Support Vector Machines

1. How is the “margin” calculated in a *maximal margin classifier*?
2. What is a support vector?
3. In your own words, what is the difference between a “soft margin” and a “hard margin”?
4. In your own words, what is the difference between a *Maximal Margin Classifier* and a *Support Vector Classifier*?
5. In the context of support vector machines, what is a *kernel*?
6. We are given observatons in dimensions. For each observation, there is an associated class label:

| Observation |  |  |  |
| --- | --- | --- | --- |
| 1 | 3 | 4 | red |
| 2 | 2 | 2 | red |
| 3 | 4 | 4 | red |
| 4 | 1 | 4 | red |
| 5 | 2 | 1 | blue |
| 6 | 4 | 3 | blue |
| 7 | 4 | 1 | blue |

1. Sketch the observations.
2. Sketch the optimal separating hyperplane.
3. On your sketch, indicate the margin for the maximal margin classifier.
4. Describe the classification rule for the maximal margin classifier.
5. Indicate the support vectors for the maximal margin classifier.
6. Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.
7. Sketch a hyperplane that **is not** the optimal separating hyperplane.
8. Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.
9. Suggest a kernel that might work well with a support vector machine for this data, and explain why you think it would work well.