TP4 : Classification binaire avec SVM, Mélange de modèles, et Modèles Probabilistes Mixtes

Exercice 1 :

Question 2

La stratification est importante ici car, elle permet dans notre cas lorsqu’il y a plus de ham que de spam. Sans cela il y aura un déséquilibre entre les 2 classes ce qui fausserait l’entrainement.

Question 4 : La courbe ROC montre le compromis entre le taux de vrais positifs TPR et le taux de faux positifs FPR

|  |
| --- |
|  |

Question 5 : le score AUC résume la performance du modele, par exemple un score AUC proche de 1 est un bon modele, un score proche de 0.5 est un modele aléatoire.

|  |
| --- |
|  |

Exercice 2 :

Question 1 :

|  |
| --- |
|  |

Question 2 :

|  |
| --- |
|  |

Question 3 :

Dans notre cas le vote soft serait le plus performant. Comme on peut le voir grâce au score AUC du vote soft qui est plus eleve que celui du vote hard. On l’observe également sur les courbes ROC ou le coin supérieur gauche du modele vote soft est plus haute (AUC plus élevé). On observe également sur la matrix de confussion un meilleur score de recall et F1-score du modele soft vote

Question 4 :

|  |
| --- |
|  |

Si l’on classe les modeles grâces au graphique ROC-AUC nous avons ce classement :

Question 5 :

Théoriquement si nous utilisons plusieurs modèles pour une même tache comme ici la détection de spam et ham. Chaque modèle peut donner sa vision et donc finalement avoir une vue plus complète des données. Chaque modèle à sa force. Par exemple comme vue dans les tp précédent le modèles naïve bayes et plus performant pour les textes courts. Et la régression logistique elle est bien calibrée pour les probabilités.

Exercice 3 :

Question 1 :

Le GMM fonctionne de manière avec plusieurs clusters ou chaque groupe suit une distribution gaussienne. Par exemple un GMM attribue à chaque point de 50% d’etre un spam et 50% d’etre un ham. Elle permet également de travailler sur des données complexes. Comme des données avec du bruit ou dans notre cas ici la classification ham/spam car un spam et un ham sont assez similaire, mais également l’analyse d’image.

Question 4 :

|  |
| --- |
|  |

Question 5 :

|  |
| --- |
|  |

Exercice 4 :

Question 1 :

Question 4 :

Question 5 :

Exercice 5 :

Question 1 :

Question 2 :

Question 3 :

Question 4 :

Question 5 :