Loss Functions for Binary Class Probability Estimation

Thomas Strohmann and Greg Grudic

- Test data: $y_1, y_2, ..., y_n \in \{0,1\}$
- Estimated Probabilities: $p_1, p_2, ..., p_n \in [0,1]$

There are the following metrics for evaluating the estimated probabilities:

- Brier Score / Squared Loss: [1] [2]
  $Loss_{BS} = n^{-1} \sum_{i=1}^{n} (y_i - p_i)^2 = n^{-1} \left(\sum_{i=1}^{n} y_i (1-p_i)^2 + (1-y_i) p_i^2 \right)$
- Log loss: [1]
  $Loss_{Log} = n^{-1} \left(\sum_{i=1}^{n} -y_i log(p_i) - (1-y_i) log(1-p_i) \right)$
- Exponential loss / Boosting loss: [1]
  $Loss_{Exp} = n^{-1} \left(\sum_{i=1}^{n} y_i \sqrt{\frac{1-p_i}{p_i}} + (1-y_i) \sqrt{\frac{p_i}{1-p_i}} \right)$
- Calibration loss: [2] Suppose we estimate k different probabilities p_j for $j = 1, ..., k$. Let n_j be the number of predictions with value p_j and r_j the fraction of corresponding y's that are 1. (formally $r_j = n_j^{-1} \sum_{i:p_i = p_j} y_i$)
  $Loss_C = n^{-1} \sum_{j=1}^{k} n_j (r_j - p_j)^2$
- Refinement loss: [2]
  $Loss_R = n^{-1} \sum_{j=1}^{k} n_j \, r_j (1 - r_j)$
- Separation Theorem: [2] $Loss_{BS} = Loss_C + Loss_R$
- Lift charts: [3] Sort probabilities in decreasing order: $p_{s_1} \geq p_{s_2} \geq ... \geq p_{s_n}$
  Apply that sorting to the y values.
  Define lift function $l(k/n) = \frac{k^{-1} \sum_{i=1}^{k} y_{s_i}}{\bar{y}} \quad k = 1, ..., n; \;\; \bar{y} = n^{-1} \sum_{i=1}^{n} y_i$
  (Notes: lift function is step function with step size $1/n$ and $l(1) = 1$)
  Maximize area under lift function. Or equivalently: $Loss_{Lift} = -\int_0^1 l(a) da$

References

[1] A. Buja, W. Stuetzle, and Y. Shen. Degrees of boosting a study of classification loss functions. *International Conference on Robust Statistics*, 2003.

[2] G. Blattenberger and F. Lad. Separating the brier score into calibration and refinement components: A graphical exposition. *The American Statistician*, 39(1):26–32, February 1985.

[3] G. Piatetsy-Shapiro and B. Masand. Estimating campaign benefits and modeling lift. *Proceedings of the Fifth International Conference on Knowledge Dicovery and Data Mining*, 1999.