

Assignment 5

Remarks:

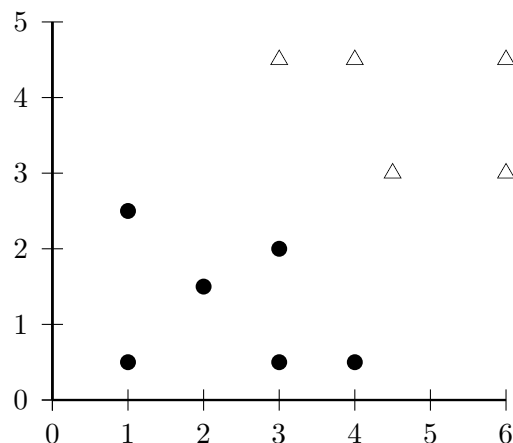
1. Please submit your work until **December, 17 12:00** on Ilias.
2. If you need any **help**, send an email to `max.kisselew@ims.uni-stuttgart.de` or drop in at my office: 1.013 (Pfaffenwaldring 5b).

Exercise 1 (IIR 15) [4 P.]

As you know from class, a Support Vector Machine (SVM) is estimated by finding the smallest vector \vec{w} so that

$$\text{sign}(y_i(\vec{w}^T x_i + b)) \geq 1. \quad (1)$$

Estimate a SVM for the data given below. Find the support vectors and the general form of the normal vector before solving the equation system 1 for the best matching \vec{w} . The Chapter 15.1 of the IR book may help you with this exercise.

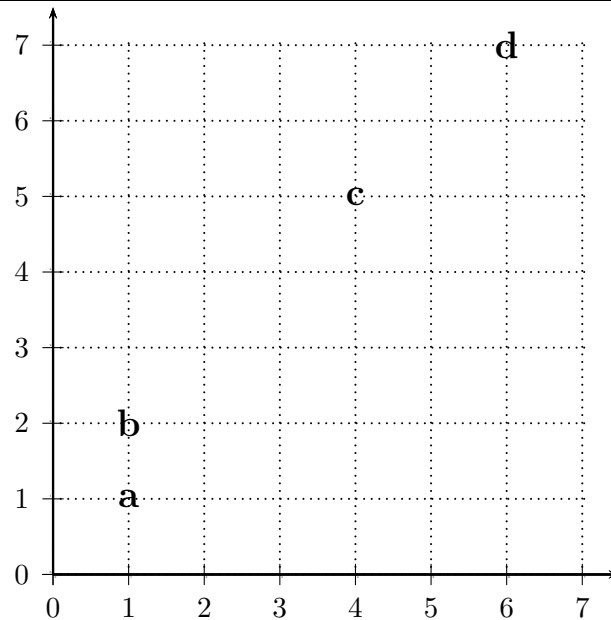


Exercise 2 (IIR 16) [4 P.]

- (i) Perform a 2-means clustering to convergence for the points below. Start with the two seeds a and b. For each iteration give (I) the coordinates of the centroids (II) the assignments of points to centroids.
- (ii) Give the coordinates of a fifth point e and two centroids with the following properties:

1. the two centroids are a local optimum; that is, one iteration of reassignment and recomputation will not change the position of the centroids
2. the two centroids are not the global optimum.

- (iii) Give two centroids that are better for the 5 points than the ones in (ii). (No need to prove global optimality, but show they are better than in (ii).)

**Exercise 3 (IIR 16) [1 P.]**

Why are documents that do not use the same term for the concept *car* likely to end up in the same cluster in *K-means* clustering?

Exercise 4 (IIR 16) [1 P.]

Two of the possible termination conditions for K-means were (1) assignment does not change, (2) centroids do not change (IR book, page 361). Do these two conditions imply each other? Why or why not?

Due date: Monday, December 17, 2012, 12:00