

Mini Projet : Réseau Bayésien (Étude de Cas dans le Domaine Médical - Diagnostic du Cancer du Sein)

Description du projet :

Dans ce projet, vous utiliserez un ensemble de données provenant du Centre Médical Universitaire, Institut d'Oncologie, à Ljubljana, en Yougoslavie. Cet ensemble de données contient des informations sur le cancer du sein et sera utilisé pour construire un réseau bayésien **afin d'aider au diagnostic de cette maladie.**

Données :

Vous utiliserez un ensemble de données contenant des informations sur 286 instances de tumeurs, certaines **étant malignes et d'autres bénignes**. Les instances sont décrites par 10 attributs :

- Class : Le diagnostic de la tumeur (no-recurrence-events, recurrence-events)
- Age : L'âge du patient (10-19, 20-29, ..., 90-99)
- Menopause : L'état de ménopause du patient (lt40, ge40, premeno)
- Tumor-size : La taille de la tumeur (0-4, 5-9, ..., 55-59)
- Inv-nodes : Le nombre de ganglions lymphatiques envahis (0-2, 3-5, ..., 36-39)
- Node-caps : La présence ou non de capsules des ganglions lymphatiques (yes, no)
- Deg-malig : Le degré de malignité de la tumeur (1, 2, 3)
- Breast : Le sein affecté par la tumeur (left, right)
- Breast-quad : Le quadrant du sein affecté par la tumeur (left-up, left-low, ..., central)
- Irradiat : Le traitement par irradiation (yes, no)

Lien de téléchargement : [Breast Cancer - UCI Machine Learning Repository](#)

Partie 01 : Apprentissage de la structure

Avant de construire la structure du réseau bayésien, consultez un médecin pour obtenir des informations sur les relations entre les différentes caractéristiques des tumeurs et leur diagnostic de cancer du sein.

Cette tâche comprend plusieurs étapes qui sont :

- **Identification des variables :**

Identifiez les variables pertinentes pour le diagnostic du cancer du sein à partir de l'ensemble de données.

- **Construction de la structure du réseau :**

À l'aide des variables identifiées et des informations obtenues lors **de la consultation avec le médecin**, construisez la structure du réseau bayésien en définissant les liens de dépendance probabiliste entre elles.

- **Représentation graphique :**

Présentez la structure du réseau bayésien sous forme graphique.

Partie 02 : Apprentissage des paramètres

Utilisez l'algorithme de maximisation de l'espérance (EM) pour apprendre les paramètres du réseau bayésien à partir des données du dataset.

Partie 03 : Inférence et diagnostic

Utilisez le réseau bayésien final pour effectuer l'inférence et le diagnostic médical. Expliquez comment vous pouvez utiliser le réseau pour estimer la probabilité d'un diagnostic particulier en fonction des caractéristiques observées de la tumeur.

Rapport : Résumez vos résultats dans un rapport qui inclut :

- Une introduction
- Une description des variables sélectionnées et de leur importance pour le diagnostic du cancer du sein, basée sur la consultation avec le médecin.
- La structure du réseau bayésien avec des explications justifiant les liens entre les variables.
- Une explication de l'algorithme EM et comment il est utilisé pour apprendre les paramètres du réseau bayésien.
- Une explication de l'inférence et du diagnostic médical basés sur le réseau bayésien final.
- Une représentation graphique du réseau bayésien (montrant la structure et les paramètres).
- Toute observation ou conclusion pertinente sur la construction du réseau et son application au diagnostic médical.