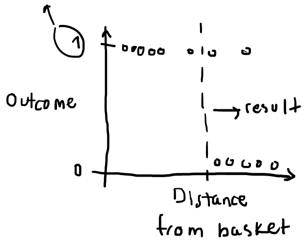
Week4

Saturday, February 26, 2022 12:32 PM

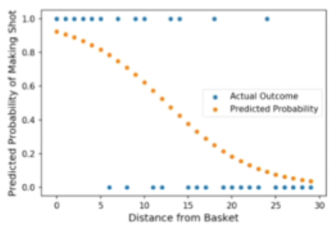
Logistic Regression

→ This a classification technique

→ Distribution convergence : when they both reach the same distribution

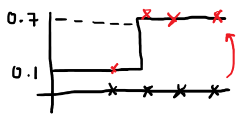


the shorter the distance, the more likely a win is → at what distance should we shoot from?



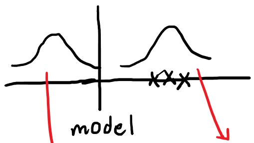
Verisimilitude(verosimilitud)

↳ how similar my data is, called π



evaluated → verisimilitude is the fact that we have an x and we evaluate on the same points (the PDF function)

$\pi = (0.1)(0.7)(0.7) = \boxed{0.049}$ ← we want to maximize it



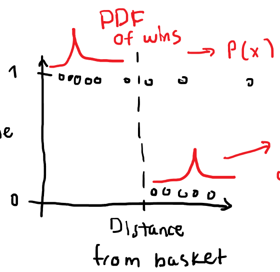
model

than this one

the model is better in this area

the more x points land on the distribution's density domain, the larger π we get, the better model → (probability density function

look for the closest points to the density function's peak.



basically binary outcome depending on one axis

logistic Regr function

$$p(x) = \frac{1}{1+e^{-nx_i+b}}$$

How do we merge them?

=

Bernoulli function technique

Bernoulli function : $f(x) = p(x_i)^{y_i}(1-p(x_i))^{1-y_i}$

(Bernoulli Trials)

filters what we consider when merging the PDF's for both win/loss

x is verisimilitude

we want to find the m, b parameters that optimize the function

Thus

$$\max_{m,b} f(x) = \max_{m,b} \overbrace{\prod_{i=\emptyset}^{N}}^{\text{Prod}} f(x_i)$$

$$= \max_{m,b} \prod_{i=1}^{N} p(x_i)^{y_i}(1-p(x_i))^{1-y_i}$$

After $\log()$,

$$= \max_{m,b} \log\Big(\overbrace{\prod p(x_i)^{y_i}(1-p(x_i))^{1-y_i}}^{\log(ab)=\log(a)+\log(b)}\Big)$$

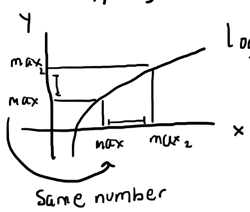
→ now Σ

$$= \max_{m,b} \sum_{i=0}^{N-1} \log\left(p(x_i)^{y_i}(1-P(x_i))^{1-y_i}\right)$$

Differentiate to maximize (and also equalize to $\emptyset$)

$$\frac{\partial}{\partial m} \sum_{i=0}^{N-1} \log\left(p(x_i)^{y_i}(1-p(x_i))^{1-y_i}\right) = 0$$

Why $\log(L(x_i, m, b))$?

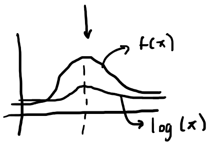


$\log(x)$

↳ monotonously incrementing

=

do not affect the maximum



log(x) "softens the curve" and reduces dimensions → since the increment softens → useful for maximizing

$$\frac{\partial}{\partial m}\sum_{i=0} \log\left(p(x_i)^{y_i}(1-p(x_i))^{1-y_i}\right) = 0$$

useful for maximizing

x_i = distance to the basket
y_i = 0 or 1 (fail or success)

$$\frac{\partial}{\partial m}\sum_{i=0}^{N-1} \log\left(p(x_i)^{y_i}\right) + \log\left(1-p(x_i)\right)^{1-y_i} = 0$$

$$\frac{\partial}{\partial m}\sum_{i=0}^{N-1} y_i\log(p(x_i)) + (1-y_i)\log(1-p(x_i)) = 0$$

$$\frac{\partial}{\partial m}\sum_{i=0}^{N-1}\left(y_i\log(p(x_i)) + \log(1-p(x_i)) - y_i\log(1-p(x_i))\right) = 0$$

Considering $p(x_i) = \underbrace{\frac{1}{1+e^{-mx_i+b}}}_{\text{logistic distribution}}$, with bias $b=0$

$$\frac{\partial}{\partial m}\sum\left[y_i\log\left(\frac{p(x_i)}{1-p(x_i)}\right) + \log(1-p(x_i))\right] = 0$$

