TD réseau de neurones

Adrien Le Coz, Pol Labarbarie
Gianni Franchi, Adrien Chan-Hon-Tong

Remarque bien classer le point x de label $y(x)$ avec un modèle f c'est avoir $y(x)f(x) > 0$ ou $y(x) = sign(f(x))$ mais pas nécessairement $f(x) = y(x)$ (f renvoie un "score" d'être dans la classe 1, et on seuille à 0 dans le cas binaire). Dans le cas multiclass, f produit R scores (1 par classe) et on voudra $\arg\max_r f(x) = y(x)$.

Notation $relu(x) = max(x, \mathbf{0}) = [x]_+$ il s'agit bien du max composante par composante (quand l'entrée est un vecteur).

Partie 1 : Extrait de l'exam de 2021 : Construction d'une famille universelle

Q1.1 Rappeler pourquoi $\phi_x(p) = relu\left(1 - \sum_d ([p_d - x_d]_+ + [x_d - p_d]_+)\right)$ vérifie $\phi_x(x) = 1$ et $\forall p, ||x - p||_1 \geq 1 \Rightarrow \phi_x(p) = 0$

Q1.2 Remontrer comment on peut construire un réseau de neurone f qui apprend par coeur la base $(x_1, y_1), ..., (x_N, y_N) \in \mathbb{R}^D \times \{-1, 1\}$ à l'aide de $\phi_{x_1}, ..., \phi_{x_N}$ (en supposant $\forall i, j,\ ||x_i - x_j||_1 \geq 1$).

Rappeler combien de neurone il faut pour cela.

Q1.3 On va construire une autre famille universelle. On suppose $x_1, ..., x_N \in \mathbb{R}^D$ avec $||x_1||_2 = ... = ||x_N||_2 = 1$ et $\forall i, j,\ x_i \neq x_j$. On introduit $\delta = relu(\max_{i \neq j} \sum_d x_{i,d} x_{j,d})$ (c'est à dire le relu du maximum des produits scalaires de x_i et x_j).

Montrer que $\delta < 1$.

On rappelle que la norme 2 c'est la racine du produit scalaire du vecteur avec lui même et qu'on a l'inégalité de Cauchy...

Q1.4 On note $\psi_n(p) = relu(\sum_d p_d \times x_{n,d} - \delta)$.

Montrer que $\forall n, \psi_n(x_n) > 0$.

Montrer que $\forall i \neq j, \psi_i(x_j) = 0$.

Q1.5 montrer comment on peut construire un réseau de neurone f qui apprend par coeur la base $(x_1, y_1), ..., (x_N, y_N) \in \mathbb{R}^D \times \{-1, 1\}$ à l'aide de $\psi_1, ..., \psi_N$ (en supposant $||x_1|| = ... = ||x_N|| = 1$ et $\forall i, j,\ x_i \neq x_j$).
On rappelle que apprendre par coeur ça veut dire que $\forall y_1, ..., y_N$ la fonction qu'on construit a le même signe que y_n en x_n ($\forall n$). Dit autrement, on ne demande que $y_n f(x_n) > 0$ mais pas forcément $f(x_n) = y_n$.

Combien de neurone il faut pour cela ?

Pour votre culture, il s'agit de la famille universelle générique la plus économe en neurones. Mais il y en a plein d'autres, ça permet de faire 1 examen différent chaque année...

Partie 2 : Apprentissage 1D

Q2 : Chercher w_1, w_2, w_3, b tel que le réseau 1D $h(x, w) = w_1[x]_+ + w_2[x - 1]_+ + w_3[x - 2]_+ + b$ vérifie

— $h(0, w) > 0$ (par exemple 1)

— $h(1,w) < 0$ (par exemple -1)
— $h(2,w) > 0$ (par exemple 1)
— $h(3,w) < 0$ (par exemple -1)
Oui, ça aussi ça permet de faire facilement un exercice d'examen...

Partie 3 : quelques résultats remarquables

Q3.1 On considère la fonction $f(x) = f((x_1 \;\; x_2)^T) = x_2 - relu(x_1 - x_2)$. Déterminez les zones où f est positive vs négative.

Q3.3 Même questions avec $g((x_1 \;\; x_2)^T) = x_2 + relu(x_1 - x_2)$ et $h((x_1 \;\; x_2)^T) = x_1 + relu(x_2 - x_1)$, que remarquez vous ?

Q4 : Considérons la base de données $((0 \;\; 2)^T, 1)$, $((0 \;\; -2)^T, 1)$, $((2 \;\; 0)^T, 1)$, $((-2 \;\; 0)^T, 1)$, $((0 \;\; 0)^T, -1)$, ainsi que les 2 réseaux

— $\psi(x) = [(0 \;\; 1)x]_+ + [(0 \;\; -1)x]_+ + [(1 \;\; 0)x]_+ + [(-1 \;\; 0).x]_+ - 1$
— $\phi(x) = 2relu((-1 \;\; 1)x - 1) + 2relu((1 \;\; -1)x - 1) - 1$

Q4.1 : Dessiner la base et donner la frontière de décision que vous considéreriez comme *naturelle* au vu de cette base de données.

Q4.2 : Montrez que les 2 réseaux apprennent la base par coeur. Dessinez les zones positives et négatives.

Q4.3 : Donnez la structure de chaque réseau.

cette base est intéressant car il est possible de l'apprendre avec un réseau de 3 neurones. Mais la solution obtenue est asymétrique. Pour obtenir une solution symétrique et bornée, il faut 5 neurones. Ici, plus de paramètres permet d'obtenir une solution plus élégante...