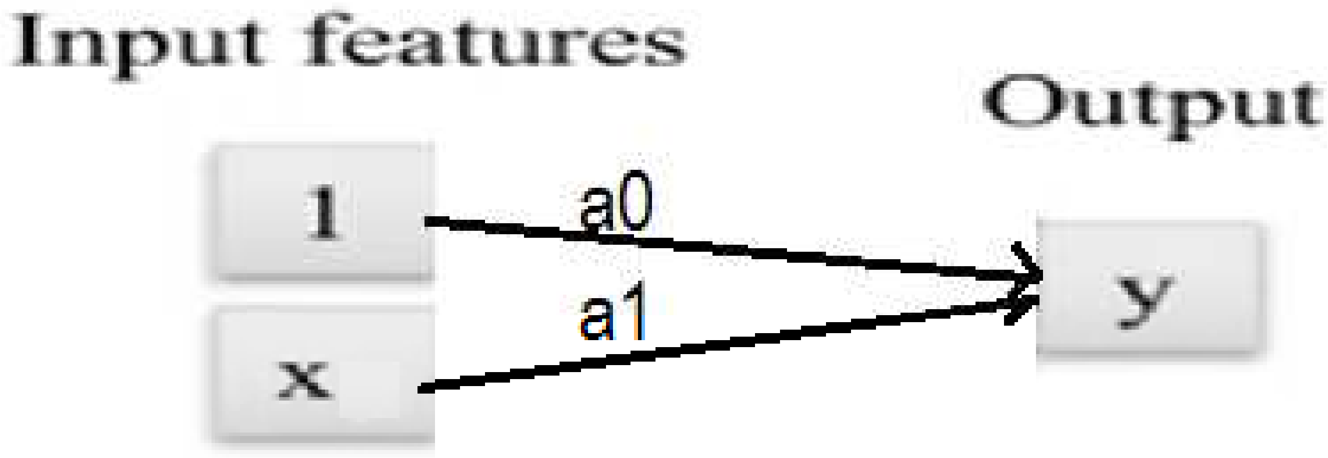
Chapitre II : Analyse Prédictive :

(Prédire une Catégorie : Classification)

II- 3- Classification Linéaire : Régression Logistique

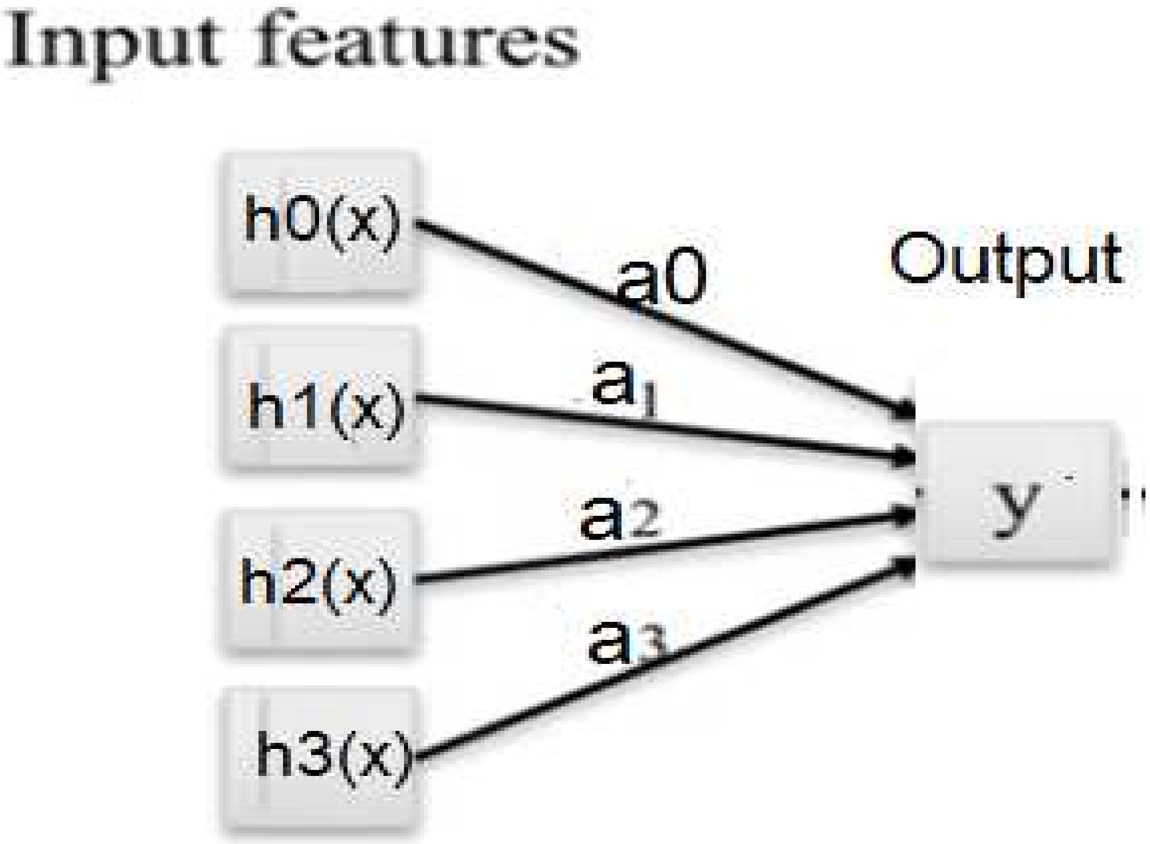
Régression Linéaire Simple:



