

CS446: Machine Learning, Fall 2018, Homework 1

Name: Aahan Agrawal (agrawal10)

Collaborated with Udit Samani (usamani2), Ankit Gohel (ankitng2)

Problem (1)

Solution:

We will repeatedly use the fact that if $\sigma(z) = \frac{1}{1+\exp(-z)}$ then

$$\frac{\partial}{\partial z} \sigma(z) = \frac{\exp(-z)}{(1 + \exp(-z))^2}$$

(a)

Let $z = w^T x^i$ and compute the gradient:

$$\begin{aligned} \nabla f &= \frac{1}{2n} \sum_{i \in [n]} 2 [y^i - \sigma(w^T x^i)] \cdot \frac{\partial}{\partial z} (-\sigma(z)) \cdot \frac{\partial}{\partial w} (w^T x^{(i)}) \\ &= \frac{1}{2n} \sum_{i \in [n]} 2 [y^i - \sigma(w^T x^i)] \cdot \frac{-\exp(-z)}{(1 + \exp(-z))^2} \cdot (x^{(i)})^T \\ &= \frac{1}{n} \sum_{i \in [n]} [y^i - \sigma(w^T x^i)] \cdot \frac{-\exp(-w^T x^i)}{(1 + \exp(-w^T x^i))^2} \cdot (x^{(i)})^T \end{aligned}$$

Take the transpose to obtain a column vector

$$= \frac{1}{n} \sum_{i \in [n]} [y^i - \sigma(w^T x^i)] \cdot \frac{-\exp(-w^T x^i)}{(1 + \exp(-w^T x^i))^2} \cdot (x^{(i)})$$

The update rule is

$$w_{i+1} = w_i - \alpha \nabla f$$

where α is some step size

(b)

Let $z = y^i w^T x^i$ and compute the gradient:

$$\begin{aligned}
 \nabla f &= \frac{1}{n} \sum_{i \in [n]} \frac{-1}{\sigma(y^i w^T x^i)} \cdot \frac{\partial}{\partial z} \sigma(z) \cdot \frac{\partial}{\partial w} y^i w^T x^i \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{1}{\sigma(y^i w^T x^i)} \cdot \frac{\exp(-z)}{(1 + \exp(-z))^2} \cdot y^i (x^i)^T \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{1}{\sigma(y^i w^T x^i)} \cdot \frac{\exp(-y^i w^T x^i)}{(1 + \exp(-y^i w^T x^i))^2} \cdot y^i (x^i)^T \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{1 + \exp(-y^i w^T x^i)}{1} \cdot \frac{\exp(-y^i w^T x^i)}{(1 + \exp(-y^i w^T x^i))^2} \cdot y^i (x^i)^T \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{\exp(-y^i w^T x^i)}{(1 + \exp(-y^i w^T x^i))} \cdot y^i (x^i)^T \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{1}{(1 + \exp(y^i w^T x^i))} \cdot y^i (x^i)^T
 \end{aligned}$$

Now take the transpose of the RHS so that we express the gradient as a column vector

$$= \frac{-1}{n} \sum_{i \in [n]} \frac{1}{(1 + \exp(y^i w^T x^i))} \cdot y^i (x^i) \quad (\gamma)$$

The update rule is

$$w_{i+1} = w_i - \alpha \nabla f$$

where α is some step size

(c)

Take the Jacobian of (γ) ; the transpose of that result will give us the Hessian:

$$\begin{aligned}
 Hf &= \frac{\partial}{\partial w} \nabla f = \frac{\partial}{\partial w} \frac{-1}{n} \sum_{i \in [n]} \frac{y^i (x^i)}{(1 + \exp(y^i w^T x^i))} \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{y^i (x^i)}{1} \frac{(-1) \exp(y^i w^T x^i) y^i (x^i)^T}{(1 + \exp(y^i w^T x^i))^2} \\
 &= \frac{-1}{n} \sum_{i \in [n]} \frac{y^i (x^i) (-1) \exp(y^i w^T x^i) y^i (x^i)^T}{(1 + \exp(y^i w^T x^i))^2} \\
 &= \frac{1}{n} \sum_{i \in [n]} \frac{y^i (x^i) \exp(y^i w^T x^i)}{(1 + \exp(y^i w^T x^i))^2} \cdot y^i (x^i)^T \\
 &= \frac{1}{n} \sum_{i \in [n]} \frac{\exp(y^i w^T x^i)}{(1 + \exp(y^i w^T x^i))^2} \cdot x^i (x^i)^T
 \end{aligned}$$

It suffices to show that any one summation component is positive semidefinite to show that the entire summation (of matrices) is positive semidefinite. We observe that these component matrices are symmetric, and so we do not take the transpose of any such matrix.

Observe that $\frac{\exp(y^i w^T x^i)}{(1 + \exp(y^i w^T x^i))^2}$ is some constant greater than 0. Let us call this constant α . Thus one matrix component that we wish to prove positive semidefinite is:

$$\alpha(x^i)(x^i)^T$$

Let z be another vector in \mathbb{R}^d . We wish to show that

$$z^T (\alpha(x^i)(x^i)^T) z \geq 0$$

The LHS is, equivalently:

$$= \alpha ((x^i)^T z)^T ((x^i)^T z)$$

which is, in fact at least 0.