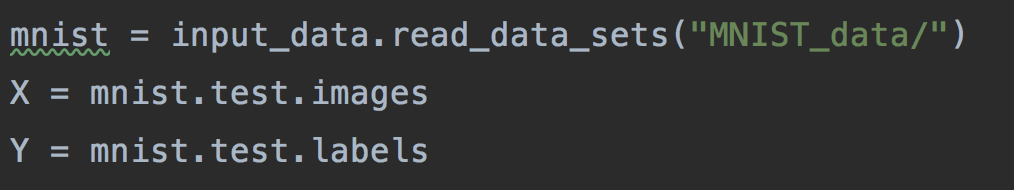
**Problem Set 3**

1. **Problem 1**
   1. Load dataset

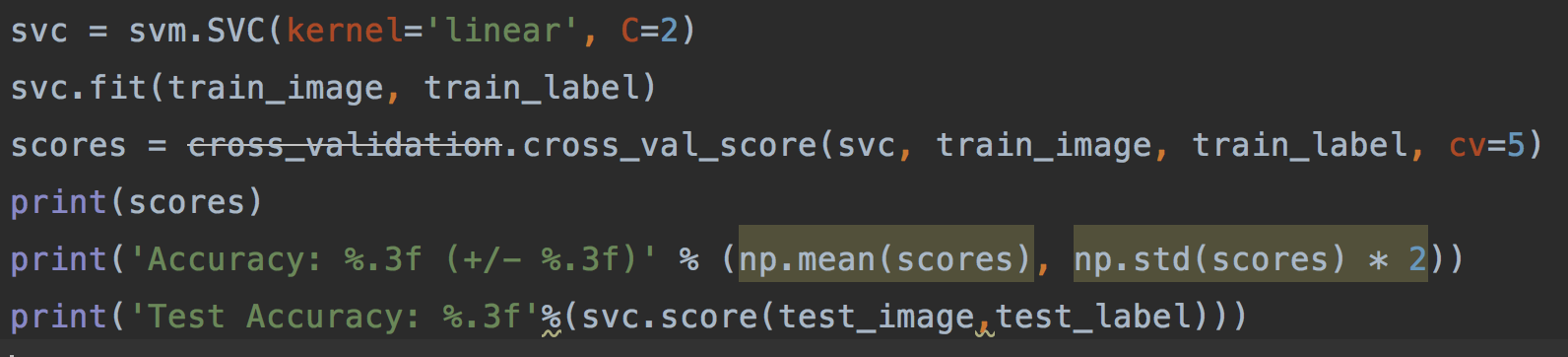
****

Because the size of mnist.train is too large, the it will spend too much time to train the model. I choose the mnist.test whose size is smaller as the total set.

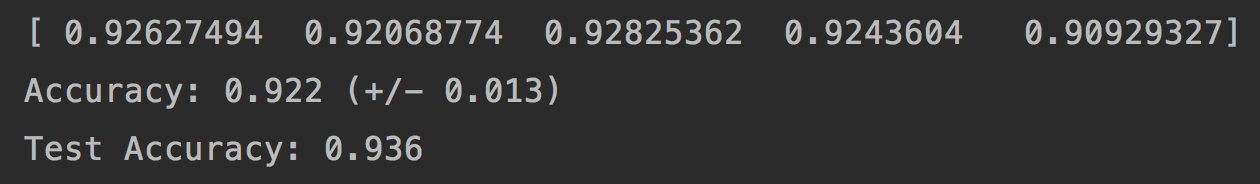
* 1. Split the dataset

****

* 1. Train and test



* 1. Result



1. **Problem 2**

Identify the Lagrange dual problem of the following primal problem

Given feature

Minimize the weihted sum between the squared length of the separating vector and the errors, whegre w is the separating vector, is the dot product, and is the error made by separating vector on feature

Primal formulation:

Set partial derivative of Lagrange function

Substituting into Lagrange function, we get dual problem of maxizing:

b = 0

C-

So the dual problem is:

* 1. **Point out what is the "margin" in both the primal formulation and the dual formulation**

Margin in primal formulation:

Margin =

Margin in dual formulation

Margin =

* 1. **what are the benefits of maximizing the margin**

One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized.

* 1. **Characterize the support vectors**

Only the vectors which can satisfy are support vectors. It means the points on the margin are the support vectors.

* 1. **Point out the benefit of solving the dual problem instead of the primal problem**

exchange max and min, we can get the dual problem

The benefit:

The primal problem is to minimize the After being converted to a dual problem, the problem is to maximize the . There are less variables in dual problem and the dual problem is easier to solve than the primal problem.