Surrogate losses $\ell : \{-1,1\} \times \mathbb{R} \to \mathbb{R}$ are convex upper bounds on the zero-one loss function for binary classification. We already encountered two of them:

- **Hinge loss** $\ell(y, \hat{y}) = [1 - y \hat{y}]_+$
- **Boosting loss** $\ell(y, \hat{y}) = e^{-y \hat{y}}$

where $y \in \{-1,1\}$ and $\hat{y} \in \mathbb{R}$.

As many surrogate losses exist, we may wonder whether some should be preferred over the others. We now define an important criterion, called *consistency*, that a surrogate loss may satisfy with respect to the function $\eta(x) = \mathbb{P}(Y = 1 \mid X = x)$ which defines the Bayes optimal predictor $f^*$.

A surrogate loss function $\ell$ is consistent if, for all $x \in X$,

$$\text{sgn}(y^*_x) = f^*(x) \quad \text{for} \quad y^*_x = \arg\min_{\hat{y} \in \mathbb{R}} \mathbb{E}[\ell(Y, \hat{y}) \mid X = x]$$

In other words, the sign of the prediction minimizing the conditional risk with respect to the surrogate loss must be equal to the Bayes optimal classification for the zero-one loss.

We now verify the consistency of the hinge loss. We have

$$y^*_x = \arg\min_{\hat{y} \in \mathbb{R}} \left( \eta(x) [1 - \hat{y}]_+ + (1 - \eta(x)) [1 + \hat{y}]_+ \right)$$

$$= \arg\min_{\hat{y} \in [-1,+1]} \left( \eta(x) [1 - \hat{y}]_+ + (1 - \eta(x)) [1 + \hat{y}]_+ \right)$$

$$= \arg\min_{\hat{y} \in [-1,+1]} \left( 1 + (1 - 2\eta(x)) \hat{y} \right)$$

$$= \begin{cases} 
-1 & \text{if } \eta(x) \leq 1/2, \\
+1 & \text{otherwise}
\end{cases}$$

$$= f^*(x)$$

In the second inequality, we could replace $\hat{y} \in \mathbb{R}$ with $\hat{y} \in [-1,+1]$ because both functions $[1 - \hat{y}]_+$ and $[1 + \hat{y}]_+$ increase or remain constant outside of the interval $[-1,+1]$.

More generally, the following result holds.

**Theorem 1.** If a surrogate loss $\ell : \{-1,1\} \times \mathbb{R} \to \mathbb{R}$ is such that for all $y \in \{-1,1\}$ the function $\ell(y, \cdot)$ is convex, differentiable at zero, and satisfies $\ell'(y,0) < 0$, then $\ell$ is consistent.

Besides the hinge loss, the boosting loss, the square loss $\ell(y, \hat{y}) = (1 - y \hat{y})^2$ and the quadratic hinge loss $\ell(y, \hat{y}) = ([1 - y \hat{y}]_+)^2$ are also consistent.