y=a0 + a1x

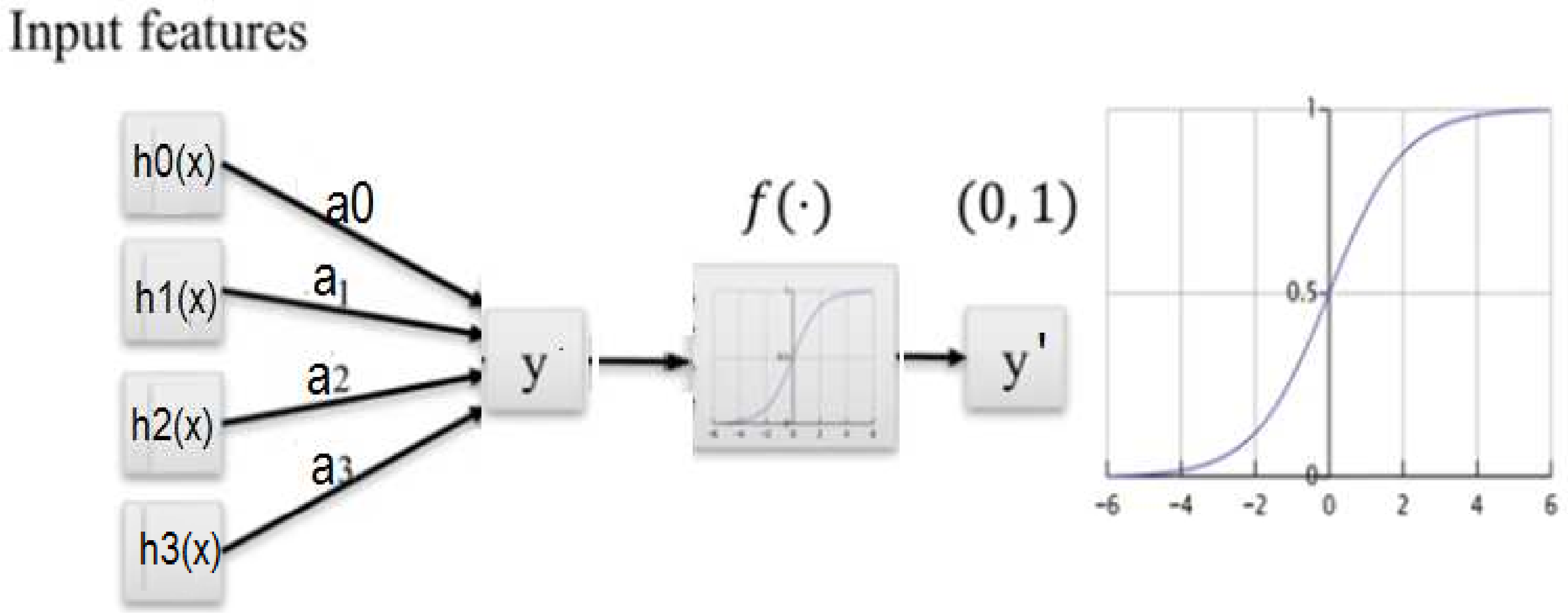
Equation d’une Droite

Régression Linéaire Multiple/Polynomiale:



Y= a0h0(x) + a1h1(x) + a2h2(x)+…+ aphp(x)

Régression Logistique



•Cette approche de prévision utilise la même forme linéaire que celle utilisée pour la régression.

•Au lieu de prédire une valeur cible continue, nous prenons la sortie de la fonction linéaire et appliquons la fonction **Sigmoid** pour produire une sortie binaire avec deux valeurs possibles (2 étiquettes de classe possibles).

•Si la valeur cible >0,5, ➔la fonction renvoie +1

•si la valeur cible <0,5 ➔la fonction renvoie -1

La Régression Logistique:

La régression logistique est un modèle de classification linéaire simple ou nous avons deux variables Y et X :

Y est la variable expliquée qualitative et

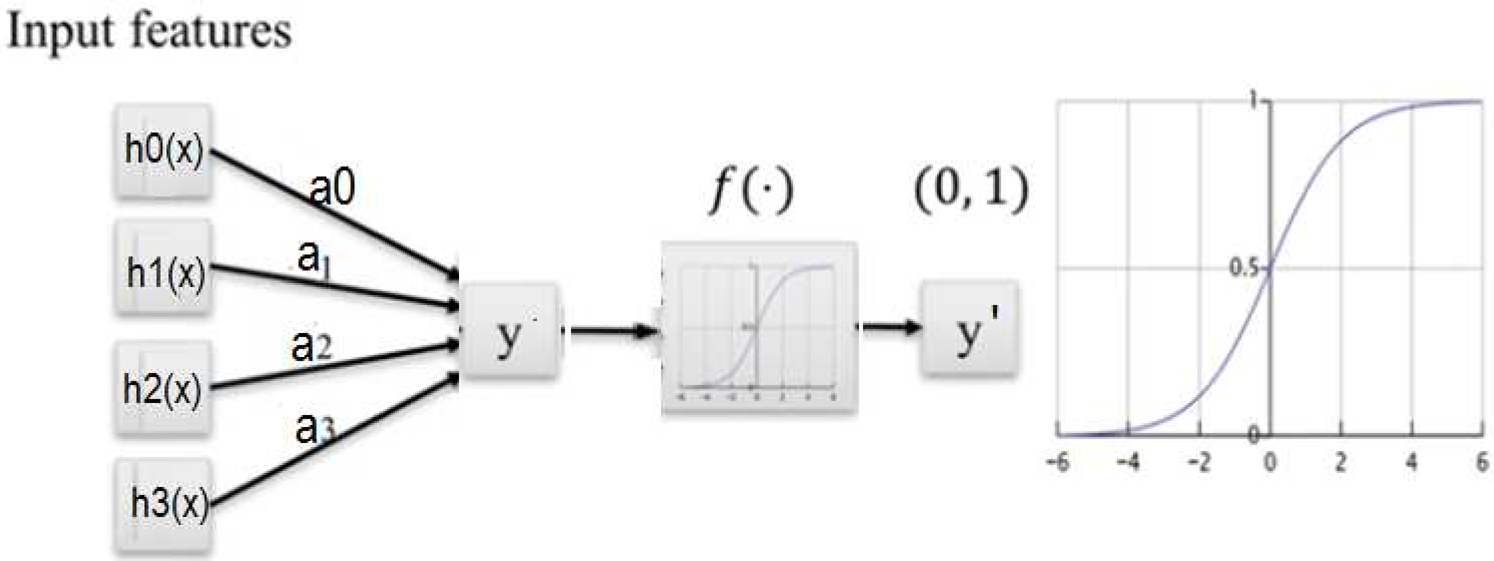
X est la variable explicative , seulement ici X est donnée avec plusieurs features , elle peut être quantitative ou qualitative(On doit codifier)

Codage des catégories ou des classes :

Si les entrées sont quantitatives (numériques) Longueur, Largeur…pas de problème de multiplication,

mais si les entrées sont qualitatives, alors on procède a une codification des entrées.

Régression Logistique



On reprend la formule générale de la régression depuis le chapitre II Regression polynomiale :

$Y = a_0h_0(x) + a_1h_1(x) + a_2h_2(x) + a_3h_3(x) + \ldots\ldots + a_p(h_p(x) = a^TH(x)$

Sachant que :

h0(x)=feature0 ;

h1(x)=feature1;

……………….

hn(x)=feature n

Le modèle de la régression Logistique :

Donc : $Y' = f(Y) = f(a^T H(x))$

$Y' = f(a_0 h_0(x) + a_1 h_1(x) + a_2 h_2(x) + a_3 h_3(x) + \ldots\ldots + a_p(h_p(x))$

La fonction 'f' s'appelle : Fonction Logistique:

$Y' = \text{logistic}[a_0 h_0(x) + a_1 h_1(x) + a_2 h_2(x) + a_3 h_3(x) + \ldots\ldots + a_p(h_p(x)]$

La fonction logistique transforme une entrée réelle en une probabilité que la classe soit positive.

