MSBD 5004 Mathematical Methods for Data Analysis
Homework 3

Due date: October 28, Friday

1. Determine whether each of the following scalar-valued functions of n-vectors is linear. If it is a linear function, give its inner product representation, i.e., an n-vector $\mathbf{a}$ for which $f(\boldsymbol{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{x}$ for all $\boldsymbol{x}$. If it is not linear, give specific $\boldsymbol{x}$, $\boldsymbol{y}$, α and β such that

$$f(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) \neq \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y}).$$

   (a) The spread of values of the vector, defined as $f(\boldsymbol{x}) = \max_k x_k - \min_k x_k$.

   (b) The difference of the last element and the first, $f(\boldsymbol{x}) = x_n - x_1$.

2. Suppose $\phi : \mathbb{R}^2 \to \mathbb{R}$ is an affine function, with $\phi(1, 0) = 1$, $\phi(1, -2) = 2$.

   (a) What can you say about $\phi(1, -1)$? Either give the value of $\phi(1, -1)$, or state that it cannot be determined.

   (b) What can you say about $\phi(2, -2)$? Either give the value of $\phi(2, -2)$, or state that it cannot be determined.

   Justify your answers.

3. This question is about the inner product representation of bounded linear functions.

   (a) Consider the function $E_{st} : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by $E_{st}(\boldsymbol{X}) = x_{st}$, where $\boldsymbol{X} = [x_{ij}]_{i,j=1}^{n}$, i.e., E_{st} obtains the (s, t)-entry of a matrix. Find a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $E_{st}(\boldsymbol{X}) = \langle \boldsymbol{A}, \boldsymbol{X}\rangle$ for all $\boldsymbol{X} \in \mathbb{R}^{n\times n}$.

   (b) Consider the function $H : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by $H(\boldsymbol{X}) = \sum_{i+j=n+1} x_{ij}$, where $\boldsymbol{X} = [x_{ij}]_{i,j=1}^{n}$. Therefore, $H(\boldsymbol{X})$ is the summation of anti-diagonals of $\boldsymbol{X}$. Find a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $H(\boldsymbol{X}) = \langle \boldsymbol{A}, \boldsymbol{X}\rangle$ for all $\boldsymbol{X} \in \mathbb{R}^{n\times n}$.

   (c) Given $\boldsymbol{a} \in \mathbb{R}^n$, consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\boldsymbol{x}) = |\langle \boldsymbol{a}, \boldsymbol{x}\rangle|^2$ for any $\boldsymbol{x} \in \mathbb{R}^n$. Obviously f is NOT linear. Nevertheless, we can convert it to a linear function on the "lifted" matrix $\boldsymbol{x}\boldsymbol{x}^T \in \mathbb{R}^{n\times n}$. More precisely, there exists a linear function $F : \mathbb{R}^{n\times n} \to \mathbb{R}$ satisfying $f(\boldsymbol{x}) = F(\boldsymbol{x}\boldsymbol{x}^T)$. Find the inner product representation of F (i.e., find $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that $f(\boldsymbol{x}) = F(\boldsymbol{x}\boldsymbol{x}^T) = \langle \boldsymbol{A}, \boldsymbol{x}\boldsymbol{x}^T\rangle$.) (*This "lifting" technique is quite useful in, e.g., SDP relaxation techniques for clustering.*)

4. Let V be a Hilbert space. Let S_1 and S_2 be two hyperplanes in V defined by

$$S_1 = \{\boldsymbol{x} \in V \mid \langle \boldsymbol{a}_1, \boldsymbol{x}\rangle = b_1\}, \quad S_2 = \{\boldsymbol{x} \in V \mid \langle \boldsymbol{a}_2, \boldsymbol{x}\rangle = b_2\}.$$

   Assume $S_1 \cap S_2$ is non-empty. Let $\boldsymbol{y} \in V$ be given. We consider the projection of $\boldsymbol{y}$ onto $S_1 \cap S_2$, i.e., the solution of

$$\min_{\boldsymbol{x}\in S_1\cap S_2} \|\boldsymbol{x} - \boldsymbol{y}\|. \tag{1}$$

(a) Prove that $S_1 \cap S_2$ is a plane, i.e., if $\boldsymbol{x}, \boldsymbol{z} \in S_1 \cap S_2$, then $(1+t)\boldsymbol{z} - t\boldsymbol{x} \in S_1 \cap S_2$ for any $t \in \mathbb{R}$.

(b) Prove that $\boldsymbol{z}$ is a solution of (1) if and only if $\boldsymbol{z} \in S_1 \cap S_2$ and

$$\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{x} \rangle = 0, \quad \forall \boldsymbol{x} \in S_1 \cap S_2. \tag{2}$$

(c) Find an explicit solution of (1).

(d) Prove the solution found in part (c) is unique.

5. Consider the regression model $y = \boldsymbol{x}^T \boldsymbol{a} + b$, where y is the predicted response, $\boldsymbol{x}$ is an 8-vector of features, $\boldsymbol{a}$ is an 8-vector of coefficients, and b is the offset term. Determine whether each of the following statements is true or false.

   (a) If $a_3 > 0$ and $x_3 > 0$, then $y \geq 0$.

   (b) If $a_2 = 0$ then the prediction y does not depend on the second feature x_2.

   (c) If $a_6 = -0.8$, then increasing x_6 (keeping all other x is the same) will decrease y.

6. Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ be given with $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. Assume $N < n$. Consider the ridge regression

$$\min_{\boldsymbol{a} \in \mathbb{R}^n} \sum_{i=1}^N \left(\langle \boldsymbol{a}, \boldsymbol{x}_i \rangle - y_i \right)^2 + \lambda \|\boldsymbol{a}\|_2^2,$$

   where $\lambda \in \mathbb{R}$ is a regularization parameter, and we set the bias $b = 0$ for simplicity.

   (a) Prove that the solution must be in the form of $\boldsymbol{a} = \sum_{i=1}^N c_i \boldsymbol{x}_i$ for some $\boldsymbol{c} = [c_1, c_2, \ldots, c_N]^T \in \mathbb{R}^N$. (*Hint: Similar to the proof of the representer theorem.*)

   (b) Re-express the minimization in terms of $\boldsymbol{c} \in \mathbb{R}^N$, which has fewer unknowns than the original formulation as $N < n$.