HW Assignment 1, Theory

1 Theory (55 + 10 points)

For each problem below, show your work. Clear and complete explanations and proofs of your results are as important as getting the right answer and are required for getting all the points. Problems marked with (∗) are mandatory only for ITCS 8156.

1. [**Linear Separability**, 25 points]
   Consider the two-dimensional XOR dataset D that contains the following 4 examples with binary labels $D = \{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), (\mathbf{x}_3, t_3), (\mathbf{x}_4, t_4)\} = \{([0,0], -1), ([1,1], -1), ([0,1], +1), ([1,0], +1)\}$, i.e., the truth table of the logical XOR function.

ϕ_1	ϕ_2	t
0	0	-1
0	1	$+1$
1	0	$+1$
1	1	-1

   (a) Prove that D is not linearly separable.
   - *Hint*: Show that there cannot be a vector of parameters $\mathbf{w}$ such that $\mathbf{w}^T\mathbf{x} \geq 0$ for all examples $\mathbf{x}$ that are positive, and $\mathbf{w}^T\mathbf{x} < 0$ for all examples $\mathbf{x}$ that are negative. Do not forget to add the bias feature $\phi_0 = 1$ to each example $\mathbf{x}$.

   (b) Will the Perceptron algorithm converge on this dataset? Explain.

   (c) Add a third feature to each example in the dataset that is the product of the two original features, i.e. for each example $\mathbf{x}$ the new feature is computed as $\phi_3(\mathbf{x}) = \phi_1(\mathbf{x})\phi_2(\mathbf{x})$. Is the new dataset linearly separable? Prove your answer.

2. [**Linear Separability**, 10 + 10 points]
   Consider the two-dimensional AND dataset D that contains the following 4 examples with binary labels $D = \{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), (\mathbf{x}_3, t_3), (\mathbf{x}_4, t_4)\} = \{([0,0], -1), ([1,1], +1), ([0,1], -1), ([1,0], -1)\}$, i.e., the truth table of the logical AND function.

ϕ_1	ϕ_2	t
0	0	-1
0	1	-1
1	0	-1
1	1	$+1$

   (a) Prove that D is linearly separable.
   - Do not forget to add the bias feature $\phi_0 = 1$ to each example.

   (b) (∗) Prove that the Perceptron algorithm makes at most 51 mistakes when trained on this dataset before it converges.

- *Hint*: Use $\mathbf{u} = [-1.5, 1, 1]/\sqrt{4.25}$ in Theorem 1 (Block, Novikoff) from this paper that proves the Perceptron convergence property.
- [Bonus, 10p] Find a different vector $\mathbf{u}$ that results in a tighter bound at (b) above, i.e. a smaller number of mistakes.

(c) [Bonus, 5p] Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors such that $\|\mathbf{v}\| = 1$. Prove that $\|\mathbf{u}\| \geq \mathbf{u}^T\mathbf{v}$.

3. [**Perceptron Analysis**, 20 points]
Let $\mathbf{w}$ by the current vector of parameters in the Perceptron algorithm, right before executing the `if` statement on line 4.

(a) Prove that if the algorithm executes the `if` clause on line 5, i.e. the model made a mistake on example $\mathbf{x}_n$, then it is true that $t_n\mathbf{w}^T\mathbf{x}_n \leq 0$.

(b) Let $\overline{\mathbf{w}}$ be the new weight vector after executing the `if` clause on line 5, i.e. $\overline{\mathbf{w}} = \mathbf{w} + t_n\mathbf{x}_n$. Prove that $t_n\overline{\mathbf{w}}^T\mathbf{x}_n \geq t_n\mathbf{w}^T\mathbf{x}_n$.

2 Submission

Submit your responses on Canvas as one file named `theory.pdf`. It is recommended to use an editor such as Latex or Word that allows editing and proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it on Canvas. Make sure that your writing is legible and easy to read.