$$Y' = P(Y = +1/X) = f(a^T H(x))$$

Le modèle de la régression Logistique :

$$Y' = P(Y=+1/X) = f(a^T H(xi))$$

➔ SI $P(Y=+1/X) >= 0.5$ (ou $a^T H(xi) > 0$) Alors

$y'(xi) = +1$

Sinon

$y'(xi) = -1$ ou 0

La fonction Logistique/Sigmoid

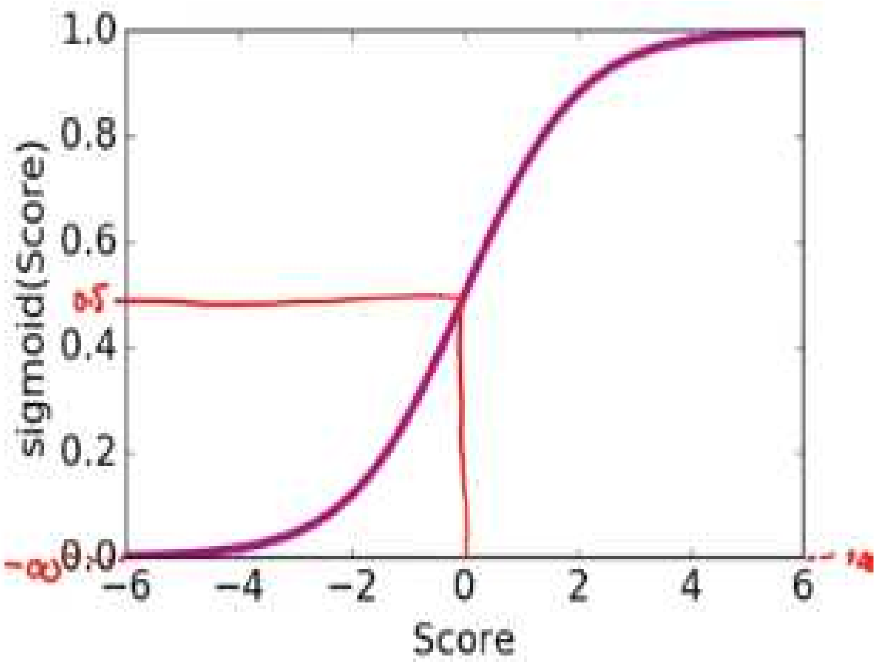
Dans notre cas :
g est la fonction f

ar :

Dans notre cas :
Score = $a^T H(x)$

$$g(score) = sigmoid(score) = \frac{1}{1 + e^{(-Score)}}$$

Elle prend ses valeurs entre 0 et 1 :



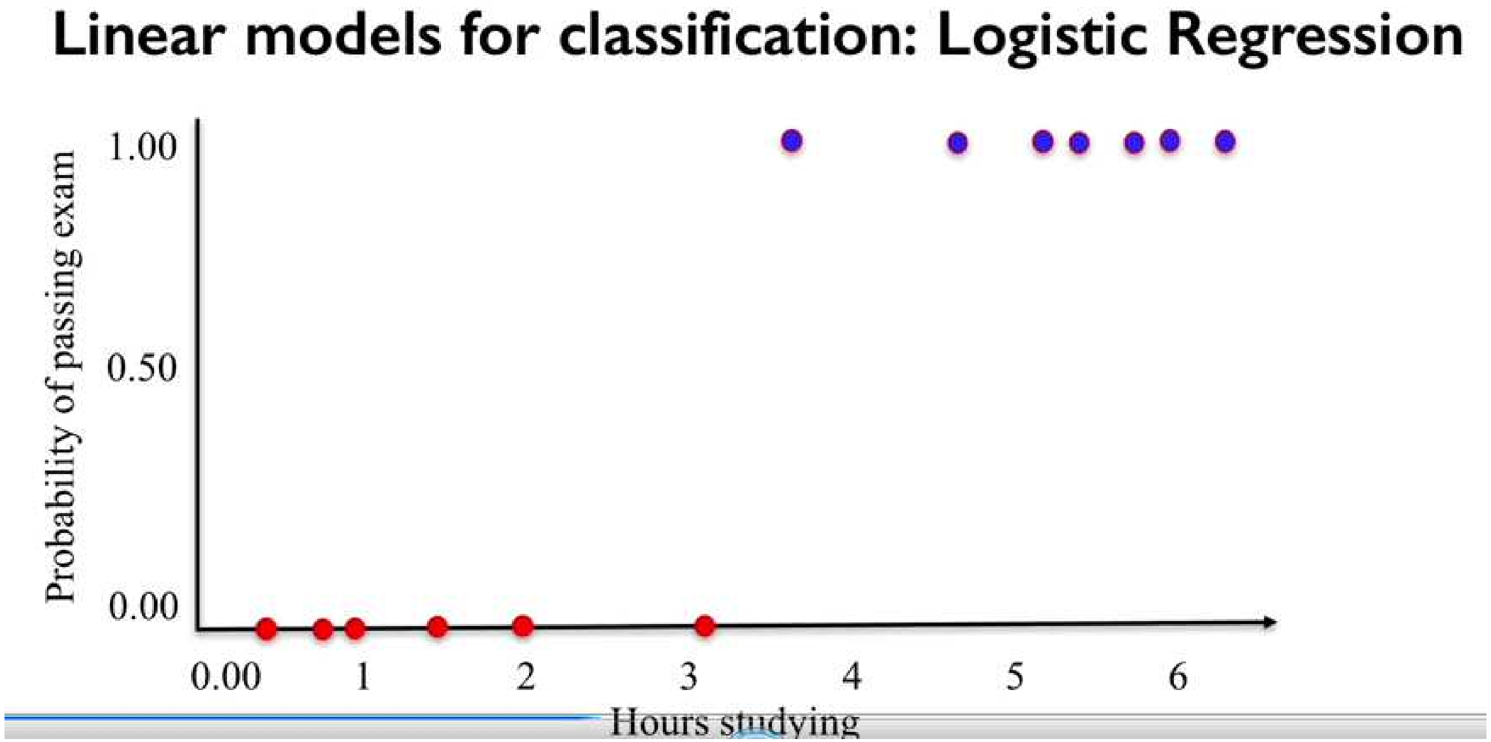
Score	-∞	0.	+∞
Sigmoid(score)	0	0.5	1

Dans **la régression logistique**,

-On ne prédit pas +1 ou -1.
-Nous prédisons une probabilité.

Quelle est la probabilité que cette évaluation soit positive?
Quelle est la probabilité d'être négatif?

