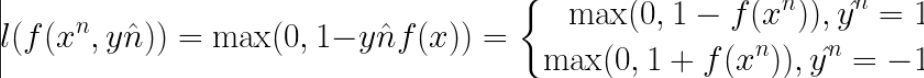
**Problem 1**

In addition to linear regression and neural networks, we can consider a combination of support vector machines (SVM) and hinge loss functions. SVM is a supervised learning model for classification. The goal of SVM is to find a hyperplane that not only correctly classifies the training data, but also maximizes the distance to the nearest training sample.

Definition of Hinge Loss:

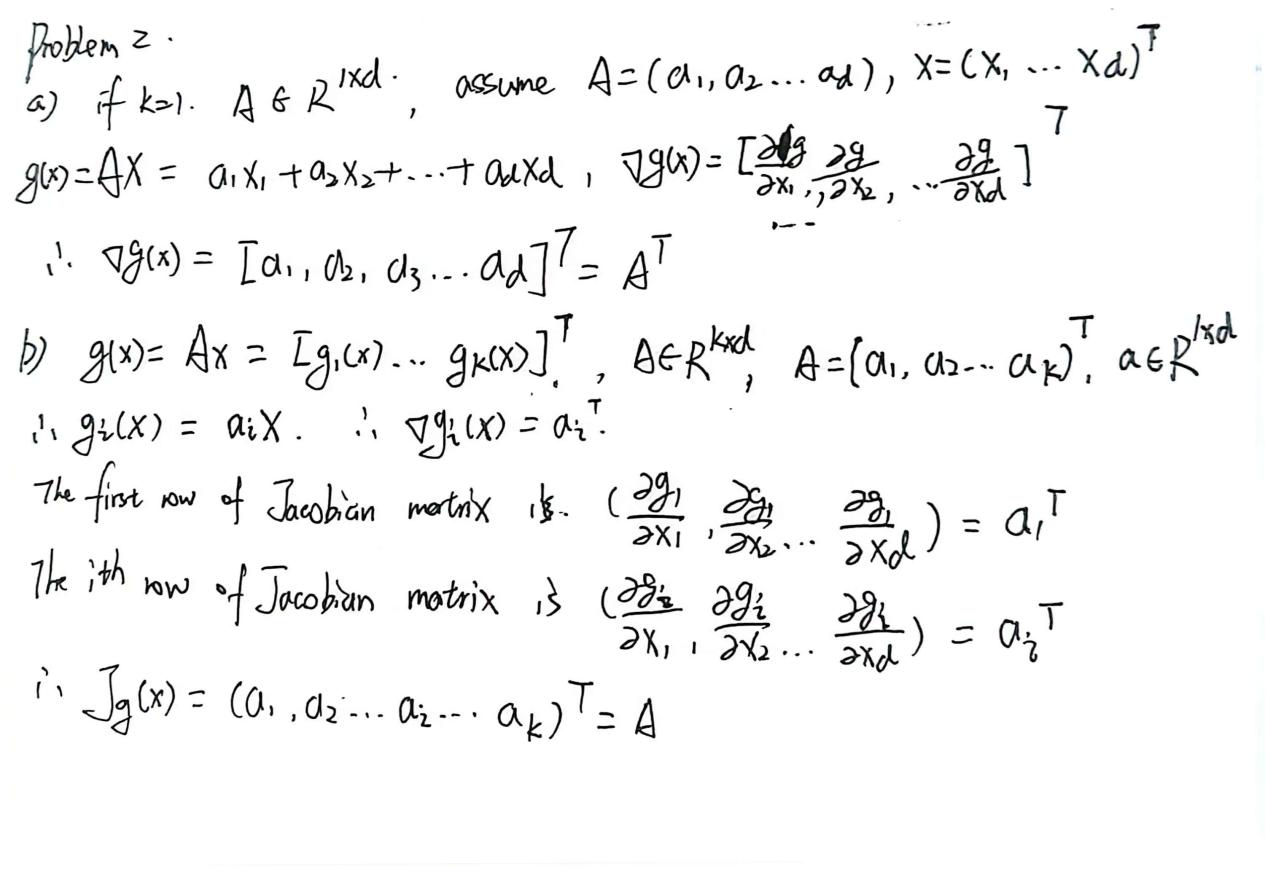


