Introduction to Data Mining: 21191441
Zhejiang University
Professor Deng Cai

May 13, 2019
Homework 2

Homework 2

Collaborators:
Name: Cai Jiayi
Student ID: 3160101462

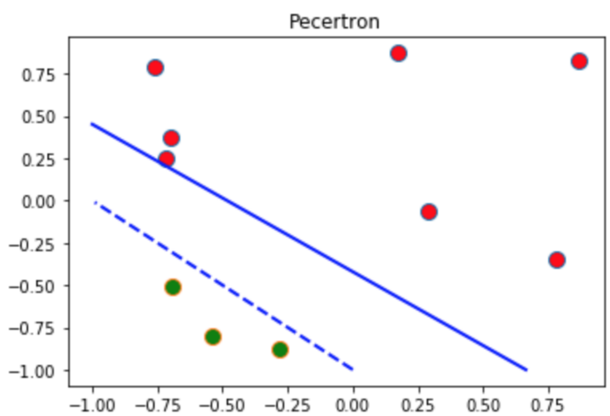
Problem 2-1. A Walk Through Linear Models

(a) Perceptron

Answer:

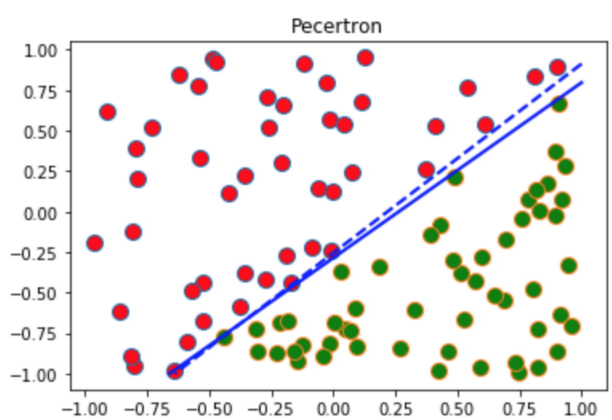
1. $n\_train = 10, n\_test = 10$

```
E_train is 0, E_test is 0.0967
Average number of iterations is 7.
```



$n\_train = 100, n\_test = 100$

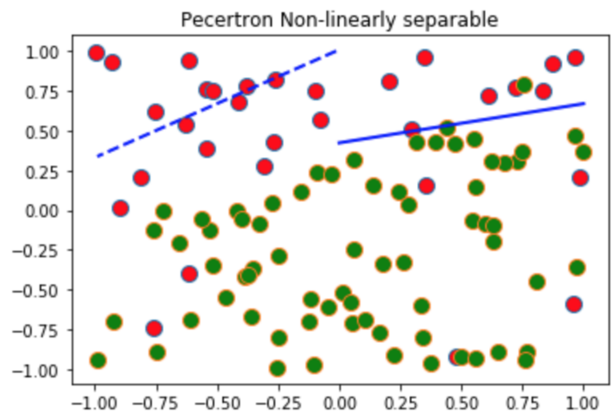
```
E_train is 0.00017, E_test is 0.01396
Average number of iterations is 41.
```



2. n_ train=10, average_ iter=7
   n_ train=100, average_ iter=41

3. If the training data is not linearly separable, the learning algorithm will never converge.

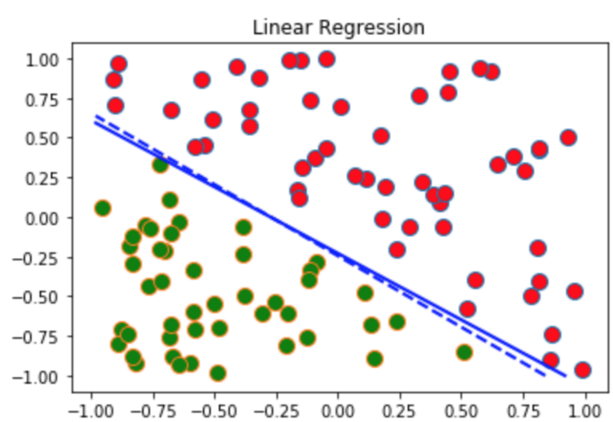
```
E_train is 0.24762, E_test is 0.19376
Average number of iterations is 1000.
```