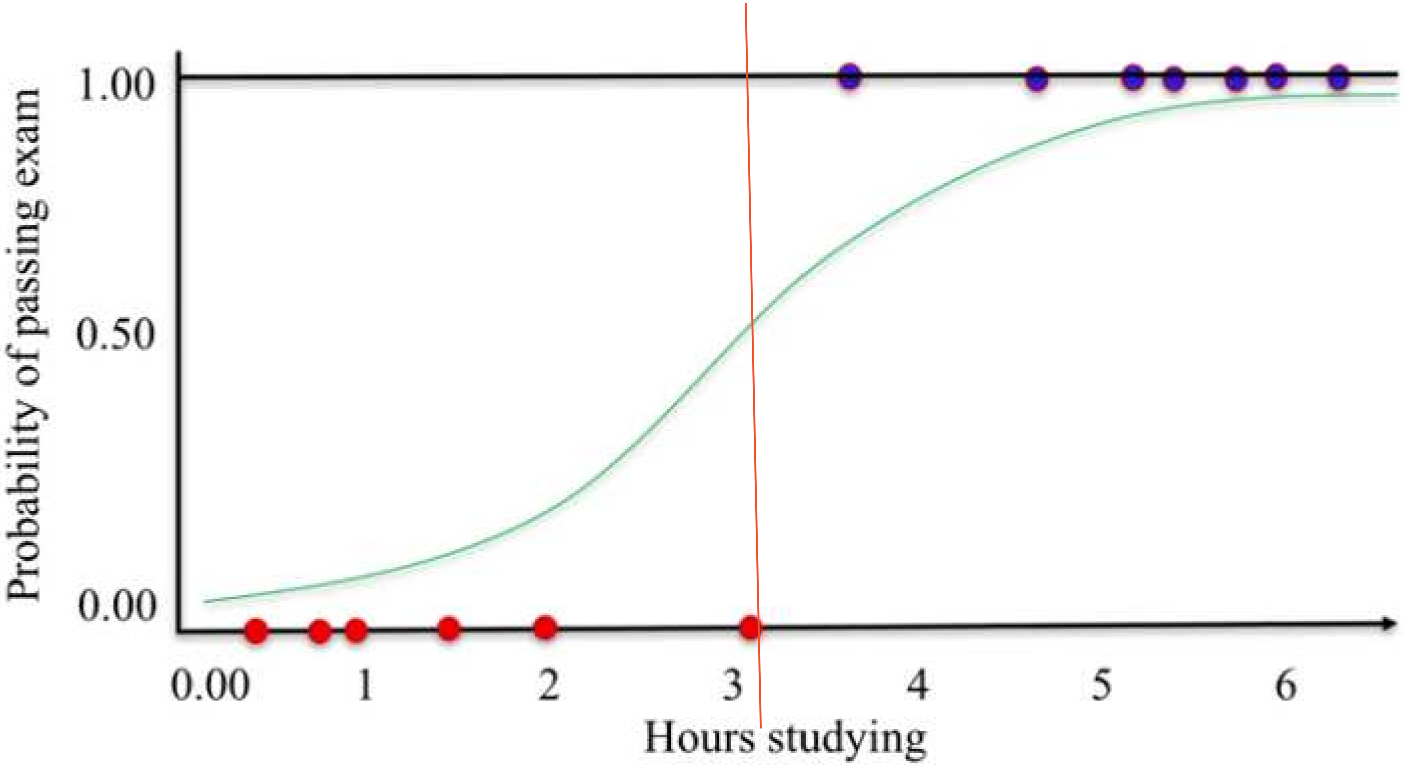
Régression Logistique:
Classification avec Un seul Feature



•On dispose d'une seule variable Heures d'étude, et on voudrait classer les étudiants selon cette variable pour savoir qui va réussir /Non

•On donne les probabilités pour chacun selon nbre Heures.

Linear models for classification: Logistic Regression



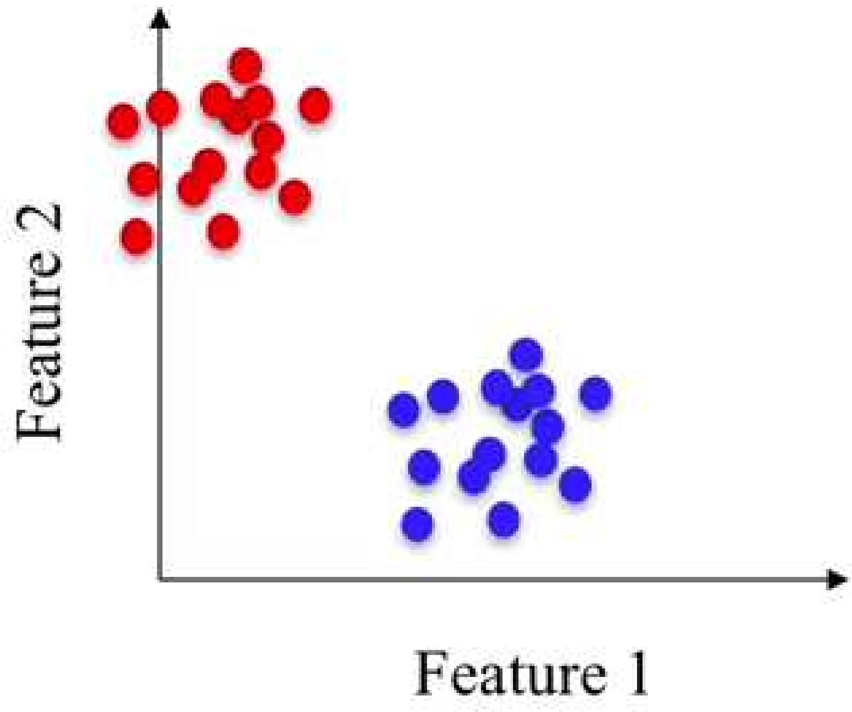
En utilisant la Régression Logistique on estime les coefficients du modèle qui s'ajuste au mieux à l'ensemble des points

Probabilité >0.5➔ Y(classe=+1)

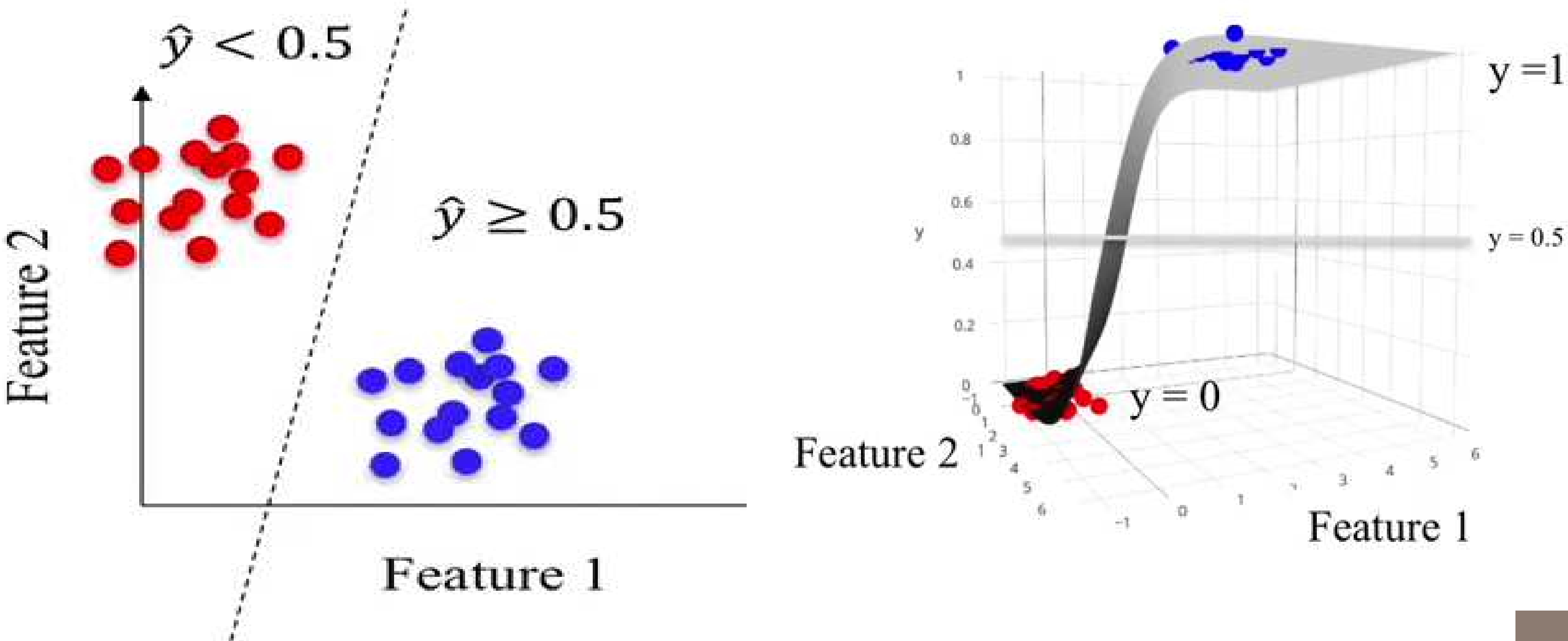
Probabilité<0.5 ➔Y(classe=0 ou -1)

