CSCI567 2014 Homework Assignment 2

Student Name Yi Duan

USC ID 9966845562

I have collaborated with Meng Sun

Yun Li

1. Linear Regression
2. Regression with heterogenous noise

5. Smooth Coefficients

8. Linearly Constrained Linear Regression

Put in and get

Put in (1) and get

1. Online Learning

According to perceptron, we need to update if . In order to minimize while the new classifies correctly on the current sample, we project on the the direction which is perpendicular to and that is the critical line of . To get , we slightly move the projection of to the opposite direction of **.** The critical value of **.** And **.**

1. Kernels
2. Let ***v*** be any vector s.t. ***v∈RN\*1***. We know that **.**
3. (**x, x’**) = <**ψ(x),ψ(x’)**> where ψ : **x** → f (**x**); Thus, K is positive semi-definite.
4. Note that the gram matrix K5 for k5 is the Hadamard product (or element-by-element product) of K1 and K2 (K5 = K1K2). Suppose that K1 and K2 are covariance matrices of (X1, . . . ,Xn) and (Y1, . . . , Yn) respectively. Then K5 is simply the covariance matrix of (X1Y1, . . . ,XnYn), implying that it is symmetric and positive definite.
5. Bias-Variance Trade-off

Given ,

(a)Suppose

(b)

(c)

(d)According to the bias-variance theorem, squared error

When is small, for example extremely to 0, bias term will become small to 0 and variance term will dominate. When become larger, the increase of bias term will be faster than variance term and bias term will dominate.

1. Programming
2. {(‘enron’:600), (‘will’:351), (‘please’:291)}
3. Unregularized logistic regression:

regularized logistic regression:

1. (a)Ionosphere



EmailSpam



(b)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| L2 norm | 0.001 | 0.01 | 0.05 | 0.1 | 0.5 |
| Ionosphere | 1.4946 | 4.8015 | 18.5577 | 38.5218 | 196.8738 |
| EmailSpam | 2.5660 | 6.9126 | 12.2558 | 16.0453 | 22.1830 |

1. (a)Ionosphere



EmailSpam



(b)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| L2 norm | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| Ionosphere | 4.6553 | 4.5648 | 4.4784 | 4.3958 | 4.3167 | 4.2409 |
| EmailSpam | 7.7083 | 7.4753 | 7.2548 | 7.0461 | 6.8487 | 6.6619 |
| L2 norm | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |  |
| Ionosphere | 4.1681 | 4.0979 | 4.0301 | 3.9645 | 3.9012 |  |
| EmailSpam | 6.4852 | 6.3182 | 6.1602 | 6.0107 | 5.8694 |  |

(c)Ionosphere











EmailSpam











1. Unregularized logistic regression:

regularized logistic regression:

1. (a)Ionosphere



EmailSphere



(b)

|  |  |
| --- | --- |
| L2 norm |  |
| Ionosphere | 1.0077e+16 |
| EmailSpam | 645.7744 |

(c)

|  |  |
| --- | --- |
| Cross-entropy value |  |
| Ionosphere | 551.6794 |
| EmailSpam | 3.2924e+03 |

1. (a)Ionosphere



EmailSpam



(b)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| L2 norm | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| Ionosphere | 4.3121e+03 | 2.1607e+03 | 1.4373e+03 | 1.0776e+03 | 857.6358 |
| EmailSpam | 12.4546 | 10.3472 | 9.2040 | 8.4370 | 7.8689 |
| L2 norm | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
| Ionosphere | 714.4020 | 612.1160 | 535.4573 | 475.8790 | 428.2429 |
| EmailSpam | 7.4227 | 7.0584 | 6.7525 | 6.4903 | 6.2618 |

(c)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cross-entropy value | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| Ionosphere | 9.3074e+05 | 4.6788e+05 | 3.1092e+05 | 2.3330e+05 | 1.8492e+05 |
| EmailSpam | 172.5871 | 159.7257 | 153.7903 | 150.3709 | 148.1985 |
| Cross-entropy value | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
| Ionosphere | 1.5414e+05 | 1.3217e+05 | 1.1572e+05 | 1.0294e+05 | 9.2728e+04 |
| EmailSpam | 146.7477 | 145.7537 | 145.0664 | 144.5944 | 144.2791 |

1. When the iteration goes up, the rate of convergence becomes higher. The larger the step size is, the faster the data converges, the larger the cross-entropy function value is and the larger the L2 norm of ***w*** is. The larger the is, the larger the L2 norm of ***w*** is.
2. Gradient descent has a relatively small cross-entropy function value and L2 norm of of ***w*** while these two values of Newton’s method are bigger. Also Gradient descent converges slower and has smaller computation time compared to Newton’s method.