Machine Learning Assignment 1

Xin Li

February 26, 2018

1 Question 1

1.1

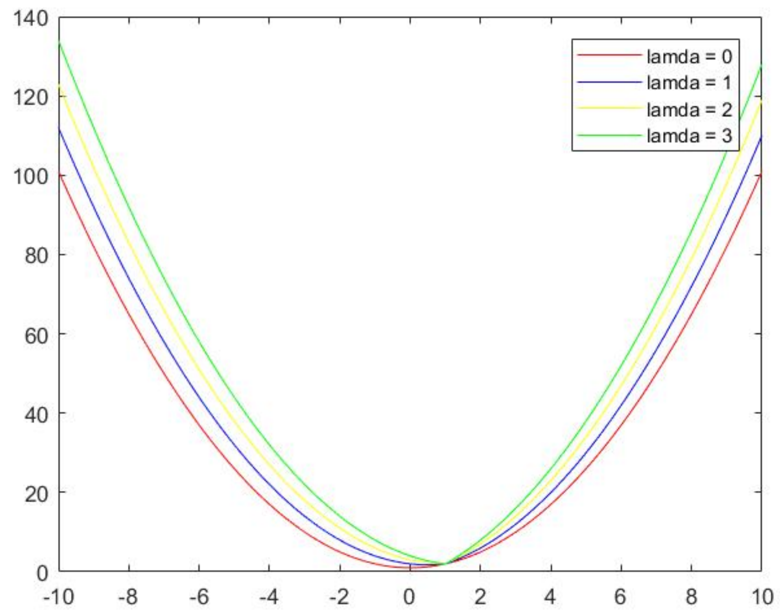

Figure 1: The loss function as a function of r_+ for all $\lambda = (0, 1, 2, 3)$

1.2

The loss function may change to:

$$L(1, r_+) = \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + \lambda * \|1 - r_+\|_1 \tag{1}$$

Thus, for $\lambda = 0$,

$$L(1, r_+) = \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 \tag{2}$$

The derivative: $\frac{dL}{dr_+} = 2r_+$. The minimizer equals to 0. And its minimum value equals 1. When $r_+ = 0$, the derivative equals to 0.

Thus, for $\lambda = 1$,

$$L(1, r_+) = \begin{cases} \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + 1 - r_+ & r_+ <= 1 \\ \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + r_+ - 1 & r_+ > 1 \end{cases} \tag{3}$$

The derivative:

$$\frac{dL}{dr_+} = \begin{cases} 2r_+ - 1 & r_+ <= 1 \\ 2r_+ + 1 & r_+ > 1 \end{cases} \tag{4}$$

The minimizer equals to $\frac{1}{2}$. And its minimum value equals 1.75. When $r_+ = \frac{1}{2}, -\frac{1}{2}$, the derivative equals to 0.

Thus, for $\lambda = 2$,

$$L(1, r_+) = \begin{cases} \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + 2 * (1 - r_+) & r_+ <= 1 \\ \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + 2 * (r_+ - 1) & r_+ > 1 \end{cases} \tag{5}$$

The derivative:

$$\frac{dL}{dr_+} = \begin{cases} 2r_+ - 2 & r_+ <= 1 \\ 2r_+ + 2 & r_+ > 1 \end{cases} \tag{6}$$

The minimizer equals to 1. And its minimum value equals 2. When $r_+ = 1, -1$, the derivative equals to 0 .

Thus, for $\lambda = 3$,

$$L(1, r_+) = \begin{cases} \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + 3 * (1 - r_+) & r_+ <= 1 \\ \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + 3 * (r_+ - 1) & r_+ > 1 \end{cases} \tag{7}$$

The derivative:

$$\frac{dL}{dr_+} = \begin{cases} 2r_+ - 3 & r_+ <= 1 \\ 2r_+ + 3 & r_+ > 1 \end{cases} \tag{8}$$

The minimizer equals to 1. And its minimum value equals 2. When $r_+ = 1.5, -1.5$, the derivative equals to 0 .

2 Question 2

The regularizer actually tries to enforce that the Manhattan distance between two representors should be small enough.
If λ gets larger and larger, the final result will stuck in a certain point. The two solution representors will stay the unchanged and $r_- = r_+$.

3 Question 3

3.1

When $\lambda = 0$, a single contour line for the general function L looks like ellipse. The contour lines look like a set of concentric ellipses. When λ is not equal to zero, a single contour line looks like the concatenation of two ellipses. The contour lines look like two sets of the concentric ellipses' concatenation(two sets have different center.).