➔ Les étudiants qui étudient plus de 3h passent l'examen, sinon Non

Régression Logistique :
Classification avec deux Features:



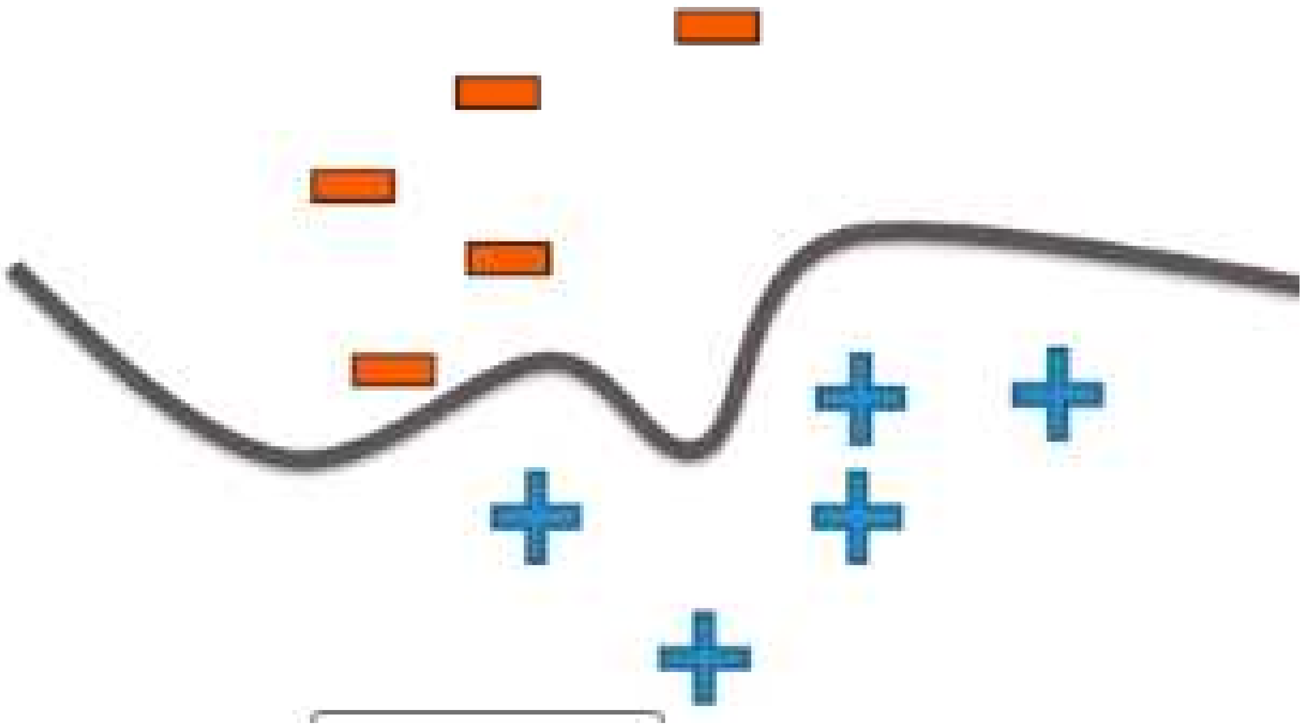
Régression Logistique :
Classification avec deux Features:



Autrement dit: La Régression Logistique donne une limite de décision linéaire entre les deux classes ici un plan (Deux Features).

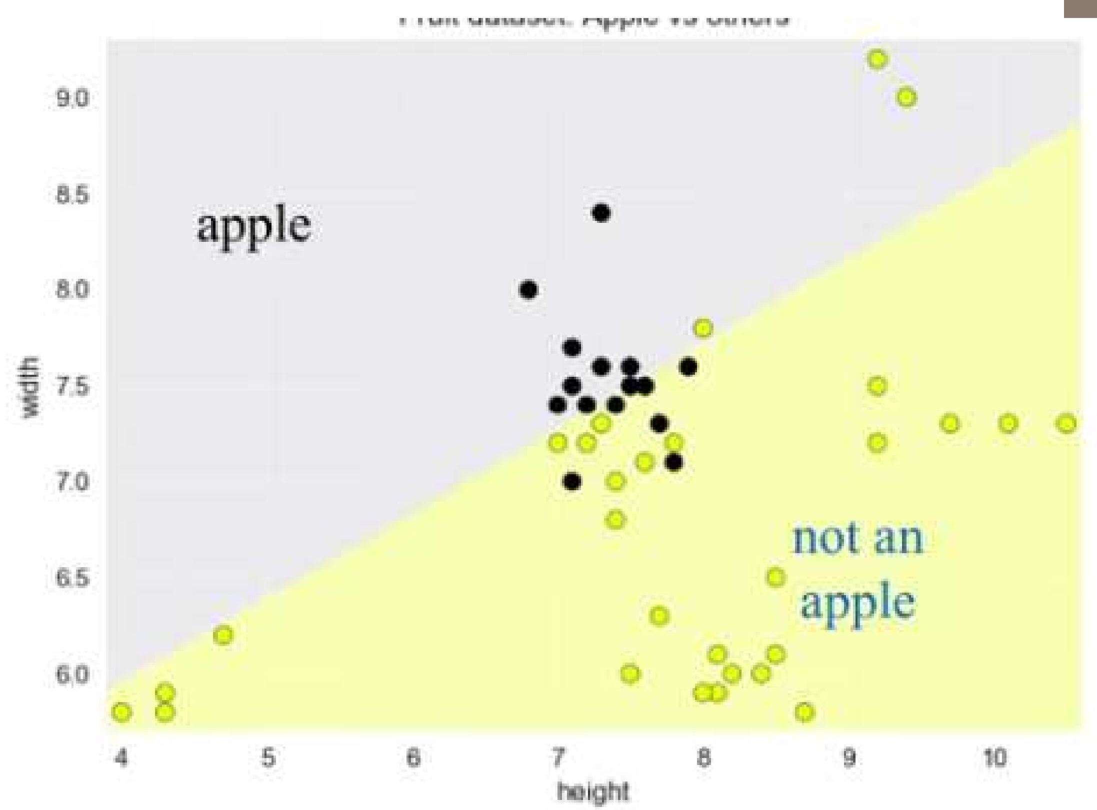
Les classifieurs <u>non linéaire :</u> la limite de décision possède des formes plus complexes :

L



Régression Logistique Simple : Classification de Fruits Pomme Vs Autre:

Fruits	Largeur	Hauteur
Pomme	7	5
Orange	9	6
Lemon	11	4
Mandarine	4	3
Pomme	8	6
………..		