Evaluating the term $\frac{P(x_i)}{1-P(x_i)}$, considering $\alpha = mx_i + b$

$$\frac{P(x_i)}{1-P(x_i)} = \frac{\frac{1}{1+e^{-\alpha}}}{1-\frac{1}{1+e^{-\alpha}}} = \frac{\frac{1}{1+e^{-\alpha}}}{\frac{\cancel{1}+e^{-\alpha}-\cancel{1}}{1+e^{-\alpha}}} = \frac{1+e^{-\alpha}}{e^{-\alpha}(1+e^{-\alpha})} = \frac{1}{e^{-\alpha}} = e^{\alpha}$$

$$= e^{mx_i+b}$$

Evaluating the term $1-P(x_i)$

$$1-P(x_i) = 1 - \frac{1}{1+e^{-\alpha}} = \frac{\cancel{1}+e^{-\alpha}-\cancel{1}}{1+e^{-\alpha}} = \frac{e^{-\alpha}}{1+e^{-\alpha}}$$

Thus,

$$= \frac{\partial}{\partial m}\sum_{i=0}^{N-1} y_i\log\left(e^{mx_i+b}\right) + \sum_{i=0}^{N-1}\log\left(\frac{e^{-(mx_i+b)}}{1+e^{-(mx_i+b)}}\right) = 0$$

Now with $b=0$ (bias)

$$\frac{\partial}{\partial m}\left(-\sum_{i=0}^{N-1} mx_iy_i - \sum_{i=0}^{N-1} mx_i + \sum_{i=0}^{N-1}\log\left(1+e^{-mx_i}\right)\right) = \underbrace{\frac{\partial}{\partial m}\ell(x,m)}_{\max_m \ell(x,m)} = 0$$

Differentiating

$$= 2\sum_{i=0}^{N-1}\left(x_iy_i + \frac{x_i(1+e^{-mx})}{1+e^{-mx}}\right)$$

$$= \boxed{\sum_{i=0}^{N-1} x_i(y_i - P(x_i))}$$

this sum will output the m that maximizes the verisimilitude function

where the objective was to find parameters m and b that maximize the verisimilitude function

$$\max_{m,b} L(x_i,m,b) = \max_{m,b}\prod f(x_i)$$

Types of Convergence

- Convergence in distribution, converges weakly or in Law (fractals)

Convergence in mean (unbiased estimator)

- Convergence 'Almost - Sure':
  - Almost surely, almost everywhere, strongly or with probability 1.

We thus said we try to find m, b parameters that maximize our function, that is, that separate the training data in almost-sure convergence. The problem is addressed with the Likelihood Method, which is defined for Bernoulli Trials as mentioned.

$$\underbrace{L(m,b)}_{} = \prod_{i=0}^{N-1} p(x_i)^{y_i}(1-p(x_i))^{1-y_i}$$

Thus $\max_{m,b} L(m,b)$

Our results for maximizing the Likelihood Function (verisimilitude) were

Our results for maximizing the Likelihood Function (verisimilitude) were

$$\frac{\partial \ell(m,b)}{\partial m} = \sum_{i=0}^{N-1} \left((y_i - p(x_i)) x_i \right) = 0$$

But unfortunately the last equation does not present a closed form, and thus we propose to use a Newton-Rapson method for numerical approximation

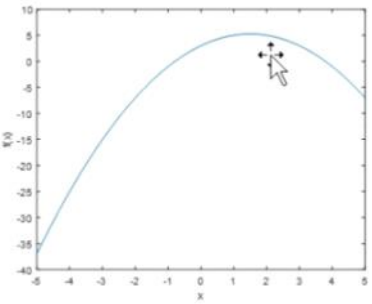
Basically we can't solve this for m since m is very inside of an exponential function, so we will use a num method for 'solving' for m.

$$m_{iteration+1} = m_{iteration} - \frac{\dfrac{\partial \ell(m_{iteration})}{\partial m}}{\dfrac{\partial^2 \ell(m_{iteration})}{\partial^2 m}}$$

similar

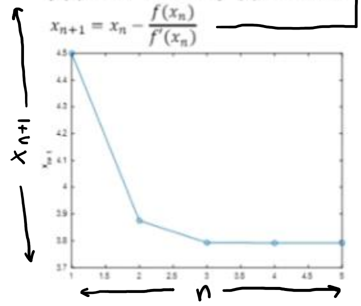
Newton Rapson Example

- Given the f(x) function, find its maximum



$$f(x) = -x^2 + 3x + 3, f'(x) = -2x + 3$$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



So, our expression of Newton Rapson to the Likelihood Function

$$m_{iteration+1} = m_{iteration} - \frac{\sum_{i=0}^{N-1} ((y_i - p(x_i)x_i)}{\sum_{i=0}^{N-1} (-x_i(1-p(x_i)) - p(x_i))}$$

Computing $\dfrac{\dfrac{\partial \ell}{\partial m}}{\dfrac{\partial^2 \ell}{\partial^2 m}}$

Example,

Given a set of training, find the parameter m and b such that separates into classes

