HW 2 Machine Learning Chengjun Yuan

1.1 For linear regression with ridge regularization:

$$RSS(\beta:\lambda) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta$$

$$= \beta^T X^T X\beta - 2\beta^T X^T Y + Y^T Y + \lambda\beta^T\beta$$

To find the $\hat{\beta}^{ridge}$ to minimize the $RSS(\beta:\lambda)$:

$$\frac{\partial RSS(\beta:\lambda)}{\partial\beta} = \frac{\partial(\beta^T X^T X\beta)}{\partial\beta} - \frac{\partial(2\beta^T X^T Y)}{\partial\beta} + \frac{\partial(Y^T Y)}{\partial\beta} + \frac{\partial(\lambda\beta^T\beta)}{\partial\beta}$$

$$= 2X^T X\beta - 2X^T Y + 0 + \frac{\partial(\beta^T(\lambda I)\beta)}{\partial\beta}$$

$$= 2X^T X\beta - 2X^T Y + 2(\lambda I)\beta$$

So, through setting $\frac{\partial RSS(\beta:\lambda)}{\partial\beta} = 0$, we can get ridge estimator:

$$(X^T X + \lambda I)\beta = X^T Y \quad \Rightarrow \quad \hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T Y$$

1.2 It <u>cannot</u> be solved through linear regression. Because $|X^T X| = 0$, there is no presence of $(X^T X)^{-1}$. So we cannot use the normal equation to do linear regression for it.

1.3 The <u>lasso</u> regularized linear regression should be chosen since it is L1 norm.

1.4.1 The linear equation with $\lambda = 0.0$ is :

betaLR=ridgeRegression.ridgeRegress(xVal,yVal,lambda=0)

The returned betaLR is [2.9714, -11.0033, 6.9623]

which means: $y = 2.9714 - 11.0033 * x1 + 6.9623 * x2$

The plots of original data and the plane from linear regression are shown in FIG. 1 & FIG. 2 below.

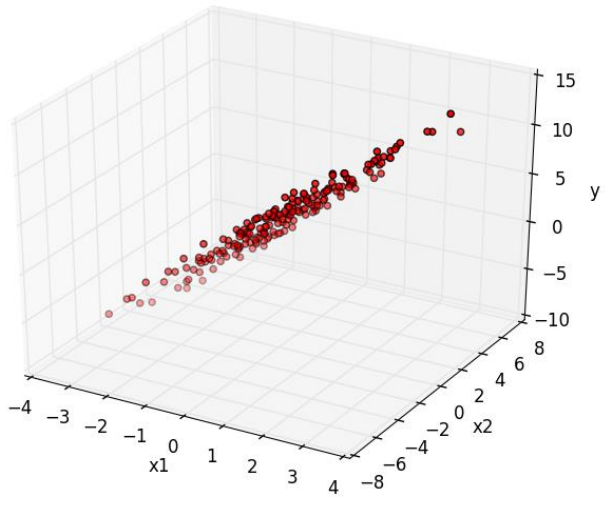

FIG. 1

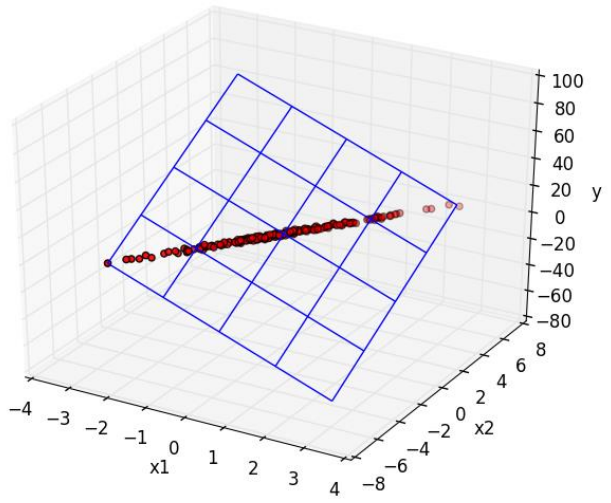

FIG. 2

1.4.2

The plot of λ vs J(β) is shown in FIG. 3 below, the best λ is **0.26**.

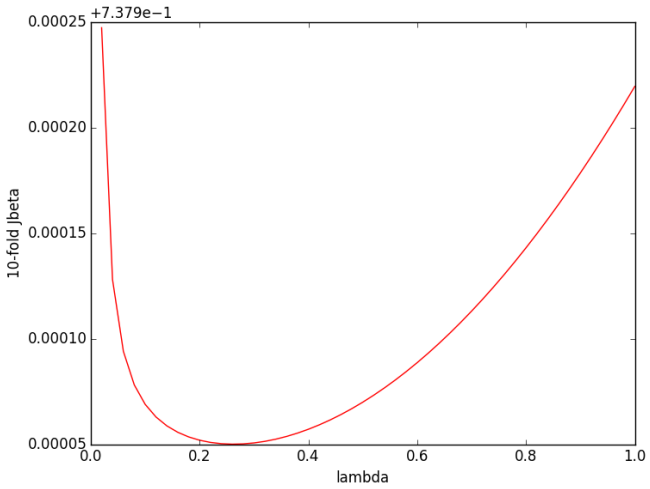

FIG. 3

Using the best λ to do ridge regression again, the data points and learned plane are shown in FIG. 4 below:

The beta is [2.96907258, 0.38675247, 1.26752001]

