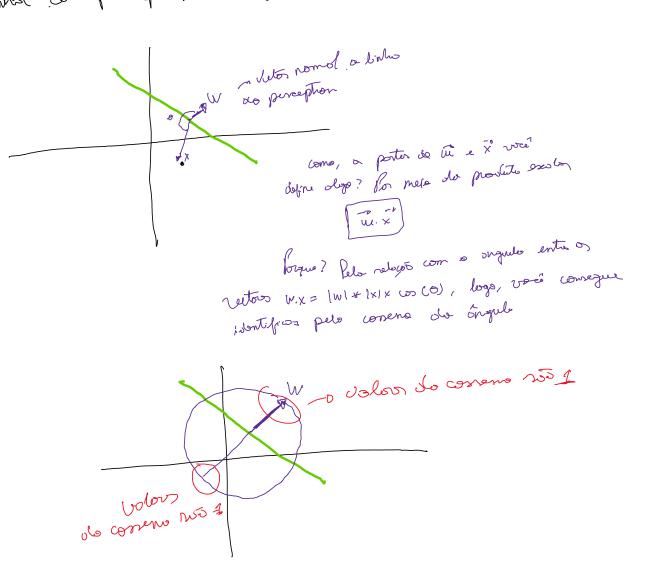
Dot product = hoduto Evenlon



WTX = W troupoter - X, outros forma de fola do produto escolar

I possa = hadita exala ala pera e des imputs 2 porm = Jogo no bunço sisgn 3 poru = Resleula os peros de ocordo com um objoritmo

- A lubar de perception i quisber por um veter Ti normal



Posse de dualizado pero voca pero voca por morto $w_1' = w_1 + n \cdot (y - \hat{y}) \cdot y_1'$ Novo selho tora oprendezado $w_1' = 0 + 0.2 \cdot (-1) \cdot L = -0.2$

(=) O De Zwyyj-h & ou sejo, o somotóno dos produtos escolosos w, vj + w, x, + w, x, = y

~ Aprender pre perception e colouber o lintur, definince or peror

-P Gradient Nemant vo Outro forma de treiros o perception

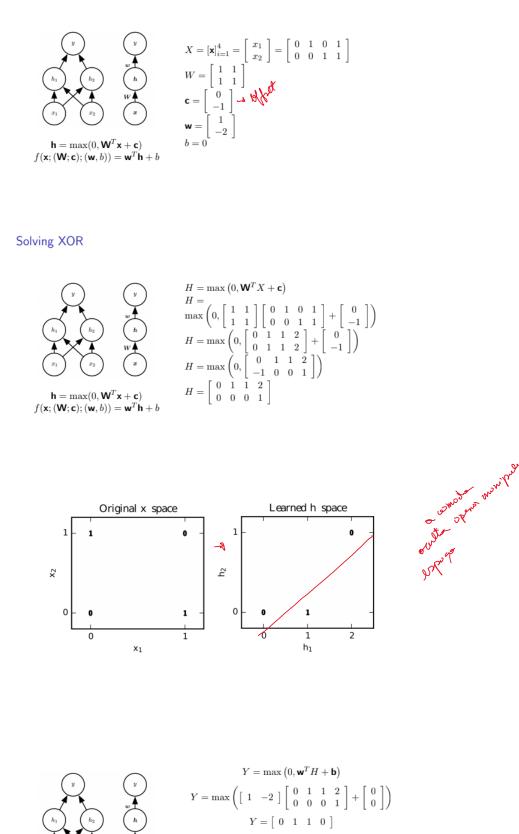
· Rosalver o problèmer de voi d'inevidade, como a XOIR

- Delta rule, de modo que re huns uma oproscionoção

Lo Utiliza o quelient descent plodos os peros

De vinicio, consideramos OCX) = W.X sem um throphold = 2 Boya treivos, premomos nos losses em uma pretrar, posse won es emo $E(\tilde{w}) = \begin{cases} \sum_{i=1}^{n} (t_{i} - o_{i}) \\ \text{potent} \end{cases}$

A ideia e surro vilimijon une função, sejo com o mor ouros min



Thank you! tvieira@ic.ufal.br

 $\begin{aligned} \mathbf{h} &= \max(0, \mathbf{W}^T \mathbf{x} + \mathbf{c}) \\ f(\mathbf{x}; (\mathbf{W}; \mathbf{c}); (\mathbf{w}, b)) &= \mathbf{w}^T \mathbf{h} + b \end{aligned}$