

# **Introduction à la Régression Logistique**

# Contexte

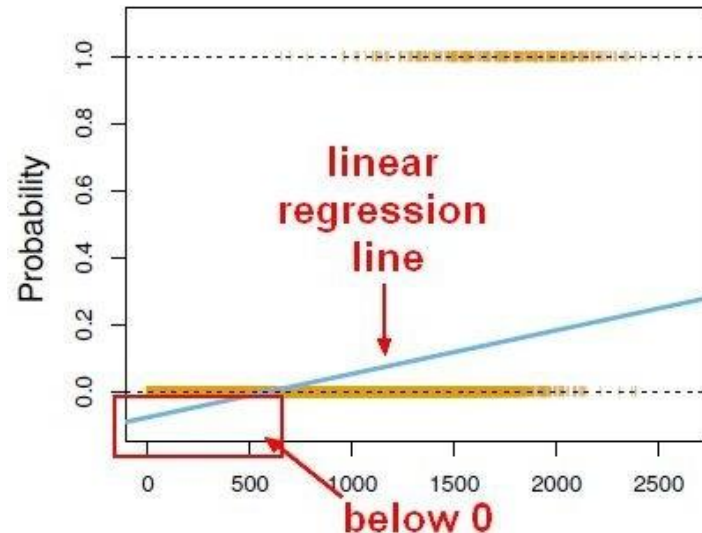
- Nous voulons en apprendre davantage sur la régression logistique comme méthode pour **Classification**.
- Quelques exemples de problèmes de classification :  
Courriels de spam ou « jambon »  
Défaut de prêt (oui/non)  
Diagnostic de la maladie
- Voici tous les exemples de classification binaire

# Contexte

- Jusqu'à présent, nous n'avons vu que des problèmes de régression où nous essayons de prédire une valeur continue.
- Bien que le nom puisse être déroutant au début, la régression logistique nous permet de résoudre des problèmes de classification, où nous essayons de prédire des catégories discrètes.
- La convention pour la classification binaire est d'avoir deux classes 0 et 1.

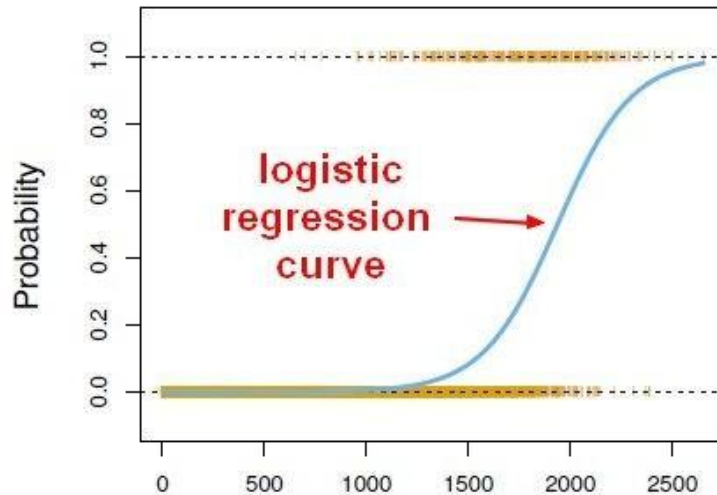
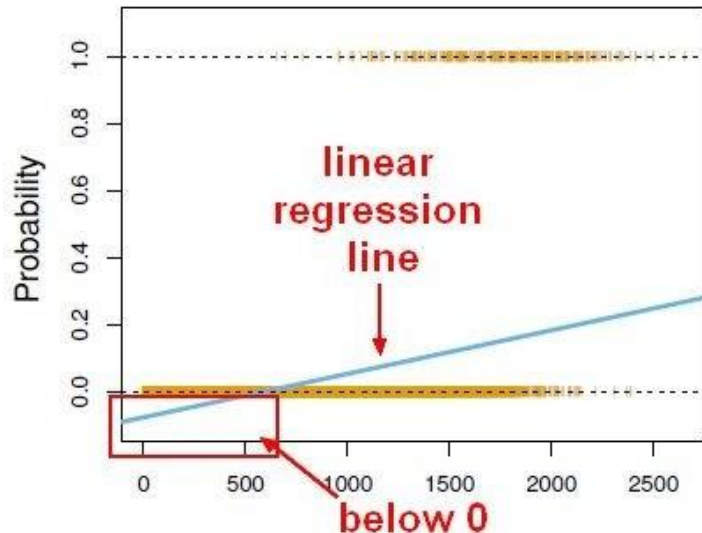
# Contexte