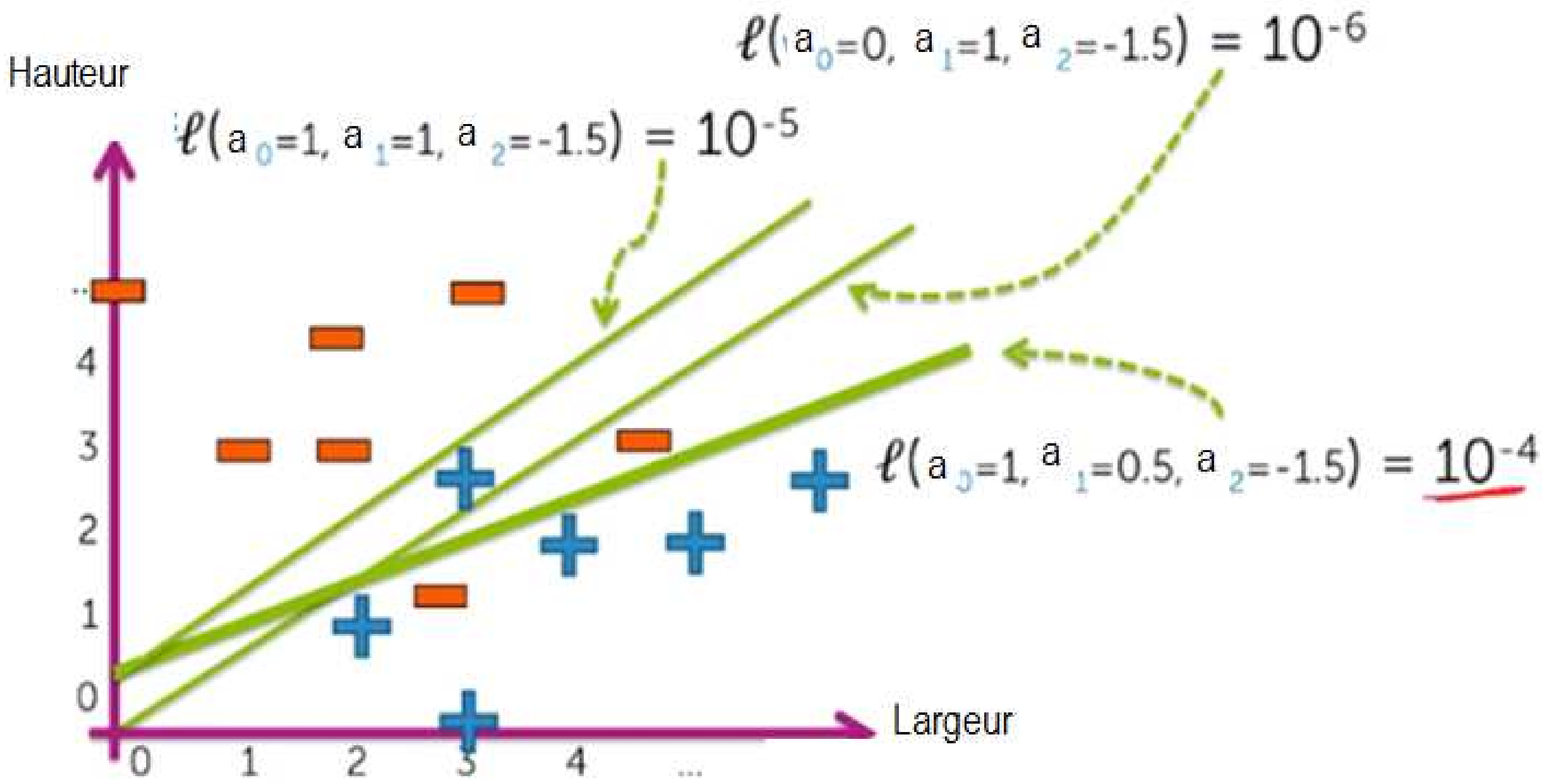


Un classifieur linéaire doit utiliser les données d'apprentissage pour apprendre ou calculer les coefficients ou les poids de chaque feature (Largeur, Hauteur).

Trouver le meilleur classifieur :

Pour trouver le meilleur classifieur, on doit **maximiser** la qualité de la métrique appelée « Likelyhood =Vraissemblance »

l(a).par rapport a tous les a0,a1,a2...



Trouver le Meilleur classifieur :

La qualité de la métrique

« Likelyhood = Vraissemblance»

$$l(a) = \prod_{i=1}^{N} P(y_i \mid \mathbf{x}_i, a)$$

Nous avons:

