Optimization in ML Assignment 1

Kushagra Dar

February 2, 2020

1 Hinge/ SVM Loss

L(w) = $\sum_{i=0}^{n} max\{0, 1 - y_i w^T x_i\}$
$Gradient:$

$$\frac{\partial L(w_i)}{\partial w_i} = \sum_{i=0}^{n} \begin{cases} -y_i x_i & \text{if } y_i w^T x_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

2 Smooth SVM Loss

L(w) = $\sum_{i=0}^{n} max\{0, 1 - y_i w^T x_i\}^2$
$Gradient:$

$$\frac{\partial L(w_i)}{\partial w_i} = \sum_{i=0}^{n} \begin{cases} (1 - y_i x_i)(-y_i) & \text{if } y_i w^T x_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

Hessian :

$$\frac{\partial^2 L(w_i)}{\partial w_i^2} = \sum_{i=0}^{n} \begin{cases} x_i & \text{if } y_i w^T x_i < 1 \\ 0 & \text{otherwise} \end{cases}$$

3 Least square loss function

L(w) = $\sum_{i=0}^{n} (y_i - w^T x_i)^2$
$Gradient:$
$L(w) = (y_i - w^T x_i)^T (y_i - w^T x_i)$

Let g: R$^m \to R$,
$g(v) = v^T v, v \in R^m$
$g'(v) = 2v$

Let h: $\mathrm{R}^n \rightarrow R^m$,

$$h(w) = y - w^T x, w \in R^n$$
$$h'(w) = -x$$

Let f: $\mathrm{R}^n \rightarrow R, (BLAS\ notation)$

$$f(w) = g(h(w)) = (y - w^T x)^T (y - w^T x)$$
$$f'(w) = g'(h(w))h'(w)$$
$$f'(w) = 2(y - w^T x)^T (-x)$$

L(w) = f(w)

$$\frac{\partial L(w_i)}{\partial w_i} = -2X^T (y - w^T X) \qquad \text{(With dimension adjustment)}$$

Hessian :

$$\frac{\partial^2 L(w_i)}{\partial w_i^2} = 2X^T X$$

4 Simple 2 layer function

L(w) = $\sum_{i=0}^{n} (y_i - max(0, w^T x_i + b))^2$

Gradient :

$$\frac{\partial L(w_i)}{\partial w_i} = \sum_{i=0}^{n} \begin{cases} 2(y_i - (w^T x_i + b))(-x_i) & \text{if } w^T x_i + b > 0 \\ 0 & \text{otherwise} \end{cases}$$