

Problem Set 2, Oct 2, 2018 (Solutions to Theory Questions)

1 MAE Subgradient (Exercise 6)

The subgradient for the function $h : \mathbb{R} \rightarrow \mathbb{R}, h(e) := |e|$ is given as

$$g : \mathbb{R} \rightarrow \{-1, 0, 1\}, g(e) := \text{sign}(e).$$

The MAE cost function is defined as $\mathcal{L}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^N |y_n - \mathbf{x}_n^\top \mathbf{w}|$.

As given in the annotated lecture notes 2, we can use the **chain-rule for subgradients**, for $\mathcal{L}(\mathbf{w}) := h(q(\mathbf{w}))$, when the outer function h is not differentiable, but q is differentiable. We write $\partial h(\mathbf{y})$ for the set of all subgradients of h at \mathbf{y} . Then any vector \mathbf{g} of the following form is a subgradient of \mathcal{L} at \mathbf{w} :

$$\mathbf{g} \in \partial h(q(\mathbf{w})) \cdot \nabla q(\mathbf{w})$$

where we can pick any element of the left, and multiply with the vector on the right (the gradient).

We now find the (sub)gradient update for a single component w_i and conclude by generalizing to the whole vector \mathbf{w} . In our case here, ∂h is the sign function. Then for $\frac{\partial \mathcal{L}(\mathbf{w})}{\partial w_i} = \frac{1}{N} \sum_{n=1}^N \frac{\partial |y_n - \mathbf{x}_n^\top \mathbf{w}|}{\partial w_i}$ we have that:

$$\frac{\partial \mathcal{L}(\mathbf{w})}{\partial w_i} = \frac{1}{N} \sum_{n=1}^N -(x_n)_i \text{sign}(y_n - \mathbf{x}_n^\top \mathbf{w}).$$

Finally we can conclude that $\frac{-1}{N} \mathbf{X}^\top \cdot \text{sign}(\mathbf{e})$ is a subgradient to \mathcal{L} at \mathbf{w} , where $\mathbf{e} := \mathbf{y} - \mathbf{X} \cdot \mathbf{w}$ and sign applied element-wise to \mathbf{e} , and \mathbf{X} is the matrix collecting all datapoints as its rows.