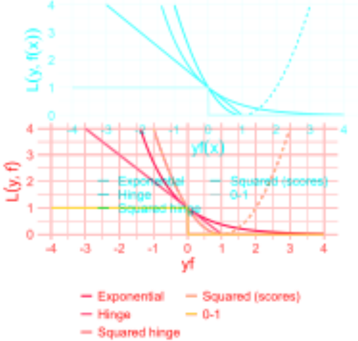
Introduction to Machine Learning

Advanced Risk Minimization

Advanced Classification Losses



Learning goals

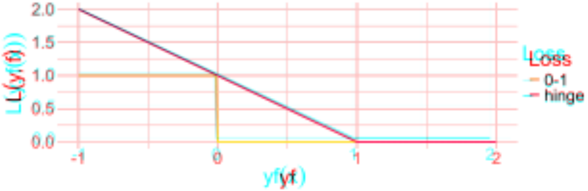
- Know the (squared) hinge loss
- Know the $L2$ loss defined on scores
- Know the exponential loss
- Know the AUC loss

HINGE LOSS

- The intuitive appeal of the 0-1-loss is set off by its analytical properties ill-suited to direct optimization.
- The **hinge loss** is a continuous relaxation that acts as a convex upper bound on the 0-1-loss (for $y \in \{-1, +1\}$):

$$L(y, f(\mathbf{x})) = \max\{0, 1 - yf(\mathbf{x})\}.$$

- Note that the hinge loss only equals zero for a margin ≥ 1, encouraging confident (correct) predictions.
- It resembles a door hinge, hence the name:

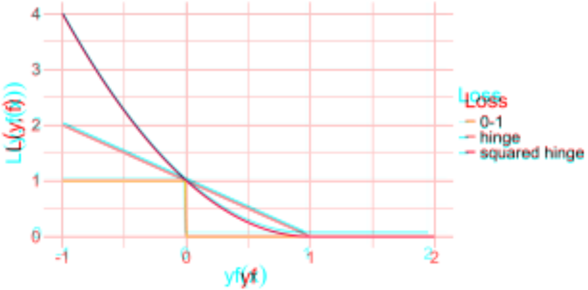


SQUARED HINGE LOSS

- We can also specify a **squared** version for the hinge loss:

$$L(y, f) = \max\{0, (1 - yf)\}^2.$$

- The $L2$ form punishes margins $yf \in (0, 1)$ less severely but puts a high penalty on more confidently wrong predictions.
- Therefore, it is smoother yet more outlier-sensitive than the non-squared hinge loss.

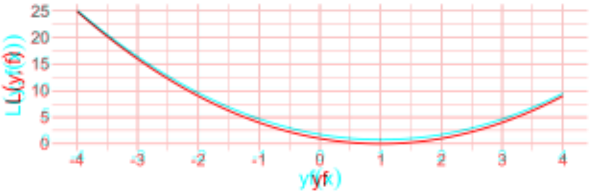


SQUARED LOSS ON SCORES

- Analogous to the Brier score defined on probabilities we can specify a **squared loss on classification scores** (again, $y \in \{-1, +1\}$, using that $y^2 \equiv 1$):

$$L(y, f) = (y - f)^2 = y^2 - 2yf + f^2 =$$
$$= 1 - 2yf + (yf)^2 = (1 - yf)^2$$

- This loss behaves just like the squared hinge loss for $yf < 1$, but is zero only for $yf = 1$ and actually increases again for larger margins (which is in general not desirable!)



CLASSIFICATION LOSSES: EXPONENTIAL LOSS

Another possible choice for a (binary) loss function that is a smooth approximation to the 0-1-loss is the **exponential loss**:

- $L(y, f) = \exp(-yf)$, used in AdaBoost.
- Convex, differentiable (thus easier to optimize than 0-1-loss).
- The loss increases exponentially for wrong predictions with high confidence; if the prediction is right with a small confidence only, there, loss is still positive.
- No closed-form analytic solution to (empirical) risk minimization.