Lecturer: Matthias Seeger
Assistants: Abdallah Elguindy, Mathias Humbert, Young Jun Ko, Xinchao Wang

Solutions to Exercise Sheet 1

Exercise 1

1.1

Substituting both points in the equation:

$$\begin{cases} 8 = 5a + b \\ 0 = 1a + b \end{cases} \implies \begin{cases} a = 2 \\ b = -2 \end{cases} \tag{1}$$

we get $x_2 = 2x_1 - 2$. Substituting the point $[0,0]^T$, we check that the line does not cross the origin.

1.2

We rewrite the line equation as $x_2 - 2x_1 + 2 = 0$ and get $w_1 = -2, w_2 = 1, b = 2$.
We have $\boldsymbol{w}^T \boldsymbol{p}_2 + b = 0$ and $\boldsymbol{w}^T \boldsymbol{p}_1 + b = 0$. Subtracting the equations, one gets $\boldsymbol{w}^T(\boldsymbol{p}_2 - \boldsymbol{p}_1) = 0$, concluding that $\boldsymbol{w}$ and $\boldsymbol{p}_2 - \boldsymbol{p}_1$ are orthogonal.

1.3

We can define $f(x)$ as

$$f(\boldsymbol{x}) = \operatorname{sgn}(y(\boldsymbol{x})) = \begin{cases} +1 \mid x_2 - 2x_1 + 2 > 0 \\ -1 \mid x_2 - 2x_1 + 2 < 0 \end{cases} \tag{2}$$

$f(x)$ remains undefined for $x_2 - 2x_1 + 2 = 0$, even though a frequent choice in practice is to sample a label at random there.
From the definition, we calculate $f([5,9]^T) = +1$, $f([10,18]^T)$ undefined (on the boundary), $f([0,-3]^T) = -1$.

1.4 Bonus

There are many possible solutions here. To check whether your solution is correct, you compute the orthogonal projection $\hat{\boldsymbol{p}}_j$ of $\boldsymbol{p}_j$ onto your line $\tilde{\boldsymbol{w}}^T \boldsymbol{x} + \tilde{b}$, $j = 1,2$, then check whether each $\|\boldsymbol{p}_j - \hat{\boldsymbol{p}}_j\| = 1$.
Here is one simple solution (to save time, you should always choose the simplest solution!). Our original equation is $\boldsymbol{w}^T \boldsymbol{x} + b = 0$, and $\boldsymbol{w}^T \boldsymbol{p}_j + b = 0$ for $j = 1,2$ (both points lie on the line). As a solution, we can pick a line *parallel* to this one, which means that $\tilde{\boldsymbol{w}} = \boldsymbol{w}$, but $\tilde{b} \neq b$. We determine $\tilde{b}$ by constructing one point on the new line. Since $\boldsymbol{w}$ is

orthogonal to the original line, we obtain this new point by starting at $\boldsymbol{p}_1$ and marching along $\boldsymbol{w}$ for distance 1:

$$\tilde{\boldsymbol{p}}_1 = \boldsymbol{p}_1 + \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}.$$

Now, since $\tilde{\boldsymbol{w}} = \boldsymbol{w}$:

$$0 = \boldsymbol{w}^T \tilde{\boldsymbol{p}}_1 + \tilde{b} = -b + \frac{\|\boldsymbol{w}\|^2}{\|\boldsymbol{w}\|} + \tilde{b} \quad \Rightarrow \quad \tilde{b} = b - \|\boldsymbol{w}\| = 2 - \sqrt{5}.$$

Note: It is *absolutely OK* to state a final solution as "$2 - \sqrt{5}$", you do not have to compute a decimal approximation (you should know that this number is negative).

Exercise 2

2.1

We use discriminants $y(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x}$, where $\boldsymbol{x} = [x, y]^T$, $\boldsymbol{w} = [w_1, w_2]^T$.
Choosing the points in the given table order, learning rate of 1 and defining class 1 as $t = +1$ and class 2 as $t = -1$, we have:

- $f(\boldsymbol{x}_1) = [1, -1][4, 3]^T = 1 > 0$, ok!
- $f(\boldsymbol{x}_2) = [1, -1][3, 4]^T = -1 < 0$, not ok, so update the weights:
  $\boldsymbol{w} \leftarrow \boldsymbol{w} + t_2 * \boldsymbol{x}_2 = [1, -1]^T + (+1) * [3, 4]^T = [4, 3]^T$.
- $f(\boldsymbol{x}_3) = [4, 3][2, 3]^T = 17 > 0$, ok!
- $f(\boldsymbol{x}_4) = [4, 3][-2, -3]^T = -17 < 0$, ok!
- $f(\boldsymbol{x}_5) = [4, 3][-4, -2]^T = -22 < 0$, ok!
- $f(\boldsymbol{x}_6) = [4, 3][-4, -3]^T = -25 < 0$, ok!
- $f(\boldsymbol{x}_7) = [4, 3][-2, 5]^T = 7 > 0$, not ok, so update the weights:
  $\boldsymbol{w} \leftarrow \boldsymbol{w} + t_7 * \boldsymbol{x}_7 = [4, 3]^T + (-1) * [-2, 5]^T = [6, -2]^T$.

Now it has converged!
Hint for solving this exercise quickly. After each update, draw the new discriminant into the figure (or a copy). This way, you can immediately spot where further errors occur, without having to do lots of math.

2.2

For the order chosen here, it would have converged already at the first update.

2.3

By visual inspection, it is clear that the data would not be linearly separable. The algorithm would run forever.

Exercise 3

3.1

In the standard form: $y(\boldsymbol{x}) = [\frac{1}{4}, \frac{1}{4}, -\frac{1}{2}, -\frac{3}{2}][x_1^2, x_2^2, x_1, x_2]^T + \frac{3}{2}$, where $\boldsymbol{w} = [\frac{1}{4}, \frac{1}{4}, -\frac{1}{2}, -\frac{3}{2}]^T$, $\boldsymbol{\phi}(\boldsymbol{x}) = [x_1^2, x_2^2, x_1, x_2]^T$ and $b = \frac{3}{2}$.
Note: As some of you pointed out, this exercise was not clearly formulated: we did not properly define "standard form". In such a situation, *any* solution you give, which reproduces $y(\boldsymbol{x})$, is correct.

3.2

Completing the squares, we rewrite $y(\boldsymbol{x})$ as

$$y(\boldsymbol{x}) = \frac{1}{4}\left((x_1 - 1)^2 + (x_2 - 3)^2 - 4\right) = \frac{1}{4}\|\boldsymbol{x} - [1, 3]^T\|^2 - 1 \tag{3}$$

The boundary is at $y(\boldsymbol{x}) = 0 \implies \|\boldsymbol{x} - [1, 3]^T\| = 2$, which represents the circle of radius 2 around the point $[1, 3]$, separating points inside and outside this circle. More precisely, the interior of the circle is $\mathcal{H}_{-1}$, everything outside the circle is $\mathcal{H}_{+1}$ (the complement of the closed circle), and the decision boundary is the margin of the circle.