USTHB 2021/2022

Faculté d'Informatique

M1 IV/SII

Examen Réseaux de Neurones et Apprentissage Automatique

Exercice 1 :

L'entreprise α-num spécialisé en vente de matériel informatique, voudrait prédire son chiffre d'affaire en fonction de son investissement en publicité.

Pour ce faire, α-num voudrait utiliser un algorithme d'apprentissage automatique en utilisant des données récoltées durant les années précédentes et présentées dans le tableau suivant :

Année	Investissement en publicité (x1000 da)	Chiffre d'affaires (x1000 da)
1	40	95,6
2	54	906,3
3	68	199,8
4	92	19,1
5	77	149

En tant qu'analyste de données, vous avez été chargé de proposer une solution au problème de α-num en répondant aux préoccupations suivantes :

1- Entre régression linéaire et régression logistique, quelle méthode serait la plus appropriée pour répondre au problème de α-num ? Justifiez votre réponse.

On suppose que la méthode choisie pour répondre au problème de α-num soit la régression linéaire, et on suppose que les paramètres initiaux soient : $\theta_{0\_init} = 0.207 \quad \theta_{1\_init} = 4.794$

2- Donner les nouveaux paramètres (θ_0, θ_1) de ce modèle de régression linéaire après 1 itération en utilisant l'algorithme de descente du gradient sans régularisation et avec un taux d'apprentissage $\alpha = 0.0004$.

3- Donner l'erreur quadratique de l'ensemble de test en supposant que le modèle entrainé ait donnée les probabilités suivantes et que les vérités terrain sont comme suit :

Investissement en publicité (x1000 da)	$h_\theta(x)$	Chiffre d'affaires (x1000 da)
23	49,162	65,1
26	55,549	76,2
43	91,746	119
58	123,684	151,8

Exercice 2 :

Afin d'orienter sa publicité, un concessionnaire automobile voudrait analyser le type de voiture acheté par une personne en fonction de son âge et de son salaire. L'objectif étant de prédire pour un âge et un salaire particulier, le type de voiture achetée (luxe, économique ou standard). Pour effectuer cette analyse, le concessionnaire a choisi le réseau de neurones de la Figure 1.

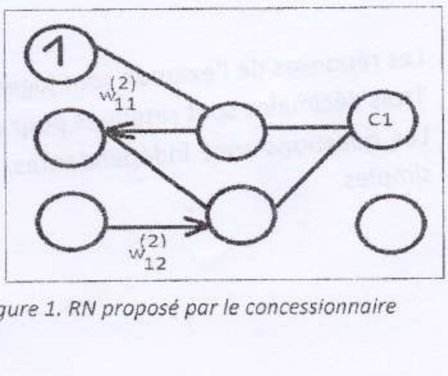

Figure 1. RN proposé par le concessionnaire

1- Le réseau de neurones proposé par le concessionnaire est incomplet et/ou incorrect. En gardant le même nombre de couches, corrigez le schéma du réseau de neurones précèdent. Notez que pour ce réseau, les hyper plans issues des fonctions linéaires sont supposées ne pas passer par l'origine. Notez aussi que ce réseau est un réseau complétement connecté.

Les paramètres de ce réseau à l'itération i étaient comme suit :

$b_1^{(2)} = 0.068$, $b_2^{(2)} = -0.018$, $w_{11}^{(2)} = 0.013$, $w_{21}^{(2)} = -0.069$, $w_{12}^{(2)} = 0$, $w_{22}^{(2)} = -0.037$
$b_1^{(3)} = -0.187$, $b_2^{(3)} = -0.294$, $b_3^{(3)} = -0.056$, $w_{11}^{(3)} = -0.524$, $w_{21}^{(3)} = -0.334$
$w_{12}^{(3)} = -0.112$, $w_{22}^{(3)} = -0.327$, $w_{31}^{(3)} = 0.134$, $w_{32}^{(3)} = -0.064$

En supposant que l'activation du réseau est une Sigmoïde,

2- Donner les prédictions de la classe voiture de luxe en utilisant les paramètres précédents pour l'exemple : (âge, salaire)=(56,153).

3- Donner la dérivée partielle du coût par rapport au poids $w_{11}^{(2)}$ en utilisant les paramètres précédents pour l'exemple : (âge, salaire, type de voiture) = (56,153, Luxe).

Nous supposons qu'à l'itération i+k, les exemples, leurs vérités et les prédictions (probabilités en pourcentage) obtenues par le réseau sont données dans le tableau suivant :

Age	Salaire	Type de voiture	% classe Luxe	% classe Economique	% classe Standard
55	72	Luxe	0.25	0.35	0.15
56	153	Luxe	0.85	0.45	0.28
28	28	Economique	0.15	0.15	0.55
52	159	Luxe	0.45	0.35	0.25
46	24	Economique	0.29	0.38	0.76
30	23	Economique	0.55	0.95	0.25
32	32	Standard	0.12	0.12	0.85
24	35	Standard	0.35	0.35	0.45
49	35	Standard	0.25	0.35	0.75

4- La somme des probabilités pour les 3 classes n'est pas égale 1, à votre avis pourquoi ?

5- Tracer la matrice de confusion de ce réseau.

6- Tracer la courbe ROC (Receiver Operating Characteristic) des trois classes pour les valeurs de seuil : 0, 0.5, 1. Que pensez-vous de ce classifieur ?

