cs229 problem set 1

Bhavit Sharma

4/23/23

Table of contents

1 Problem 1

1.1 (a)

Let's compute Hessian of $J(\theta)$ for one training sample. We have

$$J(\theta) = y \log \sigma(\theta^T x) + (1 - y) \log(1 - \sigma(\theta^T x))$$

Now, we compute the first derivate of $J(\theta)$ with respect to θ_i:

$$\frac{\partial J(\theta)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \left[y \log \sigma(\theta^T x) + (1 - y) \log(1 - \sigma(\theta^T x)) \right]$$

We need to use the fact that derivative of $\sigma(\theta^T x)$ is $\frac{\partial}{\partial \theta_i} = \sigma(\theta^T x)(1 - \sigma(\theta^T x))(x[i])$ i.e. the derivative of $\sigma(\theta^T x)$ is $\sigma(\theta^T x)(1 - \sigma(\theta^T x))x$.

Using chain rule, we have

$$\frac{\partial J(\theta)}{\partial \theta_i} = y * (1 - \sigma(\theta^T x)) * x[i] + (1 - y) * (-\sigma(\theta^T x)) * x[i]$$

Simplifying, we have

$$\frac{\partial J(\theta)}{\partial \theta_i} = (y - \sigma(\theta^T x)) * x[i]$$

So for n training samples, we have

$$\frac{\partial J(\theta)}{\partial \theta_i} = \sum_{j=1}^{n} (y_j - \sigma(\theta^T x_j)) * x_{ij}$$

Writing this in vector form, we have

$$\frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{n} \sum_{j=1}^{n} (y_j - \sigma(\theta^T x_j)) * x_j$$

Now let us compute the Hessian of $J(\theta)$ with respect to θ_i and θ_j: We know that the derivate with respect to j is

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{n} \sum_{j=1}^{n} (y_j - \sigma(\theta^T x_j)) * x_{ij}$$

So H_{ij} is

$$\begin{aligned} H_{ij} = \frac{\partial^2 J(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{\partial}{\partial \theta_i} \left[-\frac{1}{n} \sum_{k=1}^{n} (y_k - \sigma(\theta^T x_k)) * x_{kj} \right] \\ &= -\frac{1}{n} \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} (y_k - \sigma(\theta^T x_k)) * x_{kj} \\ &= -\frac{1}{n} \sum_{k=1}^{n} (\sigma(\theta^T x_k)) * (\sigma(\theta^T x_k) - 1) * x_{ki} x_{kj} \end{aligned}$$

Writing it in matrix form, we have

$$H = \frac{1}{n} \sum_{k=1}^{n} (\sigma(\theta^T x_k)) * (1 - \sigma(\theta^T x_k)) * x_k x_k^T$$

Now we want to show that the Hessian is positive semi-definite which implies that J has a local minima and it's a convex function The way it's done is by showing that for any vector v, we have

$$v^T H v \geq 0$$

Note: TODO(Bhavit): Why is this true?

$$v^T H v = \frac{1}{n} \sum_{k=1}^{n} (\sigma(\theta^T x_k)) * (1 - \sigma(\theta^T x_k)) v^T x_k x_k^T v$$

Now we can see that $V^T x_k x_k^T v$ can be written as

$$v^T x x^T v = \sum_{i=1}^{d} \sum_{j=1}^{d} v[i]x[i]x[j]v[j]$$

where d is the dimension of x. Try to write this in matrix form and you can see. Now using the hint, we can easily see that the above form is equivalent to $v^T x x^T v = (v^T x)(v^T x) > 0$.

Since $\sigma(\theta^T x_k) \in [0, 1]$, we have $H \geq 0$ always.

1.2 (b)