3.2

$$L(1, r_+) = \begin{cases} \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + \frac{1}{2} * (-1 - r_-)^2 + \frac{1}{2} * (3 - r_-)^2 + \lambda * (r_- - r_+) & r_+ <= r_- \\ \frac{1}{2} * (-1 - r_+)^2 + \frac{1}{2} * (1 - r_+)^2 + \frac{1}{2} * (-1 - r_-)^2 + \frac{1}{2} * (3 - r_-)^2 + \lambda * (r_+ - r_-) & r_+ > r_- \end{cases} \tag{9}$$

The derivative:

$$\frac{dL}{dr_+} = \begin{cases} (2 - \lambda)r_+ + (2 + \lambda)r_- - 2 & r_+ <= r_- \\ (2 + \lambda)r_+ + (2 - \lambda)r_- - 2 & r_+ > r_- \end{cases} \tag{10}$$

When λ is large enough, if we make the derivative equals 0, $r_- = r_+ = 0.5$. So the exact solution for (r_-, r_+) is $(0.5, 0.5)$.

4 Question 4

4.1

In the CVX toolbox, the loss function is considered as semidefinite-quadratic-linear-program convex problem. And it automatically choose SDPT3, which is an infeasible path-following algorithm to solve this problem.
The main step for every iteration is to compute the search direction(AHO, HKM, NT, GT direction) from the symmetrized Newton equation by computing Schur complement equations. At first, an initial point will be chosen and the relative duality gap will be calculated. In each iteration, a predictor search direction aimed at decreasing the duality gap is computed. After that, a Mehrotra-type corrector step to keep the iterates close to the central path is generated. The solver will stop when the relative duality gap and infeasibility measure are small enough[Tüt03].

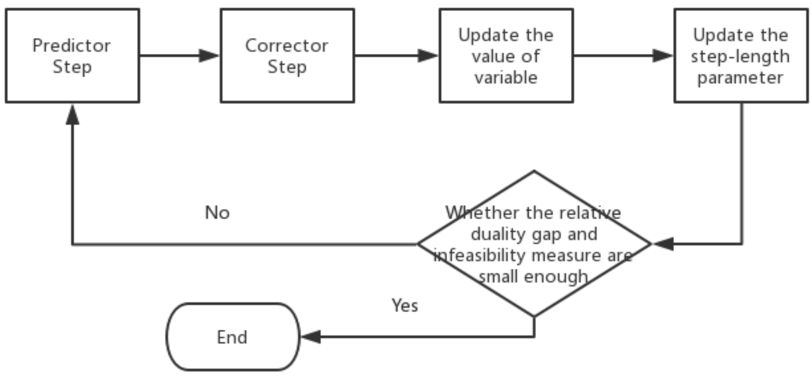

Figure 2: The flow chart for each iteration

4.2

For $\lambda = 0$, For $\lambda = \infty$,

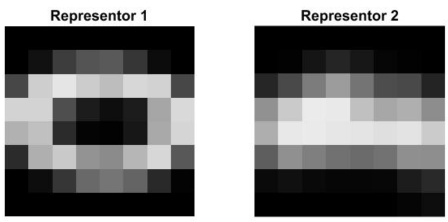

Figure 3: The result images of $\lambda = 0$

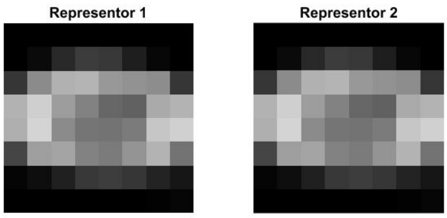

Figure 4: The result images of $\lambda = \infty$

4.3

The experiment has been repeated for 100 times to get the smooth curve.

λ	0	0.1	1	10	100	1000
Apparent Error Rate	0	0	0	0	0	0.4200
True Error Rate	0.0566	0.0641	0.0559	0.0668	0.0600	0.4340

Table 1: The result of error rate

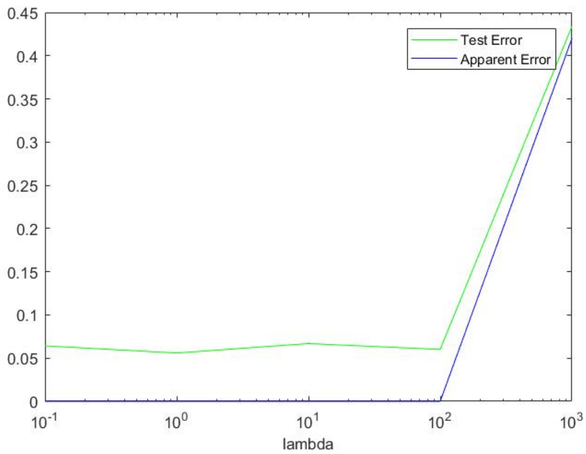

Figure 5: The regularization curves of error rate

References

[Tüt03] K. C.and Todd M. J. Tütüncü, R. H.and Toh. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical Programming*, 95(2):189–217, Feb 2003.