$X = \{-0.1, -0.5, -2, 2, 2.3, 4, 0, 5\}$

$y = \{0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\}$

- The stop criteria between iterations is > 0.05

→ Code Available, resulting in:

b=0 (bias)

m = 1.2140

error = -0.043547

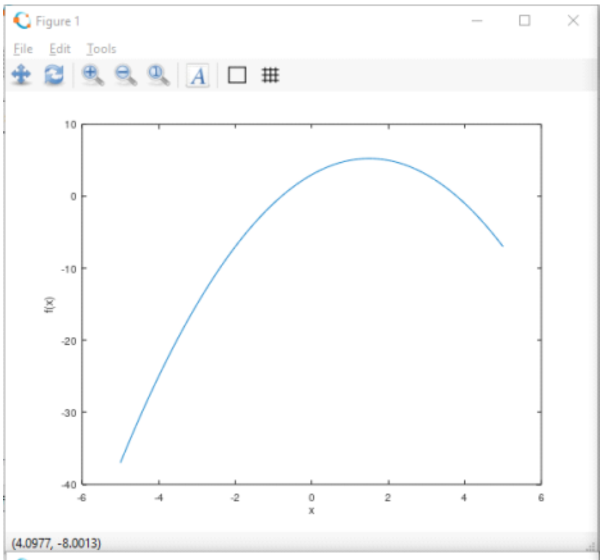


Figure 1
File Edit Tools
f(x)
x
(4.0977, -8.0013)

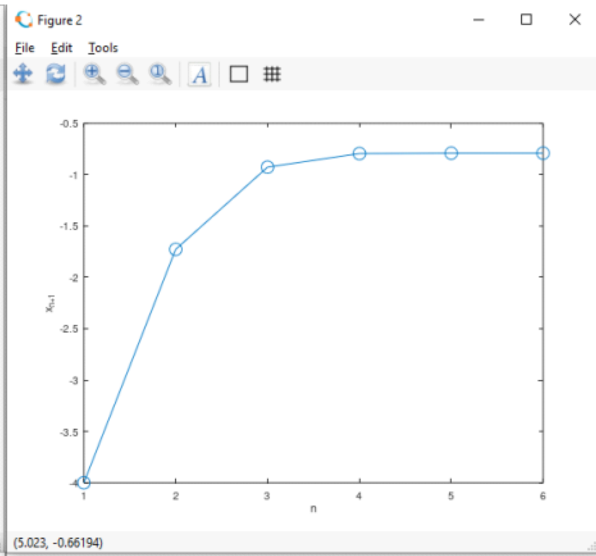


Figure 2
File Edit Tools
n
(5.023, -0.66194)

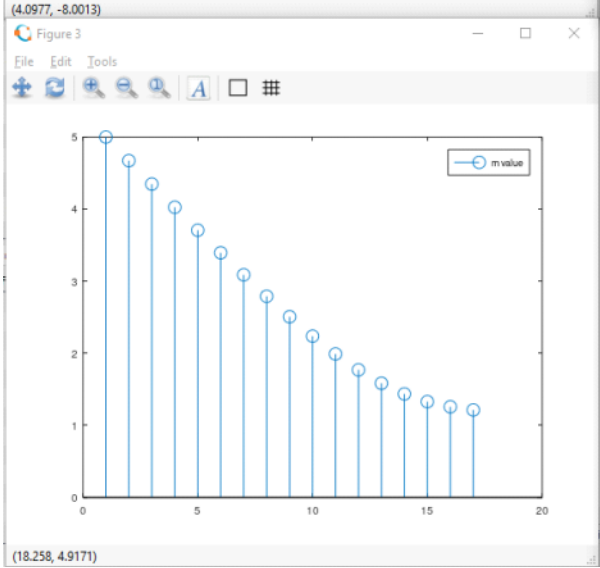


Figure 3
File Edit Tools
m value
(18.258, 4.9171)

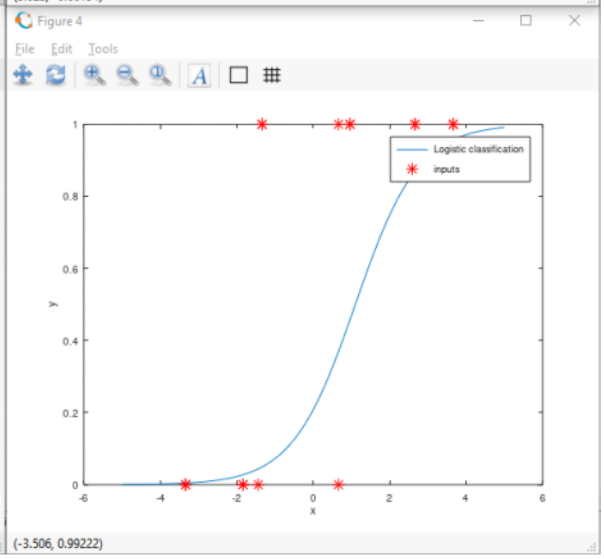


Figure 4
File Edit Tools
Logistic classification
inputs
y
x
(-3.506, 0.99222)