CS541 Homework 3

1. **Stochastic Gradient Descent**

Suppose F(w) is convex,

So

*Also,*

*= 0, or*

In SGD

So

2(a).

**3.**

, else

So, F(w) is – strongly convex, the solution is unique.

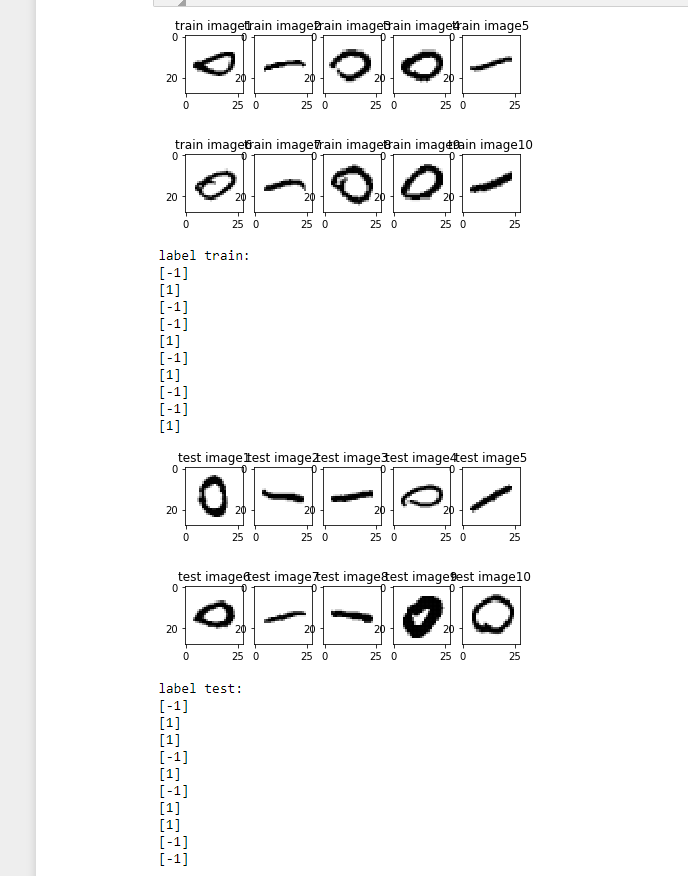
Detailed update rule:

,

,

**2.Support Vector Machine**

**Data Preparation**



* 1. **Evaluation Metric (implement by code)**
  2. **Convergence of SGD**
* **λ = 1 and rate = 1/t**

|  |  |
| --- | --- |
|  |  |

**From the data in F(w) we can see that F(w) will close to 0 when the iteration value is about 500.**

* **λ = 1 and rate = 0.01**

|  |  |
| --- | --- |
|  |  |

**From the picture and data in F(w) we find that when learning rate is 0.01, F(w) won’t close to 0.**

* **λ = 1 and rate = 1/t^2**

|  |  |
| --- | --- |
|  |  |

**From the picture we can find that F(w) won’t close to 0 when learning rate is 1/t^2.**

* 1. **Hyper-Parameter**

λ in [0.000001, 0.001, 0.1, 0.5, 1, 2, 5, 10, 20, 50, 100, 500, 1000, 10000]

|  |  |
| --- | --- |
|  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0.000001 | 0.001 | 0.1 | 0.5 | 1 | 2 | 5 |
| ||w10000||2/d | 3.8033702 | 2.79457159 | 1.0543638 | 0.02884166 | 0.00158474 | 0.000982787 | 0.000784602 |
|  | 10 | 20 | 50 | 100 | 500 | 1000 | 10000 |
|  | 0.0005003 | 0.00010653 | 0.0005128 | 0.00029024 | 6.2906E-06 | 4.59301E-06 | 2.71913E-06 |

When λ increase, the value of become smaller, the accuracy of training and testing data also decrease when λ become too big.

I select the best w when λ = 0.5 for further work.

**Visualization and accuracy for sparsity s = [10, 20, 50, 100, 200, 400]**

|  |
| --- |
|  |
|  |
|  |

From the accuracy data we can find that the value of accuracy is well when s is equal to or bigger than 100, but the accuracy is not so good when s small than 100. For our picture data, the edge of each picture is blank and only data in the middle of the picture is useful for training and testing. So when s is big enough to contain the data at the middle of the picture, the model performs as well as the whole .

2.5 Noisy Labels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Training Acc | Testing Acc | ||w||2 / d |
| 0 | 1 | 0.9987 | 0.999 | 0.00268381 |
| 0.01 | 5 | 0.9891 | 0.998 | 0.001081923 |
| 0.1 | 0.001 | 0.9305 | 0.987 | 2.412479679 |
| 0.2 | 1 | 0.9058 | 0.995 | 0.003510566 |
| 0.3 | 20 | 0.8239 | 0.965 | 0.001100146 |
| 0.5 | 5 | 0.7987 | 0.982 | 0.001801701 |
| 0.7 | 5 | 0.7487 | 0.971 | 0.002051681 |

When the noisy level increase, the value of λ that get the best testing accuracy is become bigger, the Training accuracy decrease with the increasing of noisy level and λ , decrease when λ become bigger.