Q1a

iv

$$||\nabla f(u) - \nabla f(v)|| \leq L||u - v||$$

Equivalently

$$||\nabla^2 f|| \leq L$$

$$l(w) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i \cdot x_i^T w)) + \frac{\lambda}{2}||w||^2$$

We already found the second derivative (Hessian matrix),

$$\nabla^2 l = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{y_i x_i}{1 + e^{y_i \cdot x_i^T w}}\right)^2 e^{y_i \cdot x_i^T w} + \lambda$$

Therefore,

$$||\nabla^2 l|| \leq L$$

$$\left|\left|\frac{1}{n}\sum_{i=1}^{n} \left(\frac{y_i x_i}{1 + e^{y_i \cdot x_i^T w}}\right)^2 e^{y_i \cdot x_i^T w} + \lambda\right|\right| \leq L$$

By triangle inequality,

$$\frac{1}{n}\sum_{i=1}^{n} \left|\left|\left(\frac{y_i x_i}{1 + e^{y_i \cdot x_i^T w}}\right)^2 e^{y_i \cdot x_i^T w}\right|\right| + ||\lambda|| \leq L$$

Note that, taking the extreme case,

$$\left|\left|\frac{e^{y_i \cdot x_i^T w}}{(1 + e^{y_i \cdot x_i^T w})^2}\right|\right| \leq \frac{1}{(1+1)^2} = \frac{1}{4}$$

$$\implies \frac{1}{n}\sum_{i=1}^{n} \frac{1}{4}||y_i x_i||^2 + ||\lambda|| \leq L$$

Recall that $y_i \in \{-1, 1\}$. So $||y_i||^2 = 1$. Then,

$$\frac{1}{4n}\sum_{i=1}^{n} ||x_i||^2 + \lambda \leq L$$

$$\boxed{L = \frac{1}{4n}\sum_{i=1}^{n} ||x_i||^2 + \lambda}$$

It is L-smooth for such a value.

Q1c

$$w_{t+1} = w_t - \alpha \nabla l(w_t)$$

$$\boxed{w_{t+1} = w_t + \frac{\alpha}{n}\left(\sum_{i=1}^{n} \frac{y_i x_i}{1 + e^{y_i \cdot x_i^T w_t}} + \lambda w_t\right)}$$

Q1d

For a single (batch size of 1) sample j uniformly sampled from $\{1, \ldots, n\}$,

$$w = w - \alpha \nabla l(w, x_j, y_j)$$

$$\boxed{w_{t+1} = w_t + \frac{\alpha}{n} \frac{y_j x_j}{1 + e^{y_j \cdot x_j^T w_t}} + \lambda w_t}$$

Q1e

In lecture, we proved that for gradient descent to converge,
$1 \geq \alpha L$ must be satisfied. Substituting our values for L and $\alpha = \frac{1}{4\lambda}$,

$$1 \geq \left(\frac{||X||^2}{4n} + \lambda\right)\frac{1}{4\lambda}$$

$$1 \geq \left(\frac{||X||^2}{4n} + \lambda\right)\frac{1}{4\lambda}$$

Yes, this will converge.

FIX. THIS SEEMS VERY INCORRECT

Q1f

No
Stochastic gradient descent, especially for batch sizes of 1, requires finer step sizes. We can consider a case of

Q2

Throughout, let $\mathbb{A}$ be the calculation accumulator, simply indicating that a portion of the work has been computed.

2a

$$h_w(x) = \text{sign}(x^T w)$$

Both $x, w \in \mathbb{R}^d$. We can compute the dot product as multiplying element wise and then summing. This requires d multiplications and $d-1$ pairwise sums, arriving at

$$h_w(x) = \text{sign}(\mathbb{A})$$

$$\text{Op Count: } d + (d-1) = 2d - 1$$

Taking the sign of the resultant scalar takes 1 operation,

$$h_w(x) = \mathbb{A}$$

$$\text{Op Count: } 2d - 1 + 1 = 2d$$

The calculation is complete with $\boxed{\text{num operations} = 2d}$

2b

$$l(w) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i \cdot x_i^T w)) + \frac{\lambda}{2}||w||^2$$

We can consider the $||\cdot||^2$ as a dot product $w \cdot w$. Since $w \in \mathbb{R}^d$, by above, this requires $2d-1$ operations. We then scale the resultant scalar by λ and then by $\frac{1}{2}$. This involves a further $1+1$ multiplications.

$$l(w) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i \cdot x_i^T w)) + \mathbb{A}$$

$$\text{Op Count: } 2d - 1 + 1 + 1 = 2d + 1$$

From above, $-y_i x_i^T w$ consists first of a dot product of d-dimensional vectors (again, $2d-1$ multiplications). Then, the resultant scalar is scaled by y_i and by -1 (2). Then, for a scalar, exponentiation, incrementing, and taking the log all involve one operation (3),

$$l(w) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{A} + \mathbb{A}$$

$$\text{Op Count:} (2d+1) + (2d-1) + (1+1) + (1+1+1) = 4d + 5$$

We compute each term n times, then sum all n terms for $n-1$ pairwise sums. Per above, computing each term costs $2d+4$ operations. We then divide the resultant scalar by n, invoking another operation.

$$l(w) = \mathbb{A} + \mathbb{A}$$

$$\text{Op Count: } 2d + 1 + (2d+4)n + (n-1) + 1 = 2d + 1 + (2d+5)n$$

Adding the two remaining scalar terms,

$$l(w) = \mathbb{A}$$

$$\text{Op Count: } 2d + 1 + (2d + 5)n + 1 = (2d + 5)n + 2d + 2$$

The computation is complete with $\text{num operations} = \boxed{(2d + 5)n + 2d + 2}$