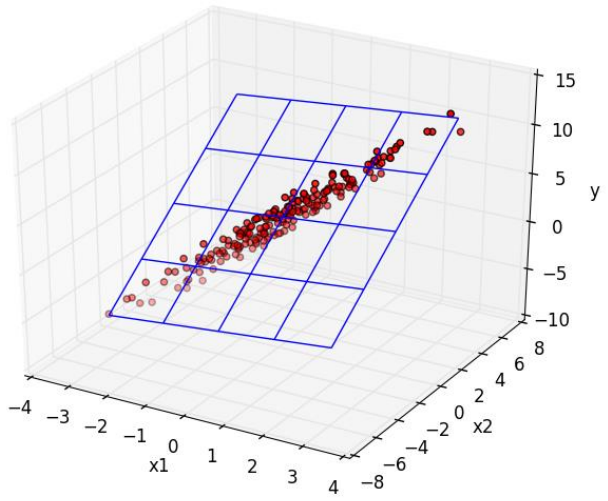

FIG. 4

1.5. The beta from linear regression is [2.9714, -11.0033, 6.9623], while the beta from ridge regression is [2.96907258, 0.38675247, 1.26752001]. Compared with the true coefficient [3, 1, 1], the ridge regression performs better than linear regression does.

The linear regression results of x1 and x2 is shown in FIG. 5. The beta is [0.0002, 2.0001]. We can see that the correlation coefficient between x1 and x2 is very large, where the multicollinearity exists. It leads to the high variance of the regressed results. The ridge regularization can fix this problem.

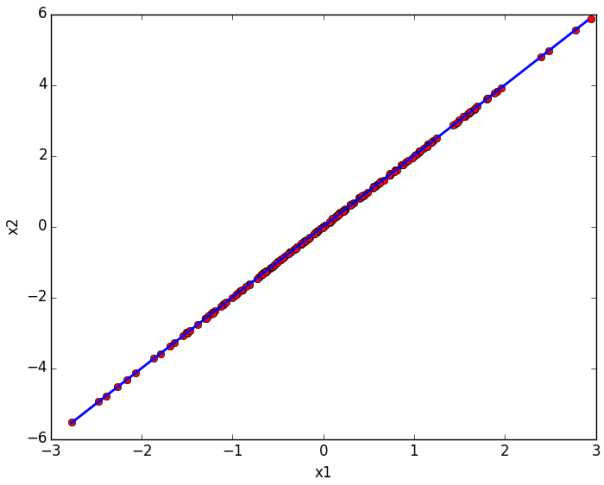

FIG. 5

2.3

Linear SVM		**Train Accuracy**	**Test Accuracy**
C	**gamma**		
0.01	No influence	0.77083	0.77612
0.1		0.81217	0.81291
1		**0.81573**	**0.81426**
10		0.81481	0.81211
100		0.81478	0.81236

Polynomial SVM		**Train Accuracy**	**Test Accuracy**
C	**gamma**		
0.01	0.0	0.75919	0.76377
	0.5	0.82774	0.82759

	1.0	0.84279	0.84307
0.1	0.0	0.76238	0.76734
	0.5	0.84340	0.84399
	1.0	0.84819	0.84903
1.0	0.0	0.81963	0.82022
	0.2	0.84165	0.84104
	0.5	**0.84902**	**0.84952**
	1.0	0.85062	0.84835
10	0.0	0.83600	0.83521
	0.2	0.84755	0.84915
	0.5	0.85034	0.84823
	1.0	0.85148	0.84725

RBF SVM		**Train Accuracy**	**Test Accuracy**
C	**gamma**		
1.0	0.0	0.83256	0.83281
1.0	0.1	0.83545	0.83398
1.0	0.2	0.84122	0.83938
1.0	0.5	0.84871	0.84835
1.0	1.0	0.84970	0.84909
10.0	0.0	0.84534	0.84534
10.0	0.1	0.84684	0.84749
10.0	**0.2**	**0.85040**	**0.84970**
10.0	0.5	0.85224	0.84915
10.0	1.0	0.85710	0.84946
20.0	0.0	0.84684	0.84749
20.0	0.1	0.84884	0.84878
20.0	0.2	0.85053	0.84921
20.0	0.5	0.85295	0.84890
20.0	1.0	0.86075	0.84958

The best performance in several kinds of kernel with the parameters are listed below.

Kernel	**C**	**gamma**	**Train Accuracy**	**Test Accuracy**
Linear	**1**	**0**	**0.81573**	**0.81426**

Polynomial	**1**	**0.5**	**0.84902**	**0.84952**
rbf	**10.0**	**0.2**	**0.85040**	**0.84970**

2.4. **False**. There are no guarantees that the support vectors remain the same. The feature vectors corresponding to polynomial kernels are non-linear functions of the original input vectors and thus the support points for maximum margin separation in the feature space can be quite different.

```
C=  1  gamma=  0.0
trainAccuracy 0.832560425048
testAccuracy 0.83281125238
C=  1  gamma=  0.1
trainAccuracy 0.835447314272
testAccuracy 0.833978256864
C=  1  gamma=  0.2
trainAccuracy 0.841221092718
testAccuracy 0.839383330262
C=  1  gamma=  0.5
trainAccuracy 0.848714720064
testAccuracy 0.848350838401
C=  1  gamma=  1.0
trainAccuracy 0.849697490863
testAccuracy 0.849087893864
C=  10  gamma=  0.0
trainAccuracy 0.845336445441
testAccuracy 0.845341195258
C=  10  gamma=  0.1
trainAccuracy 0.846841313227
testAccuracy 0.84749094036
C=  10  gamma=  0.2
trainAccuracy 0.850403857375
testAccuracy 0.84970210675
C=  10  gamma=  0.5
trainAccuracy 0.852246552624
testAccuracy 0.849149315153
C=  10  gamma=  1.0
trainAccuracy 0.857098983446
testAccuracy 0.849456421596
C=  20.0  gamma=  0.0
trainAccuracy 0.846841313227
testAccuracy 0.84749094036
```