1. Construct the smallest dataset (set of transactions) of 4 \((a,b,c,d)\) items that will give rise to the largest FP tree assuming a minsupport 1 and item ordering \(a,b,c,d\).

2. What is the maximum number of itemsets in the candidate set of Apriori for the above dataset.

3. For a dataset of \(n\) itemsets what is the worst case space requirement in terms of number of nodes.
4. Write the expression for the probability of seeing sequence “CAACCA” for the variable memory Markov model below.

5. For the training sequences below, write down the transition probabilities for the Markov Model below:

<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCAACC</td>
</tr>
<tr>
<td>ACCCACAC</td>
</tr>
<tr>
<td>CCAC</td>
</tr>
<tr>
<td>CACACA</td>
</tr>
</tbody>
</table>

6. In the model above, what is the AA to C and AA to A transition after applying Laplace correction.
7. Suppose we write the constraints for a separable SVM as $y_i(\mathbf{w}.\mathbf{x}_i + b) \geq \gamma$ instead of the usual $y_i(\mathbf{w}.\mathbf{x}_i + b) \geq 1$. Derive $c$ as a function of $\gamma$ such that the margin size will then be $\frac{c}{||\mathbf{w}||^2}$ instead of $\frac{2}{||\mathbf{w}||^2}$?

8. Suggest a modification to the SVM objective function and/or constraints to handle weighted instances where each training instance $i$ is attached with a scalar weight $u_i$ denoting the importance of the instance.

9. If you find that SVM with a RBF kernel is over-fitting your data. State with a brief reason
which of the following actions you will take:

(a) Increase or decrease $C$?

(b) Increase or decrease $\sigma$ of the RBF kernel $e^{-\frac{(x_i-x_j)^2}{\sigma^2}}$?

10. Sketch two distributions of points in 2D, one where single-link clustering is more suitable than complete-link and second where complete-link is better.
11. Suppose you are asked to build a classifier for a dataset where the number of classes is very large, say order of tens of thousands. Which of nearest neighbor, decision trees and generative models like naive Bayes do you expect to perform best in this scenario. Why?

12. In locally weighted regression, the regression function \( y = f_q(x) = w_q x + b_q \) is learnt after seeing the instance \( x_q \) so that training instances near \( x_q \) are weighted higher. Assume \( d(x, x_q) \) denotes the distance between \( x \) and \( x_q \).

- Write an expression for the square error objective function that can be minimized to learn \( f_q \).

- Draw an example of points with one dimensional data and continuous \( y \), where locally weighted regression would perform much better than linear regression.
• State two limitations of locally weighted regression compared to linear regression.

13. Consider a modification of the k-means algorithm where instead of a single representative (the centroid) we maintain some $c$ representatives (reps) per cluster. A point is assigned to the cluster where the total distance to all the reps is minimum.

• Sketch a procedure for selecting the $c$ reps out of the points assigned to a cluster at the end of each pass of the data.
• Why is using more than one rep a good idea?

• What is the running time complexity of the enhanced k-means assuming \( I \) iterations and \( T \) training instances?

14. Suppose you are asked to create a classifier for a dataset with two classes where each training instance is either marked with a positive label or not marked, which means that it can be either positive or negative. Design a training algorithm for classifying in these circumstances using a baseline SVM classifier.
Total: 45