Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical
Information Sciences

Exam on Neural Network Theory
February 6, 2020

Please note:

- Exam duration: 180 minutes
- Maximum number of points: 100
- You are not allowed to use any printed or handwritten material (i.e., books, lecture and discussion session notes, summaries), computers, tablets, smart phones or other electronic devices.
- Your solutions should be explained in detail and your handwriting needs to be clean and readable.
- Please do not use red or green pens. You may use pencils.
- Please note that the ETHZ "Disziplinarordnung RSETHZ 361.1" applies.

Before you start:

1. The problem statements consist of 5 pages including this page. Please verify that you have received all 5 pages.
2. Please fill in your name and your Legi-number below.
3. Please place an identification document on your desk so we can verify your identity.

During the exam:

4. For your solutions, please use only the empty sheets provided by us. Should you need more paper, please let us know.
5. Each problem consists of several subproblems. If you do not provide a solution to a subproblem, you may nonetheless assume its conclusion in the ensuing subproblems.

After the exam:

6. Please write your name on all the solution sheets. All sheets, including those containing problem statements, must be handed in. Please sign this page.

Family name: .... First name: ....

Legi-No.: ....

Signature: ....

Problem 1

For every vector $\boldsymbol{x} \in \mathbb{R}^2$, let x_1 and x_2 denote the first and second component of $\boldsymbol{x}$, respectively. Let $\alpha, \beta \in \mathbb{R}$ with $\alpha, \beta > 0$ be fixed parameters and consider the function $K\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined according to

$$K(\boldsymbol{x}, \boldsymbol{y}) = \alpha x_1^2 y_1^2 + \beta x_2^2 y_2^2.$$

(a) Prove that $K(\boldsymbol{x}, \boldsymbol{y})$ is a symmetric kernel.

(b) Prove that for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ and $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon, \boldsymbol{x}, \boldsymbol{y}) > 0$ depending on ε, $\boldsymbol{x}$, and $\boldsymbol{y}$ such that if $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ satisfy

$$\left\| \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} \right\|_2 < \delta, \tag{1}$$

then $|K(\boldsymbol{x}, \boldsymbol{y}) - K(\boldsymbol{u}, \boldsymbol{v})| < \varepsilon$, i.e., $K(\boldsymbol{x}, \boldsymbol{y})$ is a continuous function.

Hint. *Prove the following substeps:*

Step 1: Show that $\big|K(\boldsymbol{x}, \boldsymbol{y}) - K(\boldsymbol{u}, \boldsymbol{v})\big| \le \alpha\big|x_1^2 y_1^2 - u_1^2 v_1^2\big| + \beta\big|x_2^2 y_2^2 - u_2^2 v_2^2\big|$.

Step 2: Use the fact that (1) *implies*

$$|x_1 - u_1| < \delta, \quad |y_1 - v_1| < \delta, \quad |x_2 - u_2| < \delta, \quad \textit{and} \quad |y_2 - v_2| < \delta$$

to find a constant $C = C(\varepsilon, \boldsymbol{x}, \boldsymbol{y})$ *such that* $\delta < C$ *implies*

$$\big|x_1^2 y_1^2 - u_1^2 v_1^2\big| < \frac{\varepsilon}{2\alpha} \quad \textit{and} \quad \big|x_2^2 y_2^2 - u_2^2 v_2^2\big| < \frac{\varepsilon}{2\beta}.$$

(c) Prove that, for every $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^2$, the $k \times k$ Gramian matrix

$$\boldsymbol{K}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \begin{pmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) & K(\boldsymbol{x}_1, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_1, \boldsymbol{x}_k) \\ K(\boldsymbol{x}_2, \boldsymbol{x}_1) & K(\boldsymbol{x}_2, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_2, \boldsymbol{x}_k) \\ \vdots & \vdots & \dots & \vdots \\ K(\boldsymbol{x}_k, \boldsymbol{x}_1) & K(\boldsymbol{x}_k, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_k, \boldsymbol{x}_k) \end{pmatrix}$$

is positive semidefinite, i.e., $K(\boldsymbol{x}, \boldsymbol{y})$ is a positive semidefinite kernel.

Hint. *Write* $K(\boldsymbol{x}, \boldsymbol{y})$ *in the form* $K(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x})^\mathsf{T}\Phi(\boldsymbol{y})$ *with* $\Phi(\boldsymbol{x}) = (\sqrt{\alpha}x_1^2 \;\; \sqrt{\beta}x_2^2)^\mathsf{T}$ *and show that* $\boldsymbol{c}^\mathsf{T}\boldsymbol{K}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\boldsymbol{c} \ge 0$*, for all* $\boldsymbol{c} \in \mathbb{R}^k$.

(d) It follows from (a)–(c) that $K(\boldsymbol{x}, \boldsymbol{y})$ is a Mercer kernel. Determine the reproducing kernel Hilbert space $\mathcal{H}_K$ corresponding to $K(\boldsymbol{x}, \boldsymbol{y})$.

(e) Construct an orthonormal basis for the reproducing kernel Hilbert space $\mathcal{H}_K$ corresponding to $K(\boldsymbol{x}, \boldsymbol{y})$.

