WEEKLY PARTICIPATION 6

Roy decides he would like to fit a model of the form $y = \mathbf{x}^T\boldsymbol{\omega}$ using the ℓ_1 loss instead of ℓ_2 loss, so obtains the ERM objective

$$f(\boldsymbol{\omega}) = \frac{1}{n}\|\mathbf{X}\boldsymbol{\omega} - \mathbf{y}\|_1$$

for learning the model $\boldsymbol{\omega}$ from his training data. This is a non-differentiable convex function, so he knows he can use subgradient descent to find an optimizer.

Which of the following vectors are in $\partial f(\boldsymbol{\omega})$?

(A) $\frac{1}{n}\mathbf{X}^T\mathbf{z}$ for a $\mathbf{z} \in \partial\|\mathbf{X}\boldsymbol{\omega} - \mathbf{y}\|_1$.

(B) $\frac{1}{n}\mathbf{z}$ for a $\mathbf{z}$ satisfying $\|\mathbf{z}\|_\infty \leq 1$ and $\mathbf{z}^T(\mathbf{X}\boldsymbol{\omega} - \mathbf{y}) = \|\mathbf{X}\boldsymbol{\omega} - \mathbf{y}\|_1$.

(C) $\frac{1}{n}\mathbf{z}$ for a $\mathbf{z} \in \partial\|\mathbf{X}\boldsymbol{\omega} - \mathbf{y}\|_1$.

(D) $\frac{1}{n}\mathbf{z}$ for a $\mathbf{z}$ satisfying $\|\mathbf{z}\|_\infty \leq 1$ and $\mathbf{z}^T\boldsymbol{\omega} = \|\boldsymbol{\omega}\|_1$.

State which of options (A)–(D) give valid subgradients.