



# MLP: Visualisation de la data

Nº Identifiant	394
⌚ Duration	1 h
↗ Epic	<a href="#">Ecole: Multi-Layer Perceptron</a>
📅 Date de début	@29 septembre 2025 13:00
📅 Date de fin	@29 septembre 2025 14:00
⌚ État	Archivé
⌚ Repetition	Unique
<input checked="" type="checkbox"/> A modifier	<input type="checkbox"/>
👤 Person	Garance Colomer
⌚ Type	École
⌚ Disponibilites	Occupée
🕒 LastEdited	@29 octobre 2025 16:20

## TODO:

- Observer les graphics
- Repérer les discriminants
- Épurer les colonnes
- Écrire une observation
- Isoler les variables retenues

## Présentation des variables du dataset:

- **id:** identifiants
- **diagnosis:** Ground truth

### Moyenne (mean):

- **radius\_mean:** rayon moyen du noyau cellulaire → Moyenne
- **texture\_mean:** variation de l'intensité des pixels.
- **perimeter\_mean:** périmètre du noyau.
- **area\_mean:** aire de la cellule.
- **smoothness\_mean:** régularité des contours (variations locales de rayon).
- **compactness\_mean:** rapport périmètre<sup>2</sup> / aire – mesure de compacité.
- **concavity\_mean:** sévérité des concavités dans le contour.
- **concave\_points\_mean:** nombre de points concaves du contour.
- **symmetry\_mean:** symétrie de la cellule.
- **fractal\_dimension\_mean:** complexité du contour (analyse fractale).

#### Erreur Standard (se):

- **radius\_se:** rayon moyen du noyau cellulaire
- **texture\_se:** variation de l'intensité des pixels.
- **perimeter\_se:** périmètre du noyau.
- **area\_se:** aire de la cellule.
- **smoothness\_se:** régularité des contours (variations locales de rayon).
- **compactness\_se:** rapport périmètre<sup>2</sup> / aire – mesure de compacité.
- **concavity\_se:** sévérité des concavités dans le contour.
- **concave\_points\_se:** nombre de points concaves du contour.
- **symmetry\_se:** symétrie de la cellule.
- **fractal\_dimension\_se:** complexité du contour (analyse fractale).

#### Valeur Maximale observée (worst):

- **radius\_worst:** rayon moyen du noyau cellulaire.
- **texture\_worst:** variation de l'intensité des pixels.
- **perimeter\_worst:** périmètre du noyau.
- **area\_worst:** aire de la cellule.
- **smoothness\_worst:** régularité des contours (variations locales de rayon).
- **compactness\_worst:** rapport périmètre<sup>2</sup> / aire – mesure de compacité.
- **concavity\_worst:** sévérité des concavités dans le contour.
- **concave\_points\_worst:** nombre de points concaves du contour.
- **symmetry\_worst:** symétrie de la cellule.
- **fractal\_dimension\_worst:** complexité du contour (analyse fractale).

## Analyse des données par variable:

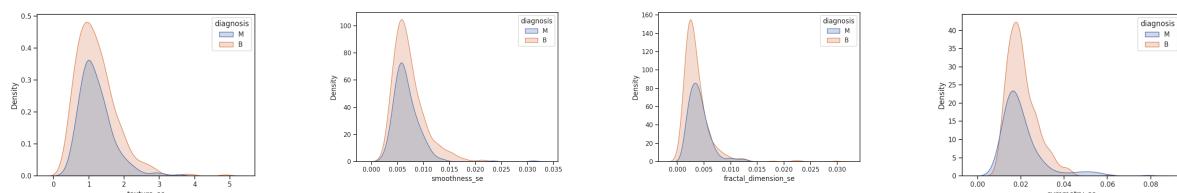
	Histogramme	Box plot simple	Box plot bi-variate	Scatterplot Matriciel:
Moyenne:				
<b>radius_mean</b>	Discriminant ok, maj entre 13 et 15	Aucun discriminant	Bon discriminant > 15, "B" < 13	
<b>texture_mean</b>	Aucun discriminant	Aucun discriminant	Aucun discriminant	
<b>perimeter_mean</b>	Bon discriminant, maj a 80 le reste au dessus de 90	Aucun discriminant	Bon discriminant, "M" se situent dessus de 100, "B" en dessous	
<b>area_mean</b>	Discriminant ok, grosse mesure sur les 500.	ok	Bon discriminant, "M" se trouvent principalement dessus de 700, "B" en dessous.	
<b>smoothness_mean</b>	Aucun discriminant	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>compactness_mean</b>	Aucun discriminant	Mauvais discriminant	Bon discriminant, "M" se situent généralement au dessus de 1000	Mauvais discriminant

