IE643 Deep Learning : Theory and Practice **July-Dec 2025**

Homework 3: For self-practice

1 Instructions

We make the following assumptions for all questions: a vector $u \in \mathbb{R}^d$ for some $d \geq 1$, is represented as $\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{pmatrix}$; a vector $u \in \mathbb{R}^d$ with $u_i = 0 \; \forall i = 1, \ldots, d$ is called a zero vector and is represented by $\mathbf{0}$; the transpose of a vector u is denoted by $u^\top$; the notation $|\alpha|$ denotes the absolute value of some $\alpha \in \mathbb{R}$.

Note: Please try all questions.

2 Questions

1. The inner product (or dot product or scalar product) between two vectors $u, v \in \mathbb{R}^d$ is defined as $\langle u, v \rangle = \sum_{i=1}^d u_i v_i$. The following properties are related to the inner product:
   (a) Prove $\langle u, v \rangle = \langle v, u \rangle, \forall u, v \in \mathbb{R}^d$.
   (b) Prove $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle, \forall u, v, w \in \mathbb{R}^d$.
   (c) Prove $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle, \forall u, v \in \mathbb{R}^d, \forall \alpha \in \mathbb{R}$.
2. A vector norm defined on $\mathbb{R}^d$ is a function $\| \cdot \| : \mathbb{R}^d \longrightarrow [0, +\infty]$, which satisfies the following properties:
   - **[Non-negativity]** $\|u\| \geq 0, \forall u \in \mathbb{R}^d$ and $\|u\| = 0$ if and only if $u = \mathbf{0}$ is the zero vector.
   - **[Absolute scaling]** $\|\alpha u\| = |\alpha| \|u\|, \forall u \in \mathbb{R}^d, \forall \alpha \in \mathbb{R}$.
   - **[Triangle inequality]** $\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in \mathbb{R}^d$.

   The following questions are related to vector norms.

   (a) Consider $\|u\|_2 = \sqrt{\langle u, u \rangle} = \sqrt{\sum_{i=1}^d u_i^2}$. Show that $\| \cdot \|_2$ is a vector norm. You must verify all the three properties described above. (Note that this is the popular Euclidean norm and is induced by the inner product definition given in the previous question.)
   (b) Consider $\|u\|_1 = \sum_{i=1}^d |u_i|$. Show that $\| \cdot \|_1$ is a vector norm.
   (c) Consider $\|u\|_p = \left[\sum_{i=1}^d |u_i|^p \right]^{\frac{1}{p}}$, where $p > 2$. Show that $\| \cdot \|_p$ is a vector norm.
   (d) Consider $\|u\|_\infty = \max\{|u_1|, |u_2|, \ldots, |u_d|\}$. Show that $\| \cdot \|_\infty$ is a vector norm.
3. Prove the Cauchy-Schwarz inequality $|\langle u, v \rangle| \leq \|u\|_2 \|v\|_2, \; \forall u, v \in \mathbb{R}^d$.

4. (a) Consider the data set $D = \{(x^t, y^t)\}_{t=1,2,3...}$, where $x^t \in \mathbb{R}^d$ and $y^t \in \{+1, -1\}$. D is said to be linearly separable by a hyperplane $H = (w^*, b^*)$ when the following holds:

$$y^t(\langle w^*, x^t \rangle - b^*) \geq \gamma > 0. \tag{3.1}$$

Show that for every data set D which is linearly separable in the sense of the condition (3.1), there exists another data set $\tilde{D} = \{(\tilde{x}^t, \tilde{y}^t)\}_{t=1,2,3...}$ where $\tilde{x}^t = \mu^t x^t + \beta^t$ and $\tilde{y}^t = y^t$, for every $t = 1, 2, 3, \ldots$, such that $\tilde{D}$ is linearly separable by a hyperplane $H = (u^*, 0)$. Find appropriate values of μ^t and β^t. Find the relation between w^* and u^*.

(b) Using your answer to part (a), justify the absence of the intercept term b^* in linear separability condition discussed in class.

5. Consider the perceptron learning algorithm with a starting point $w^1 = \begin{bmatrix}\theta \; \theta \; \ldots \; \theta\end{bmatrix}^\top$ where $\theta \in \mathbb{R}$, used to train on a linearly separable data set.

(a) Find a suitable upper bound on the number of mistakes for the choice of starting point w^1 given above.

(b) Compare and contrast the bound you obtained in part (a) with the bound discussed in class. Explain the changes observed in the bound, and explain the dependence of the bound you obtained in part (a) on the choice of w^1.

(c) Check if your bound is tight or if it can be improved.