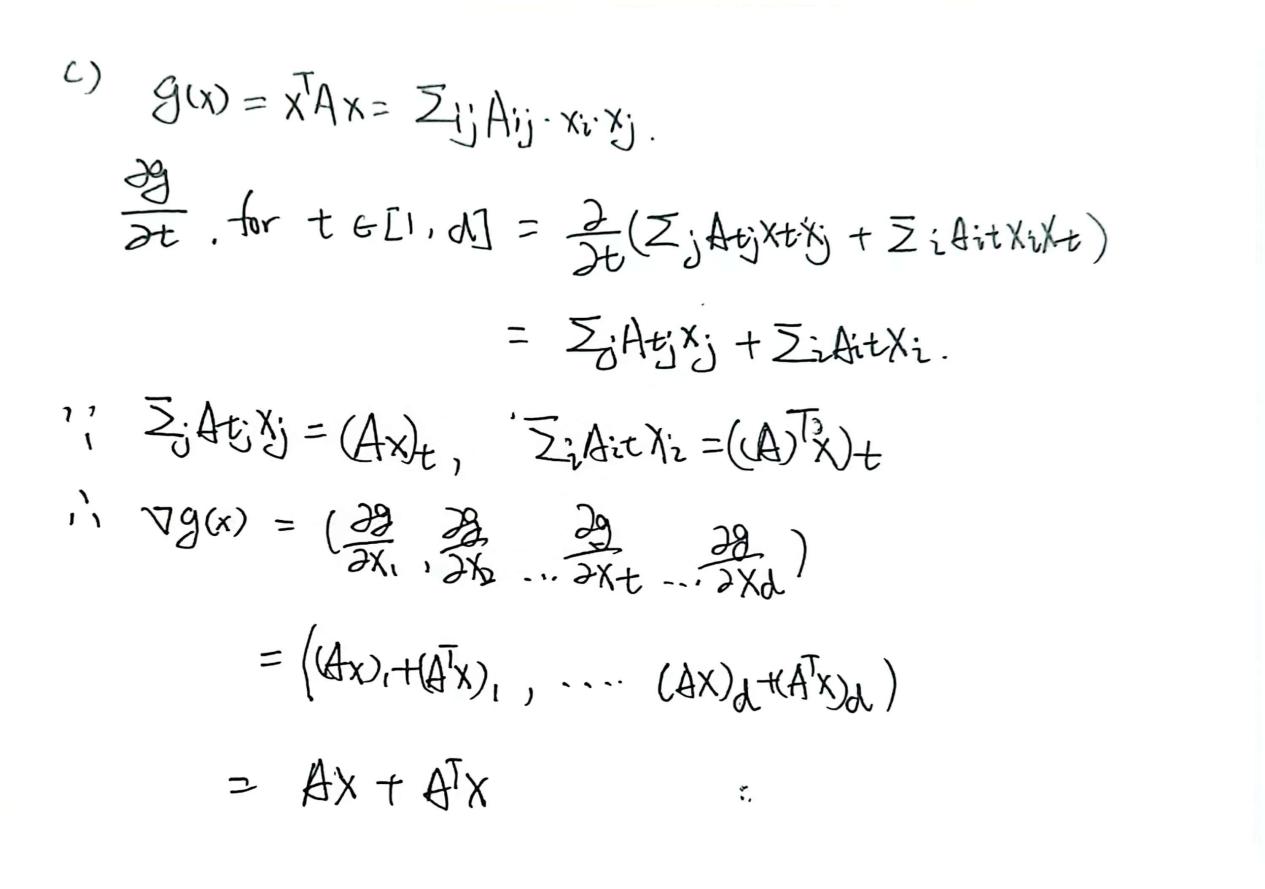
When yn = 1, f(x^n) > 1, then loss = 0;

When yn = -1, f(x^n)< -1, then loss = 0;

If the f(x^n) is greater than the threshold(=1), loss will not change(=0), therefore sometimes using sigmoid Loss with cross entropy could get better result.

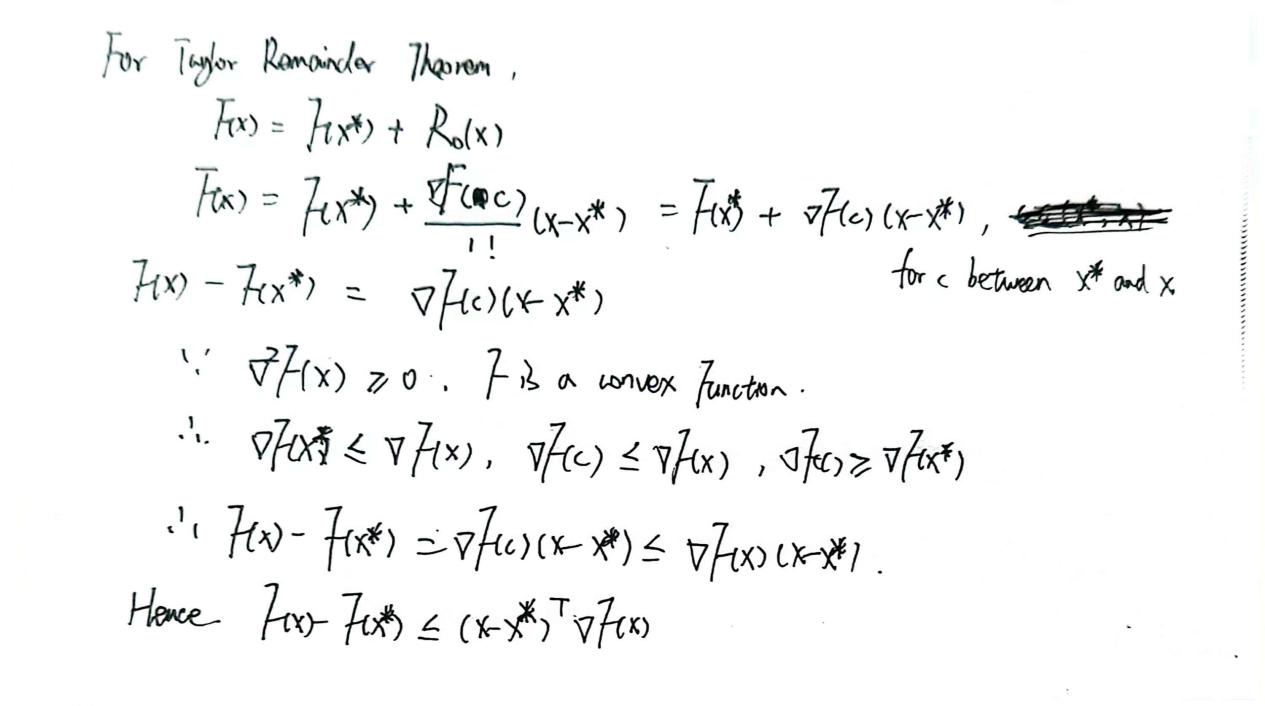
**Problem 2**

****

****

**Problem 3**

**Problem 4**

****

**Problem 5**

∇F(wt) = ∇ft(wt) = ∇() = -2

W for , =0.1