Machine Learning, 2024 Fall
Quiz 4

1 True or False [20 pts]

1. For the data $\mathbf{x} = [x_1, x_2, x_3]^T$, the model $f(x) = \text{sign}(w_0 + w_1x_1 + w_2x_2^2 + w_3x_3^2)$ is a linear classifier.
   solution: True

2. Assuming a data set $\mathcal{X} = \{\mathbf{x}_i\}$ is linearly separable, then the perceptron learning algorithm can correctly classify all the data in $\mathcal{X}$.
   solution: True

3. If a dataset is linearly inseparable in two-dimensional space, then mapping the data to three-dimensional space will definitely make the dataset linearly separable.
   solution: False

4. When we run the perceptron learning algorithm (PLA) on the dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, , N\}$, the order of the data does not affect the final weight $\mathbf{w}$ of the PLA.
   solution: False

2 Perceptron Learning Algorithm [20 pts]

For the 2-dimensional data $\mathbf{x} = (x_1, x_2)$ the linear classifier f is defined as

$$f(x) = \text{sign}(w_1x_1 + w_2x_2 + w_0)$$

$$\text{sign}(t) = \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{otherwise} \end{cases}$$

The initial values of w_1, w_2, and w_0 are all 0. Given the following training set, can the Perceptron Learning Algorithm (PLA) obtain the weights of a linear classifier f? If yes, please show the **calculation process**. If not, please explain the **reason**.

1. The training data are as follows. Obtain the converged values of w_1, w_2, and w_0.

x_1	x_2	t
0	0	0
0	1	1
1	0	1
1	1	1

   solution:

$$w_1 = 2, w_2 = 1, w_0 = 0$$

2. The training data are as follows. Obtain the converged values of w_1, w_2, and w_0.
   solution: This is linear inseparable.

x_1	x_2	t
0	0	0
0	1	1
1	0	1
1	1	0

3 Prediction and Test [20 pts]

In a binary classification problem, the true labels are $\{1, 0, 1, 0, 1, 1\}$, and the results from a linear classifier are $\{0, 0, 0, 1, 1, 1\}$.

1. Write down all the values of TP, FP, FN, and TN.
2. What are the **Recall** and **Fall-out** of this classifier?

solution:

$$TP = 2; FP = 1; FN = 2; TN = 1;$$

so the recall is $\frac{TP}{TP+FN} = \frac{2}{4}$ and fall-out is $\frac{FP}{FP+TN} = \frac{1}{2}$

4 Loss Function [20 pts]

In training a linear classifier $h = f(\mathbf{W}^T\mathbf{x})$, it is common to use the sigmoid function instead of a threshold function, and the optimization objective is cross-entropy loss. Please answer the following questions:

1. Why use the sigmoid function instead of a threshold function?
2. Why not use MSE to optimize the linear classifier?

solusion:

1. (1) The sigmoid function is differentiable. (2) The sigmoid function can position the decision boundary away from the data.
2. (1) This is because binary classification assumes the data follows a Bernoulli distribution, while MSE assumes the data follows a Gaussian distribution. (2) The cross-entropy loss function is convex and has a unique optimum, whereas MSE as a loss function is non-convex.

5 The Logistic Function [20 pts]

The Binary logistic regression is deined as:

$$P_1(y \mid \mathbf{x}) = \begin{cases} \theta(\mathbf{w}^T\mathbf{x}) & \text{for } y = 1 \\ 1 - \theta(\mathbf{w}^T\mathbf{x}) & \text{for } y = 2 \end{cases}$$

$$\theta(s) = \frac{e^s}{1 + e^s}$$

And the Multiclass softmax regression is deined as:

$$P_2(y = k \mid \mathbf{x}) = \frac{\exp(\mathbf{w}_k^T\mathbf{x})}{\sum_{j=1}^{K} \exp(\mathbf{w}_j^T\mathbf{x})}$$

Please show that the binary logistic regression and multiclass softmax regression are equicalent for $K = 2$

solution: See lecture5-classification Slides 38