7- En supposant que ce classifieur souffre de sur-apprentissage. Quelles mesures proposeriez-vous afin d'éviter le sur-apprentissage ? (citez deux solutions possibles).

Etant les ressources limitées du concessionnaire, celui-ci voudrait utiliser un classifieur de type Bayes naïf pour la classification de voitures en fonction de l'âge et du salaire d'une personne.

8- Proposez un algorithme basé sur Bayes naïf pour cette classification.

NB :

- Les réponses de l'examen sont jugées correctes en fonction du résultat ET de la méthode.
- Trois décimales sont retenues pour chaque calcul.
- Les questions sont indépendantes, il est préférable de commencer par les questions plus simples.

Bon courage

Equation normale sans et avec régularisation

$$\theta = (X^TX)^{-1}X^Ty \quad , \quad \theta = \left(X^TX + \lambda \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} X^Ty$$

Régression linéaire sans et avec régularisation

$$\mathbf{J}(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \text{ , } \boldsymbol{\Theta}\mathbf{j} \leftarrow \boldsymbol{\Theta}\mathbf{j} - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$

$$\mathbf{J}(\theta) = \frac{1}{2m} \left[\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2 \right] \text{, } \boldsymbol{\Theta j} \leftarrow \boldsymbol{\Theta j} \left(1 - \alpha \frac{\lambda}{m}\right) - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$

$$\boldsymbol{\Theta j} \leftarrow \boldsymbol{\Theta j} \left(\mathbf{1} - \boldsymbol{\alpha} \frac{\lambda}{m}\right) - \boldsymbol{\alpha} \frac{\mathbf{1}}{m} \sum_{i=1}^{m} (\boldsymbol{h_\theta(x^{(i)}) - y^{(i)}})\, \boldsymbol{x_j^{(i)}} \quad j \neq 0$$

Régression logistique sans et avec régularisation

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \quad y^{(i)} log(h_\theta(x^{(i)}) + (1 - y^{(i)})\, log(1 - h_\theta(x^{(i)}) \text{ , } \boldsymbol{\Theta}\mathbf{j} \leftarrow \boldsymbol{\Theta}\mathbf{j} - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \quad [-y^{(i)} log(h_\theta(x^{(i)}) - (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m} \sum_{i=1}^{n} \quad \theta_j^2$$

$$\boldsymbol{\Theta j} \leftarrow \boldsymbol{\Theta j} \left(\mathbf{1} - \boldsymbol{\alpha} \frac{\lambda}{m}\right) - \boldsymbol{\alpha} \frac{\mathbf{1}}{m} \sum_{i=1}^{m} (\boldsymbol{h_\theta(x^{(i)}) - y^{(i)}})\, \boldsymbol{x_j^{(i)}}$$

Réseau neuronal

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} \left[-y_k^{(i)} \log(h_\theta(x^{(i)}))_k - (1 - y_k^{(i)}) \log(1 - h_\theta(x^{(i)}))_k \right]$$

$$\delta_j^{(L)} = 2\left(\hat{a}_j - a_j^{(L)}\right) a_j^{(L)} \left(1 - a_j^{(L)}\right) \quad , \quad \delta_j^{(l)} = \sum_k \quad \delta_k^{(l+1)} w_{kj}^{(l+1)}\, a_j^{(l)} \left(1 - a_j^{(l)}\right)$$

$$\frac{\partial C}{\partial w_{jk}^{(l)}} = \delta_j^{(l)} a_k^{(l-1)} \quad , \frac{\partial C}{\partial b_j^{(l)}} = \delta_j^{(l)} \quad , \quad \Delta w_{jk}^{(l)} = \Delta w_{jk}^{(l)} + \frac{\partial C}{\partial w_{jk}^{(l)}} \quad , \Delta b_j^{(l)} = \Delta b_j^{(l)} + \frac{\partial C}{\partial b_j^{(l)}}$$

$$\frac{\partial C}{\partial w_{jk}^{(l)}} = \frac{1}{m} \Delta w_{jk}^{(l)} \quad , \frac{\partial C}{\partial b_j^{(l)}} = \frac{1}{m} \Delta b_j^{(l)} \quad , \quad w_{jk}^{(l)} \leftarrow w_{jk}^{(l)} - \alpha \frac{\partial C}{\partial w_{jk}^{(l)}} \quad , b_j^{(l)} \leftarrow b_j^{(l)} - \alpha \frac{\partial C}{\partial b_j^{(l)}}$$

Gaussienne uni variée et multi variée

$$P(x) = \prod_{j=1}^{n} \quad P(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^{n} \quad \frac{1}{\sqrt{2\pi}\sigma_j} e^{\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right)} \text{ , } P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}$$

$$p(x \mid \omega_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}\left(x - \mu_j\right)^T \Sigma_j^{-1} \left(x - \mu_j\right)\right)$$

Métriques

$$Précision_{classei} = \frac{VP}{VP+FP} \quad , \quad Rappel_{classei} = \frac{VP}{VP+FN} \quad , \quad Taux\ de\ FP_{Classei} \frac{FP}{FP+VN}$$

$$Spécificité_{classei} \frac{VN}{VN+FP} = 1 - \text{Taux de FP} \quad , \quad Mesure\ F = 2 \frac{Precision\ x\ Rappel}{Precision + Rappel}$$

Autre

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$