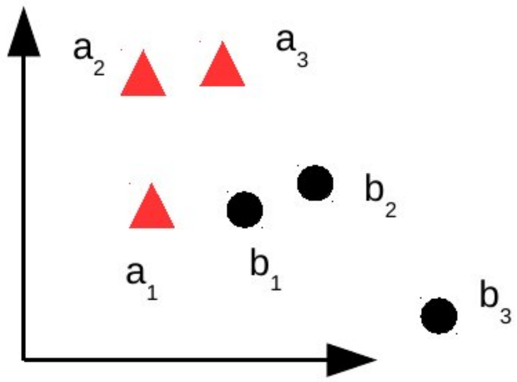
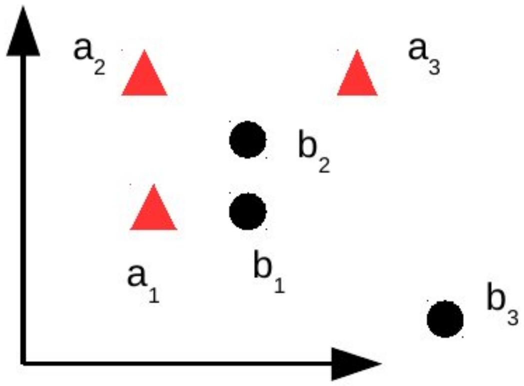
Support Vector Machines

Questions

1. Explain why minimizing the norm of the vector of weights maximizes the margin.
2. What do the constraints in the optimization problem represent?
3. What is the role of slack variables? What do they achieve?
4. What is the difference between a soft-margin and a hard-margin svm?
5. Given the dataset below, determine if a hard-margin SVM can separate the classes, and, if that is the case, identify the support vectors:



6. Same as the question before:



7. What options do you have with support vector machines if the dataset is not linearly separable?
8. What is the kernel trick and what does it achieve?
9. Why are kernels useful?
10. Consider the dataset: $\{\langle 0,0,-1\rangle,\langle 0,1,1\rangle,\langle 1,0,1\rangle\}$ where the last element of each vector is the class $t\in\{-1,1\}$. We want to use a linear (non-kernel) support vector machine

classifier to specify the decision boundary in the form of $y(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0$. Let a_1, a_2, a_3 denote the Lagrange multipliers for the constraints on x_1, x_2, and x_3 respectively.

1. Plot the data points and derive the decision boundary by inspecting the data. What can be said about the Lagrange multipliers?
2. Write the Lagrangian, apply the optimality conditions, and express the vector **w** in terms of the data points.

11. Consider the dataset: $\{\langle -1,0,-1\rangle, \langle 1,0,1\rangle, \langle 2,0,1\rangle\}$. Compute the value of the three Lagrange multipliers corresponding to each point.
12. Consider the dataset: $\{\langle -1,0,-1\rangle, \langle 1,0,1\rangle, \langle 2,0,-1\rangle\}$, that is, the same as the previous question, with class of the last point inverted. This dataset is not linearly separable, therefore we need to introduce slack variables. Assuming that the decision boundary is still $x=0$, what is the value of the slack variables ξ_1, ξ_2, ξ_3 , corresponding to each point?