

PFE

Khaled MEDJKOUH
Davidson Lova RAZAFINDRAKOTO

3 avril 2023

Table des matières

1	Introduction	3
1.1	Réseau de neurones	3
1.2	Source d'incertitudes dans un réseaux de neurones	3
1.2.1	Acquisition des données	3
1.2.2	Structure du model	3
1.3	Prédiction	3
1.3.1	Incertitude aléatoire	3
1.3.2	Incertitude épistémique	4
2	Réseau de neurones bayésiens (BNN)	5
2.1	Méthodes variationelles	5
2.2	Méthode de Laplace	5
2.3	Méthodes par échantillonnage ou Monte Carlo	5
3	Algorithme BNN-ABC-SS	6
3.1	ABC (Approximate Bayesian Computation)	6
3.2	SS (Subset Simulation)	6
3.3	Pseudo - Code	6
4	Réalisation	7
4.1	Cosinus perturbé	7
4.2	Sinus perturbé	7
5	Conclusion et perspective	7

1 Introduction

Incertitudes utile dans plusieurs domaines et la quantification duquel est nécessaire pour des évaluations de risques, détection d'anomalie et prise décision.

1.1 Réseau de neurones

- + Historique sur les réseaux de neurones
- + Explication des poids et biais et fonction d'activation
- + Quelques résultats et application
- + Ce qui nous amène à un problème

1.2 Source d'incertitudes dans un réseaux de neurones

On modélise le réseaux de neurones comme la fonction non linéaire

$$f : (x, \theta) \in \mathcal{X} \times \Theta \mapsto f(x, \theta) \in \mathcal{Y}$$

où

- $\mathcal{X} \subset \mathbb{R}^{n_e}$, l'espace des variables d'entrées
- $\mathcal{Y} \subset \mathbb{R}^{n_s}$, l'espace des variables de sorties
- $\Theta \subset \mathbb{R}^{n_p}$, l'espace des paramètres

On se donne maintenant une base de données d'entraînement $D = \{(x_i, y_i)\}_{i=1}^N \in (\mathcal{X} \times \mathcal{Y})^N$.

1.2.1 Acquisition des données

Si on prend comme données d'entrées x une mesure d'une quantité réel \tilde{x} .

Il y a une variabilité sur la mesure en fonction des circonstances et conditions $\omega \in \Omega$ dans lequel la mesure a été effectué.

La sortie y peut aussi subir des erreurs de labélisation (pour le cas d'une tâche de classification) ou aussi de mesure si c'est une quantité mesuré.

En somme on a une incertitude sur l'entrée $x|\omega \sim p_{x|\omega}$ et $y|\omega \sim p_{y|\omega}$.

S'ajoute à ça, l'incertitude sur x qui peut se propager sur le y .

1.2.2 Structure du model

Les paramètres θ et donc l'espace Θ est variable en fonction du choix de model s .

On a $\theta|D, s \sim p_{\theta|D,s}$

1.3 Prédiction

La distribution de la prédiction y^* sachant une entrée x^* est donnée par

$$p(y^*|x^*, D) = \int_{\Theta} \underbrace{p(y^*|x^*, \theta)}_{\text{Données}} \underbrace{p(\theta|D)}_{\text{Modèle}} d\theta$$

1.3.1 Incertitude aléatoire

Insert definition here

Elle affecte la partie $p(y^*|x^*, \theta)$ de la tâche de prédiction, elle est due à la variabilité (précision et erreurs) lors des mesures. Elle est donnée et inhérent au problème.

1.3.2 Incertitude épistémique

Insert definition here

Affecte la partie $p(\theta|D)$ de la tâche de prédiction, elle due :

- à la complexité du modèle,
- aux erreurs durant la phase d'entraînement,
- à la manque d'information à cause de données manquantes ou la capacité de représentation des données d'entraînement.

2 Réseau de neurones bayésiens (BNN)

Les réseaux de neurones bayésiens sont des réseaux de neurones dont les poids sont, non pas des quantités déterministes (comme dans le cas d'un NN normale) mais des distributions.

A l'initialisation, les poids suivent une loi a priori $p(\theta)$, et l'entraînement consiste à évaluer l'a posteriori de cette loi conditionnée aux données d'entraînement $p(\theta|D)$.

On ne dispose pas dans le cas générale, d'une formule analytique de cette distribution a posteriori.

Voici trois familles de méthodes pour approcher cette distribution :

- Méthodes variationnelle
- Méthodes par échantillonnage ou Monte Carlo (qu'on va voir dans la suite)
- Méthodes de Laplace

2.1 Méthodes variationnelles

Ici on approche $p(\theta|D)$ par une famille de distribution paramétrique $\{q^\gamma(\theta)\}_\gamma$ (souvent des gaussiennes). Le but est de choisir γ qui rapproche $q^\gamma(\theta)$ le plus de $p(\theta|D)$. La distance choisit ici est la divergence de Kullback-Leibler :

$$KL(q||p) = \mathbb{E}_q \left[\log \frac{q^\gamma(\theta)}{p(\theta|D)} \right]$$

Comme on ne connaît pas $p(\theta|D)$, on utilise l'ELBO (evidence lower bound) qui est égal à la divergence à une constante paramètres

$$L = \mathbb{E}_q \left[\log \frac{p(y|D, \theta)}{q^\gamma(\theta)} \right]$$

(on a en effet $KL(q||p) = -L + \log p(y|D)$)

2.2 Méthode de Laplace

$\hat{\theta}$ l'estimateur de maximum d'a priori

$$\log p(\theta|D) \approx \log p(\hat{\theta}|D) + \frac{1}{2}(\theta - \hat{\theta})^T (H + \tau I)(\theta - \hat{\theta})$$

$$p(\theta|D) \sim \mathcal{N}(\hat{\theta}, (H + \tau I)^{-1})$$

2.3 Méthodes par échantillonnage ou Monte Carlo

La formule de Bayes nous donne

$$p(\theta|D) = \frac{p(D|\theta)}{p(D)} p(\theta)$$

- $p(D|\theta)$ la vraisemblance des données D sachant le paramètre θ ,
- $p(\theta)$ la distribution a priori de θ ,
- $p(D)$ la distribution des données d'entraînement.

$$p(\theta|D) \propto$$

3 Algorithme BNN-ABC-SS

3.1 ABC (Approximate Bayesian Computation)

La méthode ABC consiste à évaluer $p(\theta|D)$ sans évaluer la vraisemblance qui peut s'avérer coûteux. Posons $\hat{y} = f(x, \theta)$ la sortie d'une évaluation de x par réseaux de neurones f avec paramètre θ . La formule de Bayes donne

$$p(\theta, \hat{y}|D) \propto p(D|\hat{y}, \theta)p(\hat{y}|\theta)p(\theta)$$

Pour simuler selon la distribution du second membre, on applique l'Algorithme de rejet.

3.2 SS (Subset Simulation)

+ celui dans le rapport d'avant

3.3 Pseudo - Code

+ celui dans le rapport d'avant

4 Réalisation

+ On se donne un base de données à étudier avec une comparasion avec des méthodes déjà existantes

4.1 Cosinus perturbé

4.2 Sinus perturbé

5 Conclusion et perspective