

Réseaux de neurones artificiels pour le classement du cancer du sein

1 Présentation du problème

Le cancer du sein est le deuxième cancer dans la population après celui de la prostate¹. Bien que le pronostic associé au cancer du sein soit en général positif, en 2012 presque 1,7 millions de cas ont été diagnostiqués. L'imagerie médicale est la première étape du diagnostic², suivi par une ou plusieurs biopsies. Cela génère une montagne de données qui, aujourd'hui, sont traitées principalement par des experts humains. Cependant, les techniques modernes d'intelligence artificielle devraient pouvoir simplifier la tâche au personnel médical.

2 Travail demandé

Le but du projet est d'entraîner des réseaux de neurones artificiels pour discriminer entre les cancers du sein bénin et malin, récurrent et non récurrent.

3 Les données

Les données (disponibles sur Arche) proviennent de [1-2]. Elles sont librement accessibles sur le UCI Machine Learning Repository³. N'hésitez pas à ouvrir ce lien car vous y trouverez plus de détails sur le problème.

Le choix de la plate-forme de développement et du langage de programmation est libre.

4 Evaluation des Projets

Les projets seront évalués par une soutenance qui aura lieu le 8 et 15 Décembre 2017 en fonction de vos disponibilités. Les fichiers avec le code source et la présentations doivent être soumis sur Arche sous forme d'archive compressée portant le nom "NomDuProjet-Nom1-Nom2-Nom3.zip", où NomX sont les noms des membres du groupe. La date limite

¹<https://curie.fr/dossier-pedagogique/cancers-du-sein-toutes-les-infos>

²<https://curie.fr/dossier-pedagogique/cancer-du-sein-la-biopsie-pour-confirmer-le-diagnostic>

³<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28original%29>

du dépôt est le **7 Décembre 2017 à 23h55** au plus tard. Au delà de cette date la soumission sera bloquée.

4.1 Soutenances

Imaginez que votre groupe est une start-up. Vous aurez 5 minutes pour faire un “pitch” de votre produit à huit investisseurs potentiels, en essayant de les convaincre à investir dans votre entreprise. Pour faire cela, il faudra que vous démontriez d’avoir maîtrisé les techniques utilisées et de savoir correctement évaluer les performances de votre produit. Les sept investisseurs seront les membres de deux autres groupes et les responsables du module. La présentation sera suivie d’une session de questions, posées par les investisseurs, durant environ 10 minutes.

Les présentations doivent se concentrer sur les **choix de conception** que vous avez fait, les **suppositions** et les **hypothèses** de votre modèle, le choix des **paramètres** et **l’optimisation de leurs valeurs**, **l’analyse des résultats** obtenus, **l’évaluation de la performance** du modèle, ses **limitations** et les **améliorations** envisagées.

Vu que le temps est très limité il ne faudra pas le gâcher en expliquant que les réseaux de neurones artificiels calculent le produit scalaire entre le vecteur d’inputs et le vecteur de poids. Ce sont des informations non nécessaires que nous connaissons tous déjà. Après tout, nous sommes tous des (futurs) ingénieurs.

5 References

- [1] Wolberg, W.H., & Mangasarian, O.L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In Proceedings of the National Academy of Sciences, 87, 9193–9196.
- [2] Zhang, J. (1992). Selecting typical instances in instance-based learning. In Proceedings of the Ninth International Machine Learning Conference (pp. 470–479). Aberdeen, Scotland: Morgan Kaufmann.