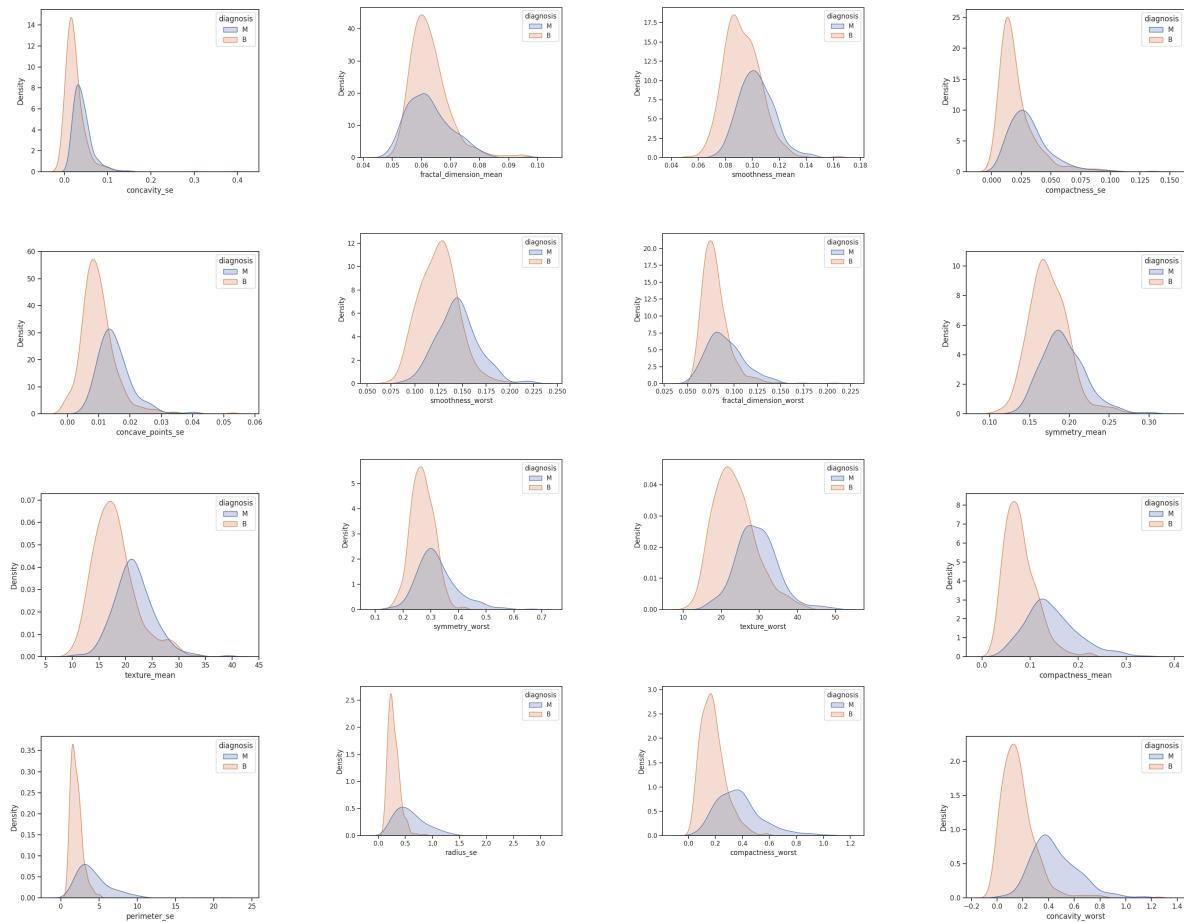
	Histogramme	Box plot simple	Box plot bi-variees	Scatterplot matriciel
			dessus de 0.10, "B" en dessous.	
<b>concavity_mean</b>	Grosse majorite a 0, le reste au dessus de 0.05	ok	Bon discriminant,moyennes distinctes, "M" au dessus de 0.1, "B" en dessous de 0.05	Mauvais discriminant
<b>concave_points_mean</b>	Discriminant ok, grosse majorite a 0.2, le reste disperse au dessus	Mauvais discriminant	Bon discriminant, les moyennes sont clairement definis, "M" au dessus de 0.60 et "B" en dessous de 0.35.	Bon discriminant
<b>symmetry_mean</b>	Aucun discriminant	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>fractal_dimension_mean</b>	Discriminant ok, separation a 0.65	Aucun discriminant	Aucun discriminant	Mauvais discriminant
Erreur Standard				
<b>radius_se</b>	Bon discriminant, maj a 0, le reste au dessus	ok +	Discriminant ok, "M" a tendance a voir des valeurs hautes	Mauvais discriminant
<b>texture_se</b>	Mauvais discriminant	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>perimeter_se</b>	Bon discriminant, maj a 0, le reste au dessus	ok	Bon discriminant, les moyennes sont distinctes, les valeurs max de "M" depassent celle de "B".	Mauvais discriminant
<b>area_se</b>	Discriminant ok, grosse mesure sur les 0-50, le reste au dessu	ok	Discriminant ok, "M" a tendance a voir des valeurs hautes	Bon discriminant
<b>smoothness_se</b>	bon discriminant, maj a 0.005, le reste au dessus	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>compactness_se</b>		ok -	Mauvais discriminant	Mauvais discriminant
<b>concavity_se</b>	Bon discriminant, majorite a 0, le reste au dessus de 0.05	ok +	Aucun discriminant	Mauvais discriminant
<b>concave_points_se</b>	Discriminant ok, grosse majorite a 0.15, le reste disperse au dessus	Aucun discriminant	Mauvais discriminant	Mauvais discriminant
<b>symmetry_se</b>	Bon discriminant, maj a 0.02	ok	Aucun discriminant	Mauvais discriminant
<b>fractal_dimension_se</b>	majorite a 0, le reste au dessus de de 0.005	ok	Aucun discriminant	Mauvais discriminant