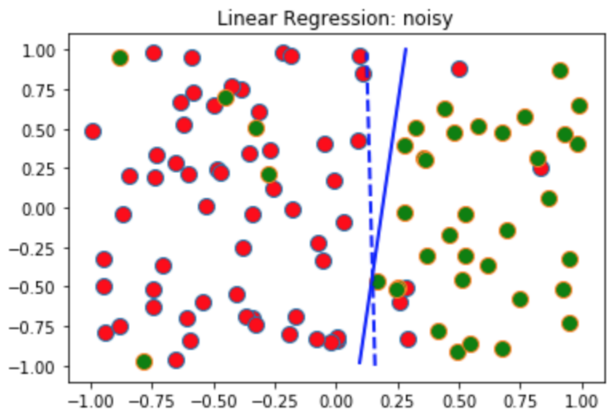
(b) Linear Regression

Answer:

```
E_train is 0.03748, E_test is 0.04599
```

1. 

```
E_train is 0.13319, E_test is 0.05918
```

2. 

3. `E_train is 0.49, E_test is 0.5496`

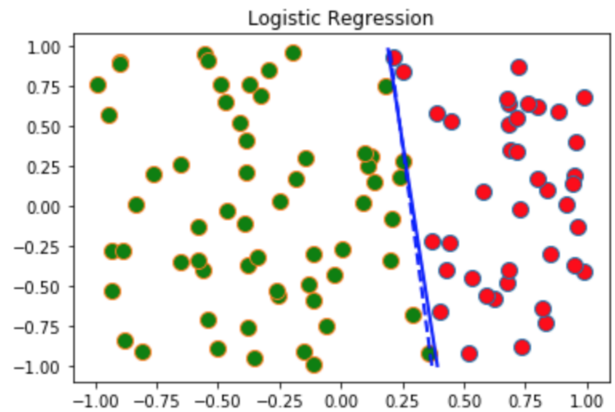
4. `E_train is 0.05, E_test is 0.066`

(c) Logistic Regression

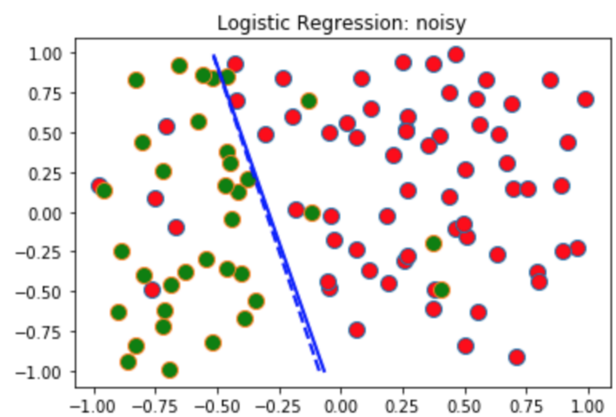
Answer:

1. $h_\theta(x) = g\left(\theta^T x\right)$
   $z = \theta^T x$
   $g(z) = \frac{1}{1+e^{-z}}\ (sigmoid\ function)$
   $P(y|x;\theta) = (h_\theta(x))^y\,(1 - h_\theta(x))^{1-y}$
   $L(P) = ylog(h_\theta(x)) + (1-y)log(1-h_\theta(x))$
   loss function:
   $l(\theta) = -\frac{1}{m}\cdot(y\log(h) + (1-y)\log(1-h))$
   $l(\theta) = -\frac{1}{m}\left(y\left[\log\left(\frac{1}{1+e^{-\theta^T x}}\right) - \log\left(\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right)\right] + \log\left(\frac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right)\right)$
   $l(\theta) = -\frac{1}{m}\left(y\theta^T x + \log\left(\frac{1}{1+e^{\theta^T x}}\right)\right)$
   $l(\theta) = -\frac{1}{m}\left(y\theta^T x - \log\left(1 + e^{\theta^T x}\right)\right)$
   gradient:
   $\nabla_\theta l = -\frac{1}{m}((yx - \left[\frac{1}{1+e^{-\theta x}}x\right]))$
   gradient descent:
   $\theta_{n+1} = \theta_n - \gamma\nabla l\left(\theta_n\right), n \geq 0$

`E_train is 0.0929, E_test is 0.0891`



`E_train is 0.201021172111, E_test is 0.145227156669`



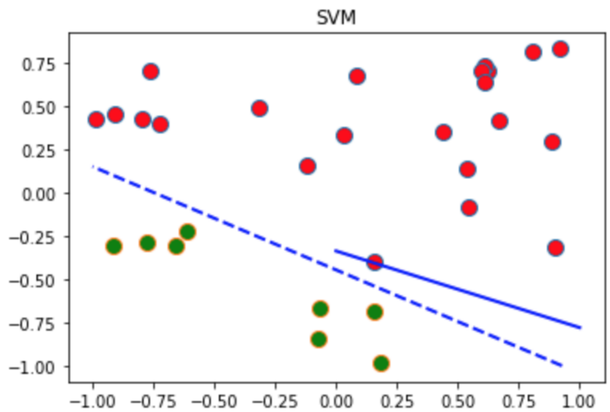
2.