- Nous ne pouvons pas utiliser un modèle de régression linéaire normal sur des groupes binaires. Il ne mènera pas à une bonne correspondance :



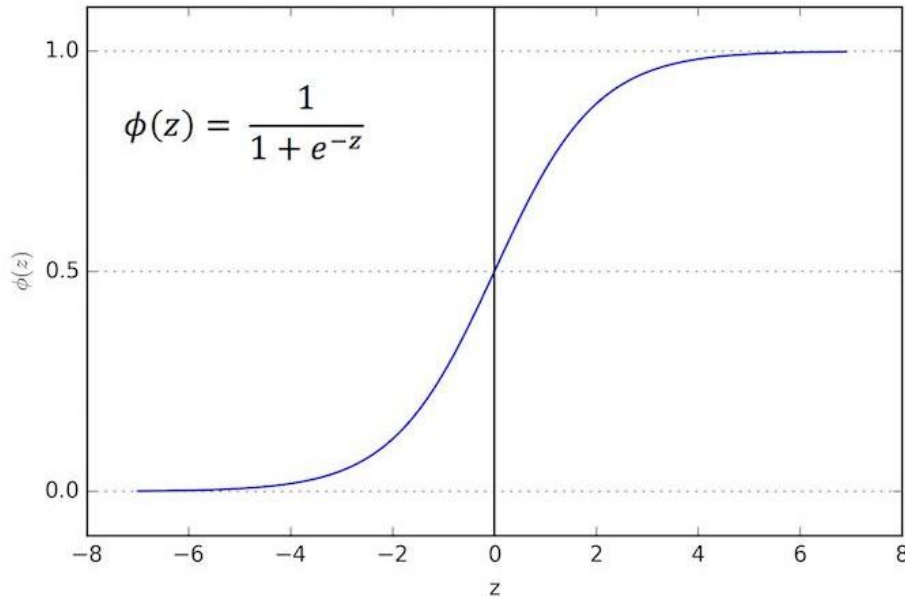
# Contexte

- Nous pouvons plutôt transformer notre régression linéaire en une courbe de régression logistique.



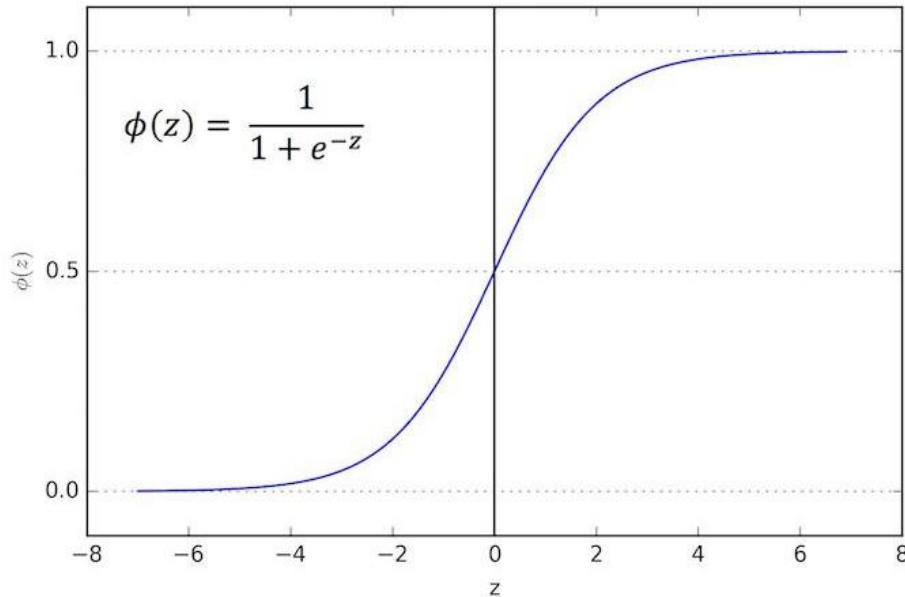
# Fonction Sigmoid

- La fonction sigmoïde (ou logistique) prend n'importe quelle valeur et la renvoie entre 0 et 1.



