TD réseau de neurones

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 $relu(x) = max(x, \mathbf{0}) = [x]_+$.

Partie 1 : Extrait de l'exam de l'an dernier : 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 $\forall p, ||x - p||_1 \geq 1 \Rightarrow \phi_x(p) = 0$

- $\phi_x(x) = relu(1 - 0) = 1$
- $[p_d - x_d]_+ + [x_d - p_d]_+ = ||x - p||_1$ donc $||x - p||_1 \geq 1 \Rightarrow \phi_x(p) = relu(1 - plusgrandque1) = relu(negatif) = 0$

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

- $f(p) = \sum_n y_n \phi_{x_n}(p)$ car $\forall n\ f(x_n) = 0+0+...+y_n \times 1+0+0+... = y_n$

Rappeler combien de neurone il faut pour cela.

- il faut 2 neurone par point et par composante en 1ère couche (donc $2ND$)
- 1 neurone par point en 2ème couche (donc N)
- 1 neurone pour tout regrouper donc, au final, $2ND$ puis N puis 1

Précisément

$$f(x) = y^T \left[\begin{pmatrix} 1 & 1 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & ... \\ ... & & & & & \end{pmatrix} \left[\begin{pmatrix} I \\ -I \\ I \\ -I \\ ... \end{pmatrix} x - \begin{pmatrix} -x1 \\ x_1 \\ -x_2 \\ x_2 \\ ... \end{pmatrix} \right]_+ + \mathbf{1} \right]_+$$

la 2ème couche c'est

line 1 : 2D 1 puis des 0

line 2 : 2D 0 puis 2D 1 puis des 0 etc

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 note $\delta = relu(\max_{i \neq j} \sum_d x_{i,d} x_{j,d})$ (c'est à dire le maximum des produits scalaires de x_i et x_j).

Montrer que $\delta < 1$.

en notant le produit scalaire avec la simple juxtaposition

- $x_i x_j \leq ||x_i|| ||x_j|| \leq 1$
- la condition d'égalité c'est qu'ils soient positivement colinéaires - mais ils ont même normes donc ils seraient égaux - or on suppose que non

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

- $\psi_n(x_n) = relu(1 - \delta)$ et $\delta < 1$ donc $\psi_n(x_n) > 0$
- $\psi_i(x_j) = relu(x_i x_j - \max_{a,b} x_a x_b) = relu(negatif) = 0$

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 $\forall i, j,\ x_i \neq x_j$).

- $f(p) = \sum_n y_n \psi_{x_n}(p)$ car $\forall n\ f(x_n) = 0 + 0 + ... + y_n \times 1 + 0 + 0 + ... = y_n$
- (quand on a construit une indicatrice, c'est toujours pareil après)

Combien de neurone il faut pour cela ?

- seulement 1 par point et 1 pour réunir c'est à dire N puis 1 ! On a gagné un étage entier...

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)

$h(x) = 1 - 2[x]_+ + 4[x - 1]_+ - 4[x - 2]_+$ marche car

- $h(0) = 1$
- $h(1) = 1 - 2 = -2$
- $h(2) = 1 - 2 \times 2 + 4 = 1$
- $h(3) = 1 - 2 \times 3 + 4 \times 2 - 4 = 1 - 6 + 8 - 4 = -1$

mais il y a une infinité de solution.

notamment $2h$ ou $\frac{1}{2}h$ marche aussi car seul le signe compte !

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.

— déjà relu est toujours positif, donc si $x_2 < 0$, f est forcément négative
— maintenant même si $x_2 > 0$, f pourrait être négative en fonction de $x_1 - x_2$:
— soit $x_1 < 2x_2$ et $f(x) > 0$
— soit $x_1 > 2x_2$, et $f(x) < 0$

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 ?

Si $x_1 > x_2$, $g(x) = x_1$ et si $x_1 < x_2$, $g(x) = x_2$ donc $g(x) = \max(x_1, x_2)$, donc $g(x) < 0 \Leftrightarrow x_1 < 0$ et $x_2 < 0$.

$$g(x) = (1 \quad -1 \quad 1) relu\left(\begin{pmatrix} 0 & 1 \\ 0 & -1 \\ 1 & -1 \end{pmatrix} x\right)$$

$g = h$ alors qu'ils s'écrivent différemment - c'est une *symétrique* cachée.

Notez au passage qu'on peut construire le max de 2 neurones avec des relu, il faudra vous en rappeler pour le cours suivant, pour relativiser les nouveautés introduites dans les CNN notamment le max pooling...

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.

C'est 4 points de label 1 en croix et 1 point de label -1 au centre.

Personnellement moi je ferais un rond centré en 0 mais à chacun sa frontière "naturelle"

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

Faut tester en chaque point pour vérifier que la base est apprise par coeur. Pour les zones, c'est un losange pour le premier - un bande pour le deuxième.

Q4.3 : Donnez la structure de chaque réseau.

$[(0\ \ 1)x]_+ + [(0\ \ -1)x]_+ + [(1\ \ 0)x]_+ + [(-1\ \ 0).x]_+ - 1$

$$= (1\ 1\ 1\ 1) relu\left(\begin{pmatrix} 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \end{pmatrix} x\right) - 1$$

On a donc 1 couche de 4 et 1 couche de 1.

$2relu((-1\ \ 1)x - 1) + 2relu((1\ \ -1)x - 1) - 1$

$$= (2\ 2) relu\left(\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} x - \begin{pmatrix} -1 \\ -1 \end{pmatrix}\right) - 1$$

On a donc 1 couche de 2 et 1 couche de 1.

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...