	Histogramme	Box plot simple	Box plot bi-variees	Scatterplot matriciel
Valeur Maximale:				
<b>radius_worst</b>	Bon discriminant, maj entre 11 et 16, le reste au dessus	ok -	Bon discriminant, "M" > 17, "B" < 15	Bon discriminant
<b>texture_worst</b>	Aucun discriminant	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>perimeter_worst</b>	Bon discriminant, maj entre 75 et 100 le reste au dessus	Aucun discriminant	Bon discriminant, "M" > 115, "B" < 100	Tres bon discriminant
<b>area_worst</b>	Discriminant ok, grosse majorite de 0 - 1000.	ok	Bon discriminant, les "M" se situent minimum au dessus de 500.	Bon discriminant
<b>smoothness_worst</b>	Discriminant ok, maj entre 0.12 et 0.14	Aucun discriminant	Aucun discriminant	Mauvais discriminant
<b>compactness_worst</b>	Discriminant ok, grosse majorite a 0.2, le reste disperse au dessus	Mauvais discriminant	Discriminant ok, les "M" se situent au dessus de 0.25, "B" en dessous.	Mauvais discriminant
<b>concavity_worst</b>	Aucun discriminant	Aucun discriminant	Discriminant ok, les "M" se situent au dessus de 0.3, "B" en dessous de 0.2 mais les valeurs min et max se chevauchent.	Mauvais discriminant
<b>concave_points_worst</b>	Aucun discriminant	Aucun discriminant	Bon discriminant, moyennes bien distinctes, "M" au dessus de 0.15 et "B" en dessous de 0.10	Bon discriminant
<b>symmetry_worst</b>	Bon discriminant, mah a 0.3, le reste au dessus	ok	Aucun discriminant	Mauvais discriminant
<b>fractal_dimension_worst</b>	Mauvais discriminant	ok	Aucun discriminant	Mauvais discriminant

