SGD logistic regression

Given data $(x_1, y_1) \cdots (x_n, y_n)$

$$\min_w \Big\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} \log(1+e^{-y_i w^T x_i})}_{\text{loss}} + \underbrace{\frac{\lambda}{2}\|w\|^2}_{\text{regularization}} \Big\} = f(w)$$

$$f(w) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\Big(\log(1+e^{-y_i w^T x_i}) + \frac{\lambda}{2}\|w\|^2\Big)}_{f_i(w)}$$

$$\nabla f_i(w) = \begin{bmatrix} \frac{\partial f_i(w)}{\partial w_1} \\ \frac{\partial f_i(w)}{\partial w_2} \\ \vdots \end{bmatrix}$$

$$\frac{\partial f_i(w)}{\partial w_1} = \frac{\partial}{\partial w_1}\Big(\log(1+e^{-y_i w^T x_i}) + \frac{\lambda}{2}\|w\|^2\Big)$$

$$= \frac{-y_i x_{i1} \cdot e^{-y_i w^T x_i}}{1+e^{-w^T x_i y_i}} + \lambda w_1$$

$$= \left(\frac{-y_i e^{-y_i w^T x_i}}{1+e^{-w^T x_i y_i}}\right) \cdot x_{i1} + \lambda w_1$$

$$= \left(\frac{-y_i e^{-y_i w^T x_i}}{1+e^{-w^T x_i y_i}}\right) \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix} + \lambda \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \left(\frac{-y_i e^{-y_i w^T x_i}}{1+e^{-w^T x_i y_i}}\right) \cdot x_i + \lambda \cdot w$$

SGD:

for $t = 1, 2, \cdots$

  Sample $i \in \{1, \cdots, n\}$

  $w \leftarrow w - \eta_t (\nabla f_i(w))$

end.

SVM (with square hinge loss)

$$\min_{w} \frac{1}{n}\sum_{i=1}^{n} \max(1-y_i w^T x_i, 0)^2 + \frac{\lambda}{2}\|w\|^2$$

$$f_i(w) = \underbrace{\max(1-y_i w^T x_i, 0)^2} + \frac{\lambda}{2}\|w\|^2$$

if $1-y_i w^T x_i \geq 0$: $\max(1-y_i w^T x_i, 0)^2$

$= (1-y_i w^T x_i)^2$

$\frac{\partial f_i(w)}{\partial w_1} = -2 \cdot y_i x_{i1} \cdot (1-y_i w^T x_i)$

$$\nabla f_i(w) = (1-y_i w^T x_i)(-2y_i)\begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{bmatrix}$$

$= (\quad 1 \cdot x_i$

if $1-y_i w^T x_i < 0$

$\Rightarrow \max(1-y_i w^T x_i, 0)^2$

$= 0$

$\Rightarrow \nabla f_i(w) = 0$

$$\Rightarrow \nabla f_i(w) = \begin{cases} \text{if } 1-y_i w^T x_i \geq 0, \Rightarrow -2y_i(1-y_i w^T x_i)x_i \\ \text{otherwise} \Rightarrow 0 \end{cases}$$

$+w$

$f_i(w) = loss(w^T x_i)$

$\nabla f_i(w) = \underbrace{loss'(w^T x_i)} \cdot \underbrace{x_i}$