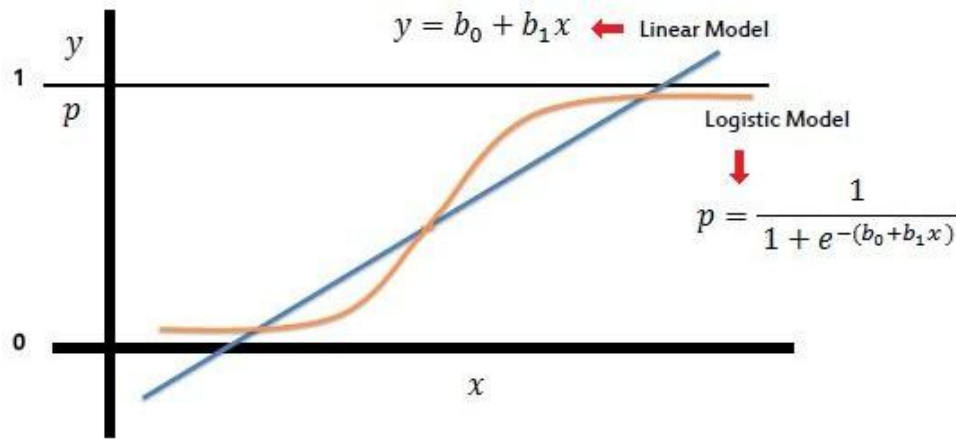
# Fonction Sigmoid

- Cela signifie que nous pouvons prendre notre solution de régression linéaire et la placer dans la fonction sigmoïde.



# Fonction Sigmoide

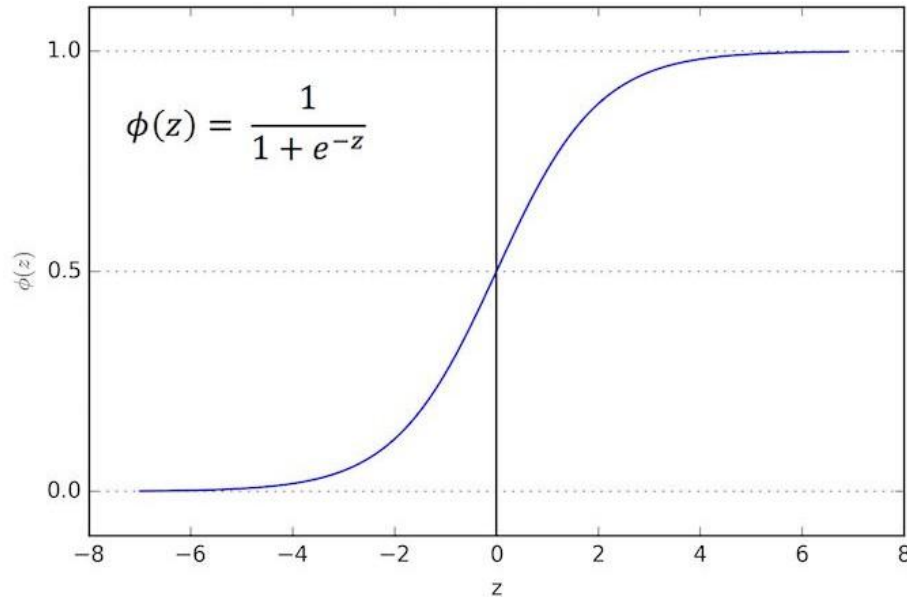
- Cela signifie que nous pouvons prendre notre solution de régression linéaire et la placer dans la fonction sigmoïde.





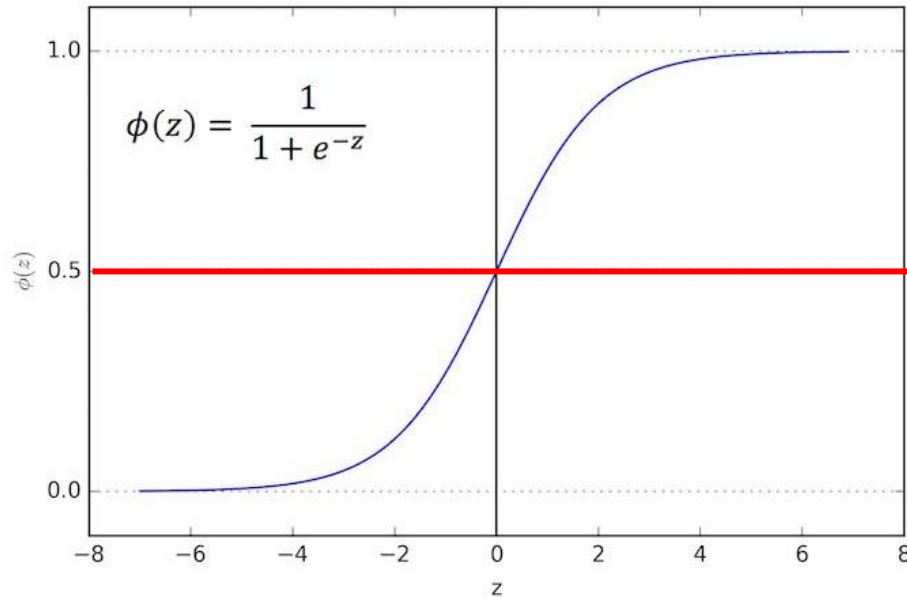
# Fonction Sigmoidé

- Cela donne une probabilité de 0 à 1 d'appartenir à la classe 1.



