CS 534: SPRING 2018

Final Exam

This exam has 7 questions for a total of 130 points.

Name :

|  |  |  |
| --- | --- | --- |
| Question | Points | Score |
| 1 | 30 |  |
| 2 | 10 |  |
| 3 | 20 |  |
| 4 | 15 |  |
| 5 | 20 |  |
| 6 | 20 |  |
| 7 | 15 |  |
| Total: | 130 |  |

Be sure to show all work. Attach extra sheets as necessary. Collaboration is not allowed.

1. (30 points) Implement a code that combines the kernel idea with the logistic regression. Use the IRIS binary case that is not linear separable to test your implementation (Please report for each fold confusion matrix, precision, recall, accuracy, and f1-score):
2. (10 points) Use the PCA implementation to visualize at least four different kind of non-linear spaces obtained by the IRIS binary case that is not linear separable.
3. (20 point) write a code allowing you to decide the best hyper parameter for the K-means. Use iris data (4 features) and iris data with the polynomial features (polynomial degree 2 and 3)
4. (15) Create a procedure that balance by oversampling an unbalanced dataset, duplicates are not allowed in the new dataset. Please use iris binary cases. Please note that one of the clustering methods studied will be required.
5. (20 points) Explain the main differences between K-means, K-medoids, DBScan and Gaussian Mixture Models (examples are welcome).
6. (20 points) Compute the frequent itemsets (with threshold 0.5) of the following transactional dataset and the relative association rules (confidence 0.6).

|  |
| --- |
| Transactions |
| a,b,c,d |
| a,b,d,f |
| b,c,d |
| b,d f |
| a,b,c |
| a,b,c,f |
| a,b |

1. (15 points) answer the following questions:
   1. How you can use association rules for classification?
   2. What can be one of the advantages to use the association rules for classification?
   3. Do you know some other model that can produce rules for the classification (explain what is, and what should be the procedure to extract the rules)
   4. Explain why bias is needed to learn something.