$Y' = P(Y=+1/X) = f(a^T H(xi))$ ➔

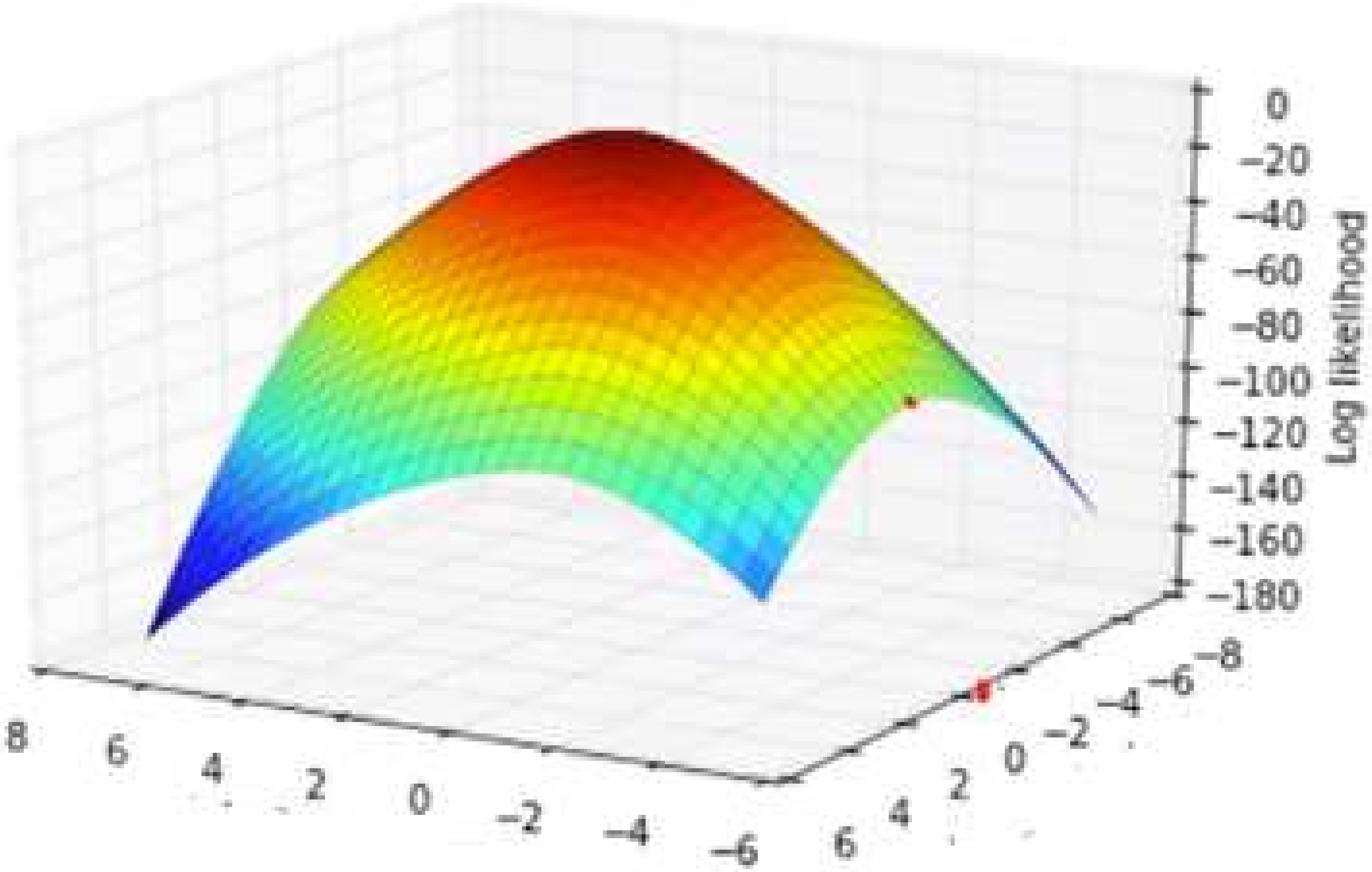
$$l(a) = \prod_{i=1}^{N} g(a^t H(x_i))$$

Meilleur classifieur = Maximiser (l(a))

L(a) =fonction +eurs variables a0,a1,..aD

➔Utiliser un <u>Algorithme d'Optimisation</u>

➔ <u>tel que Ascent Gradient</u>



<u>l’Algorithme Ascent Gradient</u> :

Début

t=1

Initialiser les coefficients (a);

*Initialiser ε:tolérance ;

Initialiser : η le ‘pas’ ;

While (not converged)

a(t+1) 🡄 at + η grad(l(a))

t🡄t+1

Fin

Apprentissage de l'algo AscG pour la régression logistique:

Réalité

Prédiction

$$\frac{\partial l(a)}{\partial aj} = \sum_{i=1}^{n} hj(xi)(v[yi = +1] - p(yi = +1/xi, a))$$

Différence entre la réalité et la prédiction

$$v[yi = +1]$$

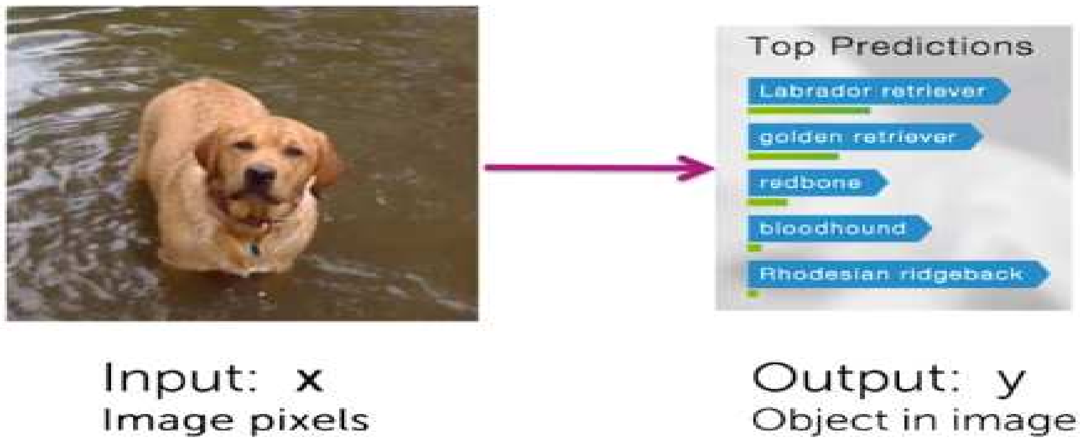
est appelé indicateur de la fonction = 1 si yi=+1

=0 sinon

Multi classes Classification :

(L'approche :une classe versus toutes les classes):

On voudrait classer une image ou un objet dans l'image , par exemple :

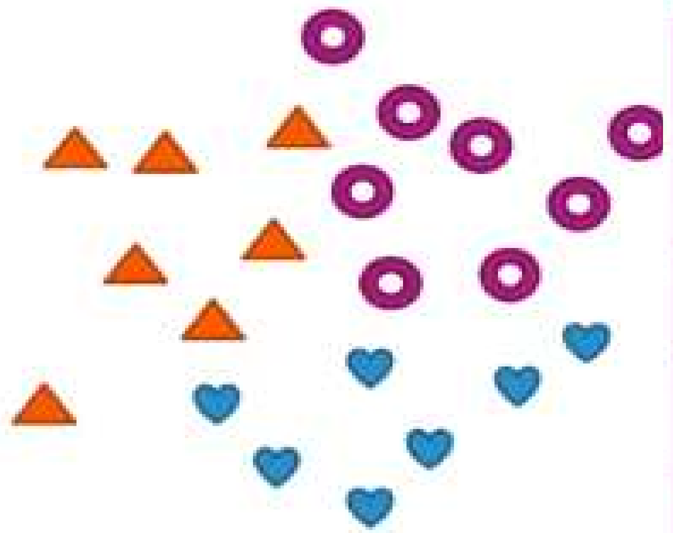


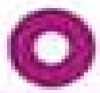
Est-ce qu'il s'agit d'une oie d'un chien, d'une camera…..