(d) Support Vector Machine

Answer:

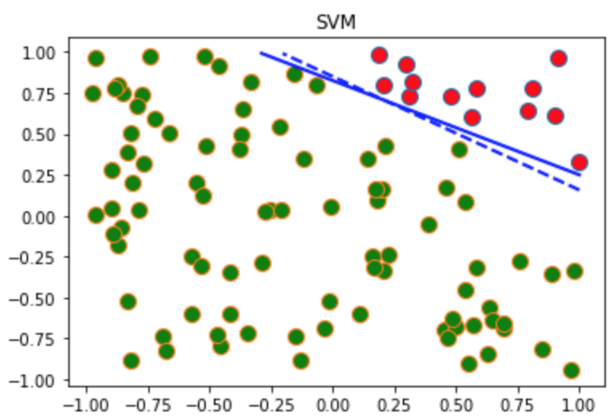
1. $\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$
   $y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1$

```
E_train is 0.0, E_test is 0.0343
Average number of support vectors is 2.977.
```




```
E_train is 0.0, E_test is 0.01036
Average number of support vectors is 2.997.
```

2.



3. d = y*np.dot(w.T,X)
   num = np.sum(np.absolute(d-1)<0.0001)
   When nTrain = 100, the average number of support vectors is 2.997.

Problem 2-2. Regularization and Cross-Validation

(a) Implement Ridge Regrssion, and use LOOCV to tune the regularization parameter λ.

Answer:

1. Ridge Regression
$\boldsymbol{a}^* = \operatorname{argmin} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{a}\right)^2 + \lambda \sum_{j=1}^{p} a_j^2$
$\boldsymbol{a}^* = \left(XX^T + \lambda \boldsymbol{I}\right)^{-1} X\boldsymbol{y}$

```
0.001 16
0.01 11
0.1 15
0.0 16
1.0 32
10.0 67
100.0 90
1000.0 94
('The lmbda chosen by LOOCV is ', 0.01)
```

2.
```
Without regularization, the sum of omega square is  5533.060166528366
With regularization, the sum of omega square is  1046.7918406339231
```

3.
```
Without regularization, the training error is:  0.0
With regularization, the training error is:  0.0
Without regularization, the testing error is:  0.14264188849824208
With regularization, the testing error is:  0.07935710698141638
```

(b) Implement Logistic Regression, and use LOOCV to tune the regularization parameter λ.

Answer:

loss function:
$l(\theta) = -\frac{1}{m} \cdot (y \log(h) + (1-y)\log(1-h)) + \frac{1}{m}\frac{1}{2}\theta^T\theta$
$l(\theta) = -\frac{1}{m}\left(y\theta^T x - \log\left(1 + e^{\theta^T x}\right)\right) + \frac{1}{m}\frac{1}{2}\theta^T\theta$
gradient:
$\nabla_\theta l = -\frac{1}{m}((yx - \left[\frac{1}{1+e^{-\theta x}}x\right])) + \frac{1}{m}\theta$
gradient descent:
$\theta_{n+1} = \theta_n - \gamma \nabla l\left(\theta_n\right), n \geq 0$

```
The lmbda chosen by LOOCV is  0.001
```

```
Without regularization, the training error is:  0.0
With regularization, the training error is:  0.005
Without regularization, the testing error is:  0.18583626318432947
With regularization, the testing error is:  0.17629331993972877
```

Problem 2-3. Bias Variance Trade-off

Let's review the bias-variance decomposition first. Now please answer the following questions:

(a) True of False

Answer:

1.
2.
3.
4.
5.

1.False. If the model is suffering from high bias, you should add more parameters to the model.
2.False. Models with high variance may be overfitting.
3.True.
4.False. If the regularization parameter is too large, the performance will be worse.
5.False.