CS680, Spring 2020, Assignment 1

Zhijie Wang, [redacted]

May 17, 2020

Exercise 1

The implementation is attached as cs680-a1.ipynb.

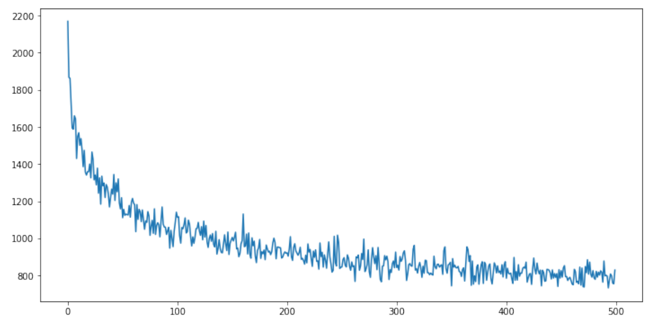

Figure 1: run for 500 passes

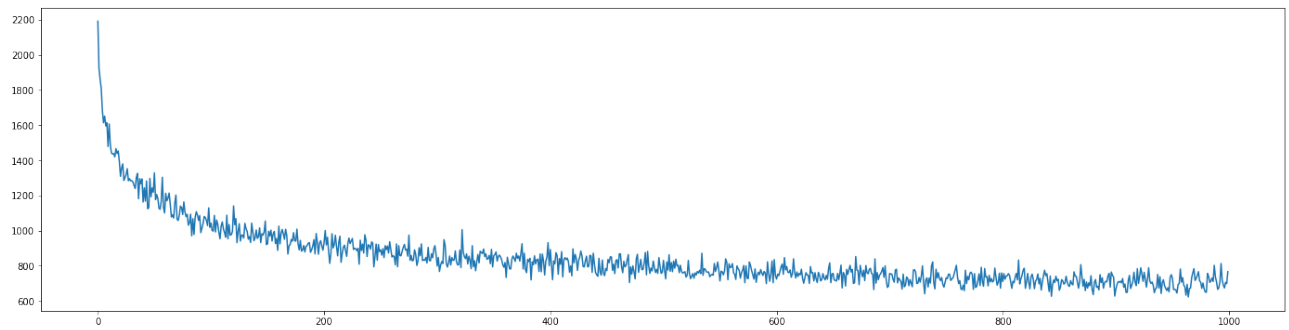

Figure 2: run for 1000 passes

As we can see from the figures, with the increase of passes, the number of mistakes is decreasing, after about 500 passes, the number of mistakes become stationary and didn't have much changes.

Exercise 2

1. **Proof:**

   Recall the expression of ordinary linear regression.

$$\min_{\mathbf{w}} \frac{1}{n} \|X\mathbf{w} - \mathbf{y}\|_F^2 \tag{1}$$

   Here $X\mathbf{w}$ and $\mathbf{y}$ are vectors, therefore (1) is equal to

$$\min_{\mathbf{w}} \frac{1}{n} \|X\mathbf{w} - \mathbf{y}\|_2^2 \tag{2}$$

Let $X \quad \begin{bmatrix} X\mathbf{w} \\ \sqrt{2\lambda} I_{d\times d} \end{bmatrix}$, $\mathbf{y} \quad \begin{bmatrix} \mathbf{y} \\ 0_d \end{bmatrix}$, then

$$
\begin{aligned}
\frac{1}{n}\|X\mathbf{w} - \mathbf{y}\|_2^2 \quad & \frac{1}{n}\left\| \begin{bmatrix} X\mathbf{w} \\ \sqrt{2\lambda} I_{d\times d}\mathbf{w} \end{bmatrix} - \begin{bmatrix} \mathbf{y} \\ 0_d \end{bmatrix} \right\|_2^2 \\
& \frac{1}{n}\left\| \begin{bmatrix} X\mathbf{w} \\ \sqrt{2\lambda}\mathbf{w} \end{bmatrix} - \begin{bmatrix} \mathbf{y} \\ 0_d \end{bmatrix} \right\|_2^2 \\
& \frac{1}{n}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \frac{2\lambda}{n}\|\mathbf{w}\|_2^2 \\
& \frac{2}{n}\left[\frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2\right]
\end{aligned} \tag{3}
$$

Obviously, $\min_{\mathbf{w}} \frac{2}{n}\left[\frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2\right]$ have the same solution as $\min_{\mathbf{w}} \frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2$.

2. In the step (2) and (3), we are adding d new data points. In detail, we are adding a identity matrix with size of $d \times d$ multiplied with scalar λ to X, and adding d 0-label to $\mathbf{y}$, respectively. Therefore, X become a matrix with size of $n \times (n + d)$, $\mathbf{y}$ is a vector with size of $n + d$.

   If the λ approaches to 0, then this question almost become a least-square, $\min_{\mathbf{w}} \frac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2$. If the λ approaches to infinity, the minimization problem will almost rely on $\lambda\|\mathbf{w}\|_2^2$, and we will finally get a model with $\mathbf{w} \quad 0_d$.