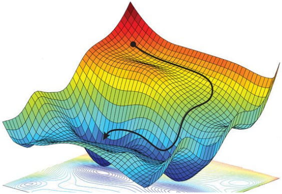
Introduction to Machine Learning

Advanced Classification Losses



Learning goals

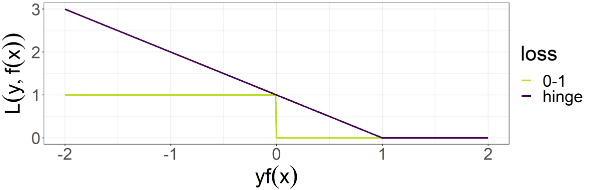
- Know the exponential loss
- Know the AUC loss

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:

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

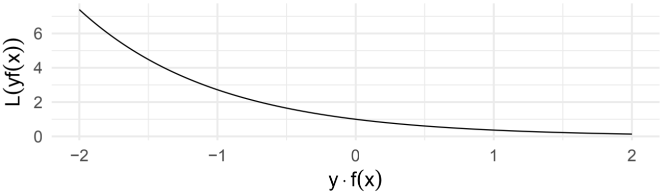
- Note that the hinge loss only equals zero for a margin ≥ 1, encouraging confident (correct) predictions.
- It resembles a door hinge, hence the name:



CLASSIFICATION LOSSES: EXPONENTIAL LOSS

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

- $L(y, f(\mathbf{x})) = \exp(-yf(\mathbf{x}))$, 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



CLASSIFICATION LOSSES: AUC-LOSS

- Often AUC is used as an evaluation criterion for binary classifiers
- Let $Y \in \{-1, 1\}$ with observations n_{-1} number of negative and n_1 of positive samples
- The AUC can then be defined as

$$AUC = n_{-1}^{-1} n_1^{-1} \sum_{i:y_i=1} \sum_{j:y_j=-1} I(f_i > f_j)$$

- This is not differentiable wrt f due to $I(f_i > f_j)$
- But the indicator function can be approximated by the distribution function of the triangular distribution on $[-1, 1]$ with mean 0
- However, direct optimization of the AUC is usually not as good as optimization wrt a common loss and tuning via AUC in practice