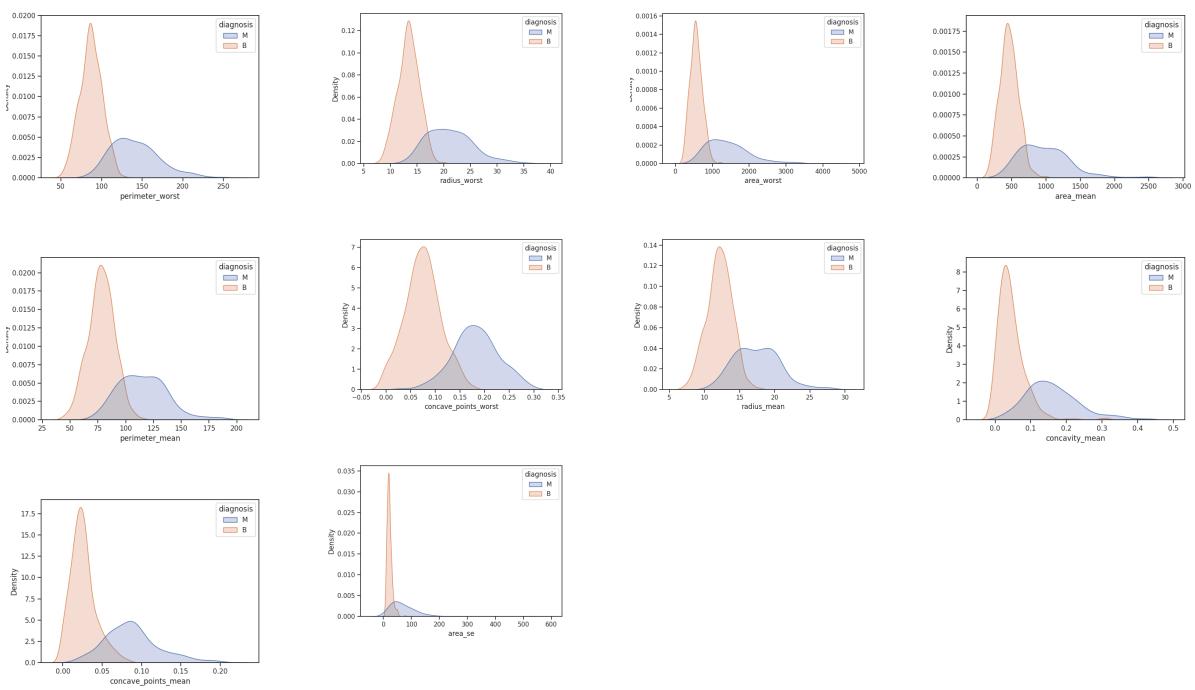
## KDE:

### Variables non discriminantes:





## Variables discriminantes



# Analyse des graphics par état pour les variables discriminantes:

État \ Variable	perimeter_worst	radius_worst	area_worst	area_mean	perimeter_mean	concave_points_worst	radius_mean
Malin	85 - 250 M: 120 - 160	13 - 35 M: 17 - 24	500 - 4500 M: 1000 - 1700	400 - 2500 M: 700 - 1250	75 - 190 M: 100 - 130	0 - 1.2 M: 0.35 - 0.6	11 M
Benin	50 - 135 M: 75 - 100	8 - 21 M: 12 - 15	250 - 1250 M: 450 - 600	200 - 1000 M: 400 - 550	45 - 120 M: 70 - 90	0 - 0.9 M: 0.1 - 0.25	2 M
Écart normalisé*	0.2	0.24	0.17	0.26	0.29	0.5	0.

\*L'écart normalisé est une différence normalisé par l'entendue

## Matrice de corrélation:

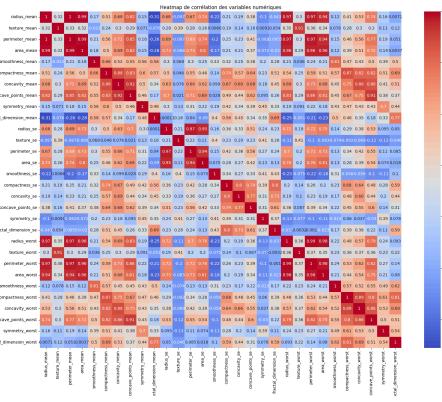
### Observations:

Les variables `area`, `radius` et `perimeter` semblent particulièrement corrélées.

Les variables `concavity_mean` et `concave_points_mean` sont corrélées.

### Radius\_mean:

- `perimeter_mean = 1`
- `area_mean = 0.99`
- `radius_worst = 0.97`
- `area_worst = 0.94`



### Perimeter\_worst:

- `radius_mean = 0.97`
- `perimeter_mean = 0.97`
- `area_mean = 0.96`
- `radius_worst = 0.99`
- `area_worst = 0.98`

### Area\_worst:

- `area_mean = 0.96`
- `radius_worst = 0.98`

### Concave\_point\_worst:

- `concave_point_mean = 0.91`

### Concavity\_mean:

- `concave_point_mean = 0.92`

## Première conclusion:

Après analyse des données je peux conclure que les variables les plus propices à la détection d'une tumeur maligne sont:

- Périmètre
- Radius
- Area
- Convavity

Néanmoins les trois premières étant fortement corrélées il n'est pas nécessaire de les inclure dans l'algorithme.

Aux vues des écarts normalisés il serait judicieux de garder comme variable discriminante `area_worst` ou `area_se`.