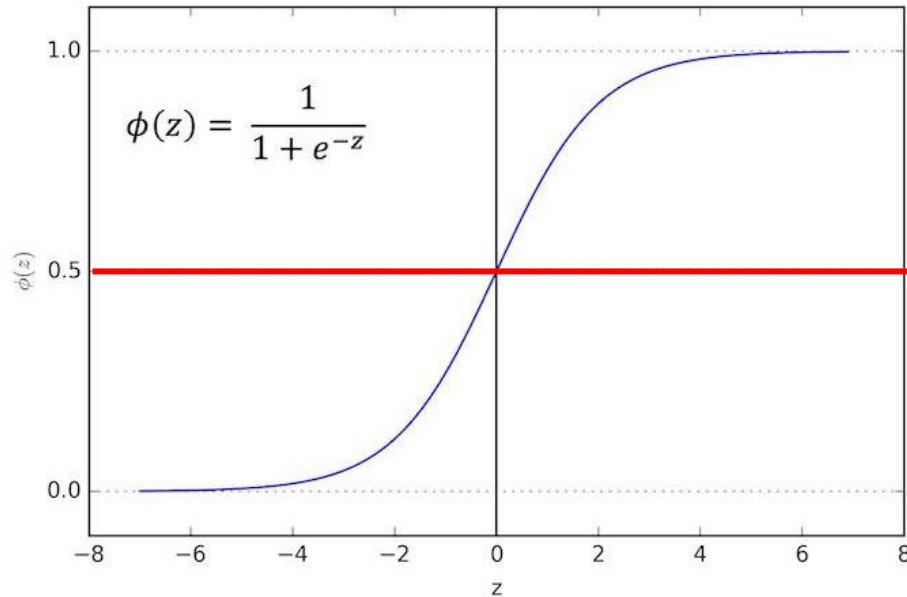
# Fonction Sigmoid

- Nous pouvons fixer un point de coupure à 0,5, tout ce qui est en dessous de ce seuil correspond à la classe 0, tout ce qui est au-dessus de cette classe 1.



# Revue

- Nous utilisons la fonction logistique pour produire une valeur allant de 0 à 1. En fonction de cette probabilité, nous assignons une classe.



# Evaluation du Modèle

- Après avoir entraîné un modèle de régression logistique sur certaines données d'entraînement, vous évaluerez les performances de votre modèle sur certaines données de test.
- Vous pouvez utiliser une matrice de confusion pour évaluer les modèles de classification.

# Evaluation du Modèle

- Nous pouvons utiliser une matrice de confusion pour évaluer notre modèle.
- Par exemple, imaginez faire des tests de dépistage.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Exemple : Test de présence de la maladie  
NO = negative test = False = 0  
YES = positive test = True = 1

# Matrice de Confusion

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Terminologie de base :

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

# Matrice de Confusion

n=165		Predicted: NO	Predicted: YES	
Actual: NO		TN = 50	FP = 10	60
Actual: YES		FN = 5	TP = 100	105
		55	110	

Accuracy:

- Dans l'ensemble, à quelle fréquence est-il **correct**?
- $(TP + TN) / \text{total} = 150/165 = 0.91$

# Matrice de Confusion

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

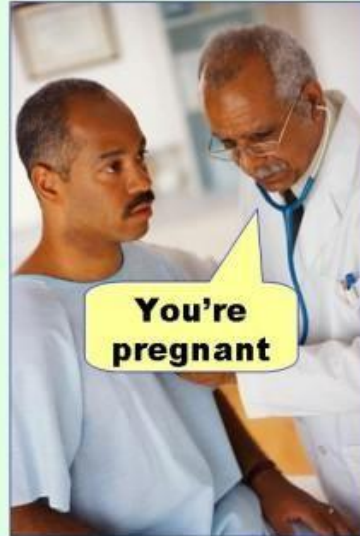
Misclassification Rate  
(Taux d'erreur):

- Dans l'ensemble, à quelle fréquence incorrecte?
- $(FP + FN) / \text{total} = 15/165 = 0.09$



# Matrice de Confusion

**Type I error**  
(false positive)



**Type II error**  
(false negative)