Formulation :

Nous avons C classes de formes :➔

y peut être 1,2,….C



Data point	x[1]	x[2]	y
$\mathbf{x}_1, y_1$	2	1	
$\mathbf{x}_2, y_2$	0	2	
$\mathbf{x}_3, y_3$	3	3	
$\mathbf{x}_4, y_4$	4	1	

Nous avons N points de données. Et on voudrait apprendre :

$$\hat{P}(y=\triangle\ |x)$$

$$\hat{P}(y=\heartsuit\ |x)$$

$$\hat{P}(y=\circ\ |x)$$

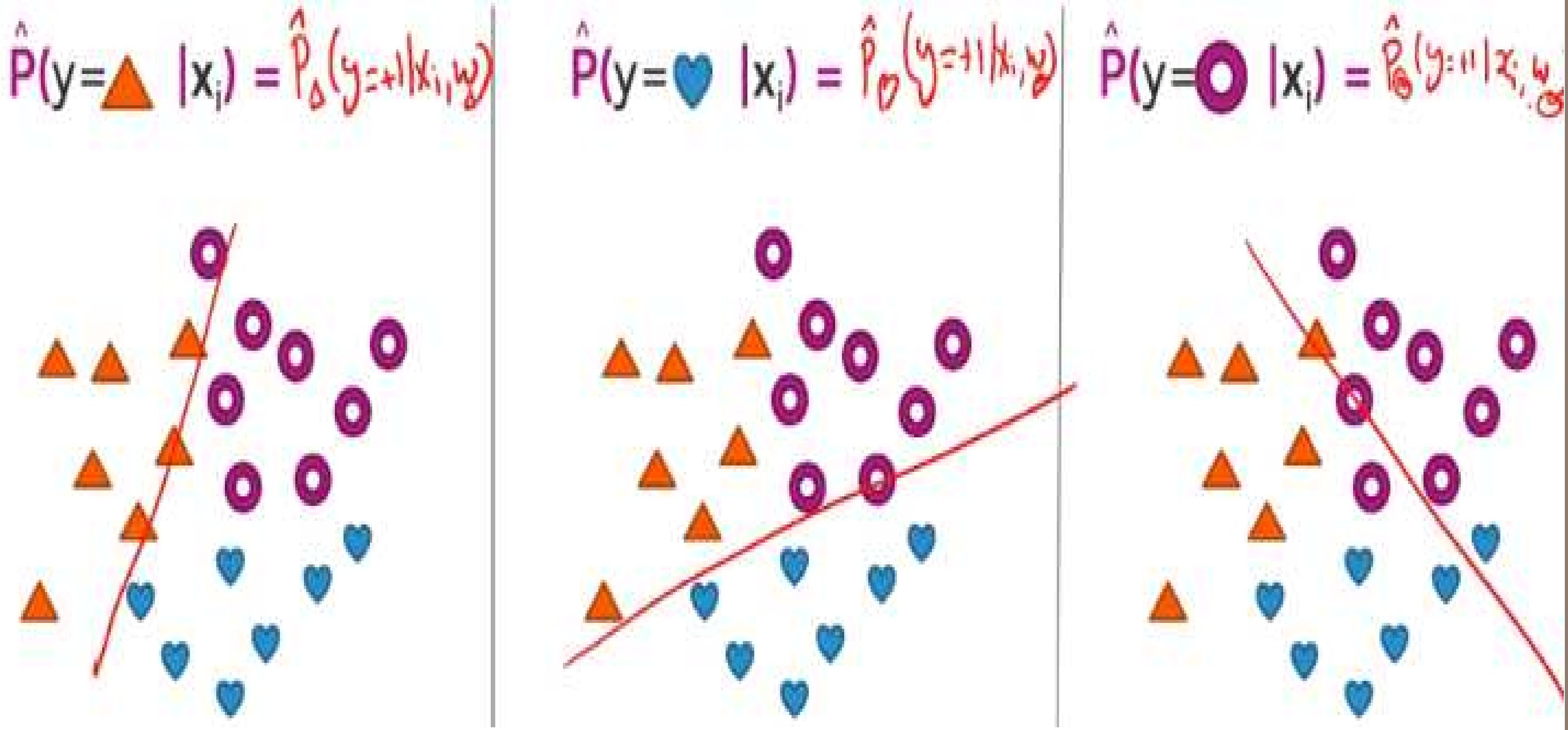
L'approche « 1 versus All(tout) »: (1 contre tous)

Cette approche Estime:

-- P(y=triangle/x) en utilisant le modèle de 2 classes.

➔+1 : Classe avec (yi=triangle)

➔-1 : Classe avec (yi= cercle ou yi=coeur)



•Ensuite on assignera xi a la classe qui a la plus grande probabilité.

•On aura donc 3 classifieurs logistique Simple**== Au Nombre de Classes**.

En général, l'algorithme de la Multiclassification est le suivant :

Début

max_prob= 0;

$\hat{Y} = 0$

For c=1,..,C:

if $\hat{P}_c = (y = +1|X_i)$ > max_prob:

$\hat{Y} = c$

max_prob = $\hat{P}_c = (y = +1|X_i)$

Fin.

contoso

Quiz Regression Logistique:

La Regression Logistique est un classifieur Non Linéaire

V F

La Régression Logistique **est** une Régression Linéaire

V F

Dans La Régression Logistique **la variable explicative X peut être qualitative**

V F

Quelle est la fonction qui compresse la valeur réelle en [0, 1]?

1-Logistic function
2-Absolute value function
3-Zero Function

Si $aT.h(\mathbf{x})>0$, qu'est ce qui est vrai pour $P(y=+1|\mathbf{x})$?

1- $P(y = +1 \mid x) < 0.5$

2- $P(y = +1 \mid x) >= 0.5$

3- on peut rien dire concer nant $P(y = +1 \mid x)$

L'algorithme Descente de Gradient est utilisé pour trouver la meilleure séparation entre classes.

V **F**

La Regression Logistique est un classifieur Binaire

V F

Vous Envisagez de former un classificateur multiclasse « 1 contre tous » pour le problème de la reconnaissance des chiffres à l'aide de la régression logistique.

On utilise la base données DIGIT qui contient les chiffres de 0 à 9.

Combien de classificateurs de régression logistique devrons-nous former ?

l'algorithme Ascent Gradient :

Début

Initialiser a(1)=0 (ou bien aléatoirement ou bien intelligemment)

Initialiser ε:tolérance

Initialiser : η le 'pas'

While ||grad(l (a(t)))||> ε (not converged)

For j=0..D

Partial[j]=$\sum_{i=1}^{n} hj(xi)(v[yi=+1]-\ p(yi=+1lxi,a))$

$a_j(t+1)$ 🡐 $a_j(t)$ + η partial[j]

t🡐t+1

Fin

Partial[j] est la dérivée de l(a) par rapport à a_j,

Elle est définie comme suit :

$$\nabla l(a) = \left(\frac{\partial l(a)}{\partial a0}, \quad , \frac{\partial l(a)}{\partial aj}, \ldots . \right)$$

$$\frac{\partial l(a)}{\partial aj} = \sum_{i=1}^{n} hj(xi)(v[yi = +1] - p(yi = +1/xi, a))$$

Différence entre la réalité et la prédiction

$v[yi = +1]$ est appelé indicateur de la fonction = 1 si yi=+1

=0 sinon

Apprentissage de l'algo AscG pour la régression logistique:

$$\frac{\partial l(w)}{\partial wj} = \sum_{i=1}^{n} hj(xi)(v[yi = +1] - p(yi = +1 l xi, w))$$

Différence entre la réalité et la prédiction

cette derivative pousse les coefficients a augmenter si yi=+1 si les exemples d'apprentissage positifs
Et a diminuer si yi=-1 si les exemples d'apprentissage négatifs