WEEKLY PARTICIPATION 1

Recall that we setup the problem of learning a linear separator $y = \text{sign}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle)$ for binary classification as follows (SVM):

$$\boldsymbol{w}_\star = \arg\min_{\boldsymbol{w}} \frac{1}{n} \sum_{i=1}^{n} \phi(y^{(i)} \langle \boldsymbol{w}, \boldsymbol{x}^{(i)} \rangle) + \frac{\lambda}{2} \|\boldsymbol{w}\|_2^2,$$

to achieve the following goals:

- Encourage the prediction to have the correct sign on all the training data, and be confident, $y^{(i)} \langle \boldsymbol{w}, \boldsymbol{x}^{(i)} \rangle \geq 1$. Recall that the specific value of 1 is not important, just that it is encouraged that $y^{(i)} \langle \boldsymbol{w}, \boldsymbol{x}^{(i)} \rangle$ be bounded away from zero for all the training data.
- Encourage the model to be stable, i.e. $\|\boldsymbol{w}\|_2$ to be small.

We chose ϕ to be the hinge loss $\phi(t) = \max\{1 - t, 0\}$ because it encourages the first point to be true. Which of the following choices similarly encourage the first property?

(A) $\phi(t) = \log(1 + \exp(-t))$

(B) $\phi(t) = \exp(-t)$

(C) $\phi(t) = 1 - \tanh(t)$

Answer by listing none, one, two, or all of these options.