Prof. Dr. Jan Rolfes, Ehsan Waiezi, Lars Weidne

Mathematics of Learning – Worksheet 8

- The exercise sheets will be uploaded every Monday. Solution sketches will be uploaded one week later.
- You can hand in your own solutions via StudOn and we correct them - this is not mandatory. Please hand in small groups of 2-3 students.
- For questions, please use the forum on StudOn since other students may have similar questions. If you have a more personal question about the exercises please send an email to ehsan.waiezi@fau.de or lars.weidner@fau.de respectively.

Basics [Hyperplanes].

We want to consider a few properties of sets of the kind

$$H := \{x \in \mathbb{R}^p : a^T x = b\}$$

for vectors $a \in \mathbb{R}^p$ and vectors $b \in \mathbb{R}$. These sets are called *hyperplanes*.
a) Prove: H is empty, if and only if $a = 0^p$ and $b \neq 0$.
b) Prove: H is convex, i.e., if $x, y \in H$, then for any real number $\lambda \in [0, 1]$, also $\lambda x + (1 - \lambda)y \in H$.
c) Prove: H is affine, i.e., if $x, y \in H$, then for any real numbers $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 + \lambda_2 = 1$, also $\lambda_1 x + \lambda_2 y \in H$.

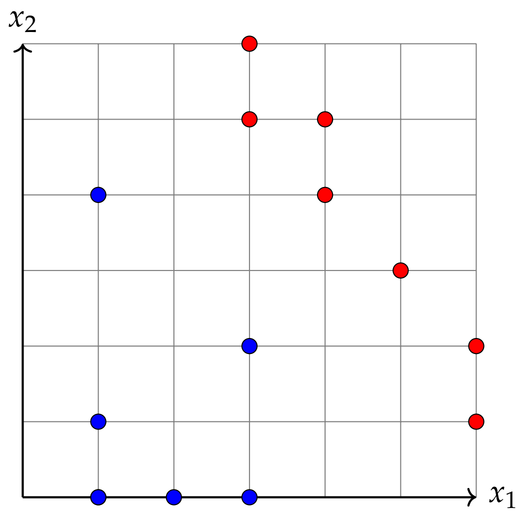
If we make a slight adaptations to the sets under consideration, i.e., we change $=$ by $\leq$, we get so-called halfspaces:

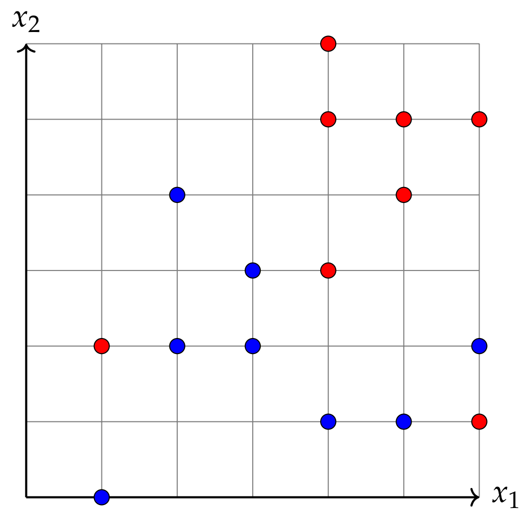
$$S := \{x \in \mathbb{R}^p : a^T x \leq b\}$$

d) Prove: Halfspaces are convex, but (in general) not affine.

Exercise 1 [Find vectors for 2d point classification].

Consider these two pictures. Find lines which separate the two classes (red points and blue points) as good as possible for every figure. You are allowed to be creative, but please explain your understanding of “as good as possible”. Write down the definition of your lines formally correct as hyperplanes.





Exercise 2 [The path to support vector machines].

Let $\{(x_1, y_1), ..., (x_n, y_n)\} \subset \mathbb{R}^p \times \{-1, 1\}$ be n pairs of labeled data (we exclude the border case where only one label exists since it is not interesting) within $\mathbb{R}^p$. Consider the following two optimization problems:

$$\begin{aligned} \max_{\beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}, M \in \mathbb{R}} & \; M \\ \text{s.t. } y_i(x_i^T \beta + \beta_0) \geq M & \quad \text{for all } i = 1, ..., n \\ ||\beta|| = 1 & \end{aligned}$$

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}} & \; ||\beta||^{(2)} \\ \text{s.t. } y_i(x_i^T \beta + \beta_0) \geq 1 & \quad \text{for all } i = 1, ..., n \end{aligned}$$

a) Try to describe in words what the optimization problems do.

b) Prove or disprove: Whenever $\hat{\beta}, \hat{\beta}_0$ solves the right problem, then $\frac{\hat{\beta}}{||\hat{\beta}||}, \frac{\hat{\beta}_0}{||\hat{\beta}||}$ solves the left problem.

c) Interpret the points of the previous task as labeled input data ($p = 2$, $n = 13/16$, red is 1 and blue is -1). Try to solve the left as well as the right optimization problem for both data sets (*Remark: This looks quite simple, but it is not, since you have to solve a constrained quadratic optimization problem. There are two ways to go which I can see: 1. set up the KKT-System and try to solve it explicitly, maybe with the help of some nonlinear equation system solver, 2. Try to get in touch with optimization software: search e.g., for pyomo, neos server. Try to do either 1 or 2.*).