Hint. *Note that* $\mathcal{H}_K$ *is finite-dimensional with corresponding inner product* $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ *satisfying the reproducing property* $\langle K_{\boldsymbol{y}}, f \rangle_{\mathcal{H}_K} = f(\boldsymbol{y})$*, for all* $f \in \mathcal{H}_K$ *with* $K_{\boldsymbol{y}}\colon \mathbb{R}^2 \to \mathbb{R}$ *defined according to* $K_{\boldsymbol{y}}(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{y})$*, for all* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}_K$*. A basis* $\{e_i : i \in \mathcal{I}\} \subseteq \mathcal{H}_K$ *is an orthonormal basis if* $\langle e_i, e_j \rangle_{\mathcal{H}_K} = 0$*, for all* $i, j \in \mathcal{I}$ *with* $i \neq j$ *and* $\langle e_i, e_i \rangle_{\mathcal{H}_K} = 1$*, for all* $i \in \mathcal{I}$.

Problem 2

In this problem, we explicitly solve the support vector machine algorithm to construct the best possible straight line separating two sets of vectors in $\mathbb{R}^2$. Specifically, consider the vectors

$$\boldsymbol{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \boldsymbol{x}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{x}_4 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and let $y_1 = 1$, $y_2 = 1$, $y_3 = -1$, and $y_4 = -1$.

(a) Draw a picture depicting $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4$ and a straight line separating $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ from $\{\boldsymbol{x}_3, \boldsymbol{x}_4\}$.

(b) Consider the optimization problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^2, b \in \mathbb{R}} \tfrac{1}{2}\|\boldsymbol{w}\|_2^2$$
$$\text{such that } y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle_2 - b) \geq 1, \quad \text{for } i = 1, \ldots, 4.$$

Write down the Lagrange function $L(\boldsymbol{w}, b, \boldsymbol{c})$ corresponding to this optimization problem, where $\boldsymbol{c} = (c_1 \ldots c_4)^\top$ is a vector containing the Lagrange multipliers $c_1, \ldots, c_4 \in \mathbb{R}$.

(c) Consider the Lagrange function $L(\boldsymbol{w}, b, \boldsymbol{c})$ from subproblem (b). The corresponding Lagrange dual function has the form

$$g(\boldsymbol{c}) = \boldsymbol{a}^\top \boldsymbol{c} - \frac{1}{2}\boldsymbol{c}^\top \boldsymbol{A}\, \boldsymbol{c},$$

where $\boldsymbol{a} \in \mathbb{R}^4$ and $\boldsymbol{A} \in \mathbb{R}^{4\times 4}$. Compute $\boldsymbol{a}$ and $\boldsymbol{A}$ explicitly.

(d) Consider the Lagrange dual function $g(\boldsymbol{c})$ from subproblem (c). Solve the corresponding Lagrange dual problem

$$\max_{\boldsymbol{c} \in \mathbb{R}^4} g(\boldsymbol{c})$$
$$\text{such that } c_i \geq 0, \quad \text{for } i = 1, \ldots, 4 \text{ and } \sum_{i=1}^{4} c_i y_i = 0$$

and identify the corresponding support vectors in the set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4\}$.

Hint. *The function $f(t) = 2t - t^2/2$ is strictly concave on $(0, \infty)$ and, therefore, maximized at the point satisfying $\mathrm{d}f(t)/\mathrm{d}t = 0$.*

(e) Compute a solution $(\tilde{\boldsymbol{w}}, \tilde{b})$ of the optimization problem in subproblem (b) and write down the expression for the corresponding hard margin binary classifier.

Problem 3

Consider the function

$$f(x) = \begin{cases} 0, & x \leq 0; \\ 4x, & 0 < x < \frac{1}{4}; \\ 1, & \frac{1}{4} \leq x \leq \frac{3}{4}; \\ 4 - 4x, & \frac{3}{4} < x < 1; \\ 0, & x \geq 1. \end{cases}$$

(a) Plot $f(x)$.

(b) Realize $f(x)$ through a depth-2 ReLU network (see the Handout for the definition of a depth-L ReLU network). Specify the width and the connectivity of the resulting network.

(c) Consider the function $h(x) = f(4x) + f(4x - 3)$. Realize $h(x)$ through a depth-2 ReLU network. Specify the width and the connectivity of the resulting network.

(d) In subproblem (c) $h(x) = f(4x) + f(4x - 3)$ was realized by a ReLU network of depth 2. Find a deeper ReLU network that realizes $h(x)$ with a width of 4. Specify the depth and the connectivity of the resulting network.

Hint. *Try to write $h(x)$ as a composition of two functions.*

Problem 4

(a) Let $(\mathcal{X}, \rho)$ be a metric space and $\mathcal{C}$ a compact set in $\mathcal{X}$. State the definitions of the ϵ-covering number $N(\epsilon; \mathcal{C}, \rho)$ and the ϵ-packing number $M(\epsilon; \mathcal{C}, \rho)$.

(b) Let $\mathcal{C}$ be the ℓ_2-ball $\mathcal{C} = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$. By considering the volumes of ℓ_2-balls in $\mathbb{R}^d$, show that for any $\epsilon \leq 2$, $M(\epsilon; \mathcal{C}, \|\cdot\|_2) \leq (\frac{4}{\epsilon})^d$.
Hint. *The volume of the ℓ_2-ball in $\mathbb{R}^d$ scales as r^d, where r is its radius.*