School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 1, 2017)
Workshop exercises: Week 12

1. Revise **Support Vector Machines**, paying particular attention to the terms "linear separability" and "maximum margin".

   (a) What is the significance of allowing "some margin of errors", indicated by ξ in the lectures?

   (b) Why are we interested in "kernel functions" here?

   (c) Why are SVMs "binary classifiers", and how can we extend them to "multi-class classifiers"?

2. What is **Clustering**?

   (a) What is the difference between "partitional" and "hierarchical" clustering? What are some other distinctions that we can draw between clusterings?

   (b) How does the k-**means** algorithm cluster data? Given the following dataset:

id	*apple*	*ibm*	*lemon*	*sun*
`A`	4	0	1	1
`B`	5	0	5	2
`C`	2	5	0	0
`D`	1	2	1	7
`E`	2	0	3	1
`F`	1	0	1	0

   Apply k-means, using the Manhattan distance, and seeds `A` and `D`. What would happen if we had used different instances as seeds?

3. For the following set of instances:

a_1	a_2	a_3	c
hot	windy	dry	Yes
mild	windy	rainy	No
hot	windy	rainy	Yes
cool	still	dry	Yes
cool	still	rainy	No
hot	still	dry	No
mild	still	dry	Yes

   (a) Calculate the **confidence** and **support** of the Association Rule {still, Yes}→{dry}.

   (b) Discuss how you would continue mining for effective **Association Rules**, according to some thresholds τ_c for confidence and τ_s for support.