Multi-Class Loss Functions

Adrian Barbu

Multi-Class Setup

- L≥2 possible classes
- Training set

$$D = \{(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \{1, ..., L\}, i = \overline{1, N}\}$$

- Linear model with parameters W∈ M$_{d \times L}$

$$\boldsymbol{u} = (u_1, ..., u_L) = \boldsymbol{x}^T W$$

- Loss function

$$L_D(W) = \sum_{i=1}^{N} L(\boldsymbol{x}_i^T W, y_i) + \lambda \|W\|_F^2$$

Multi-Class Loss Functions

- **Loss functions**
  - **Logistic (softmax) loss**

  $$L(\boldsymbol{u}, y) = -\ln \frac{e^{u_y}}{\sum_{k=1}^{L} e^{u_k}}$$

  - **Vapnik's loss**

  $$L(\boldsymbol{u}, y) = \sum_{k \neq y} \ell(u_y - u_k)$$

  - **Crammer's loss**

  $$L(\boldsymbol{u}, y) = \ell(u_y - \max_{k \neq y} u_k)$$

  - **Lee's loss**

  $$L(\boldsymbol{u}, y) = \sum_{k \neq y} \ell(-u_k)$$

Classification Loss Functions

- Logistic loss:

$$\ell(x) = \log(1 + e^{-x})$$

- Huberized SVM loss:

$$\ell(x) = \begin{cases} 0 & \text{if } x > 1 + h \\ \dfrac{(1 + h - x)^2}{4h} & \text{if } |1 - x| \leq h \\ 1 - x & \text{if } x < 1 - h \end{cases}$$
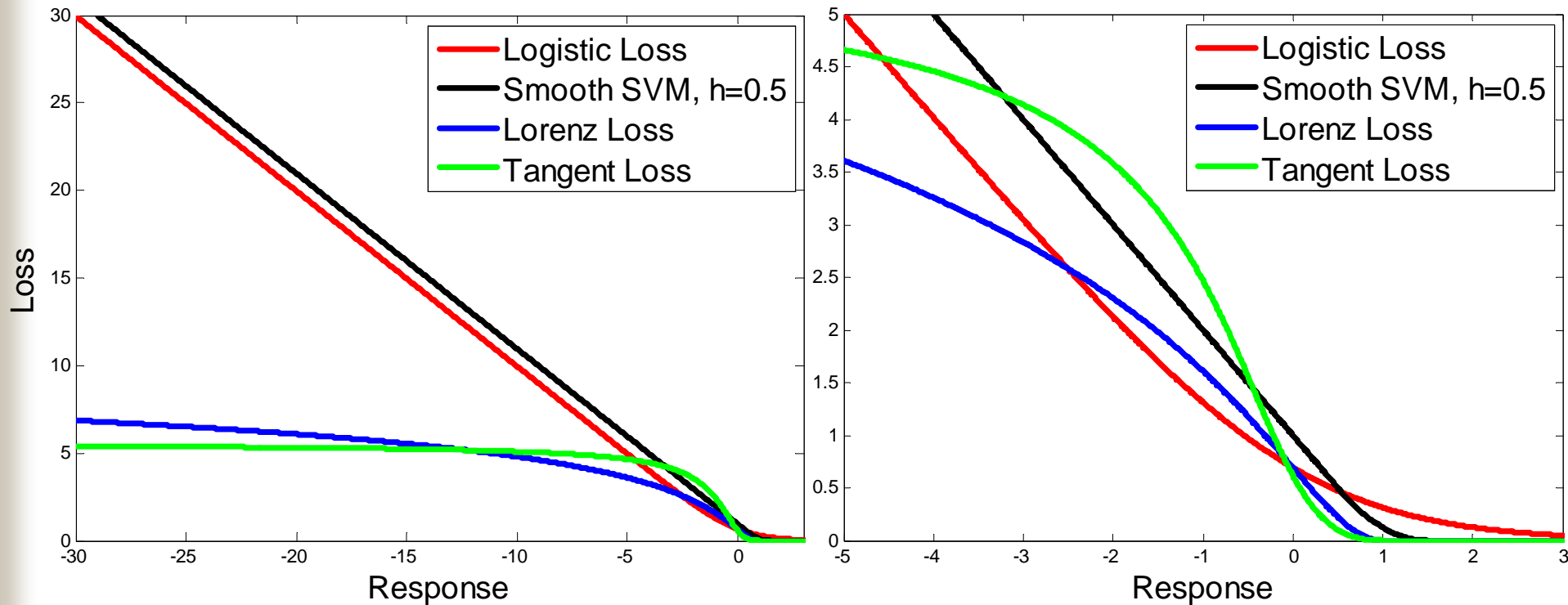
- Lorenz loss

$$\ell(x) = \begin{cases} 0 & \text{if } x > 1 \\ \ln(1 + (x - 1)^2) & \text{else} \end{cases}$$

Classification Loss Functions

- Tangent loss

$$\ell(x) = \begin{cases} 0 & \text{if } x > 1 \\ (\text{atan}(x) - \pi/4)^2 & \text{else.} \end{cases}$$

References

- Barbu, A., She, Y., Ding, L., Gramajo, G.: Feature selection with annealing for computer vision and big data learning. IEEE Trans. PAMI 2016
- A. Barbu, L. Lu, H. Roth, A. Seff, R.M. Summers. An Analysis of Robust Cost Functions for CNN in Computer-Aided Diagnosis. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2016.
- Chapelle, O.: Training a support vector machine in the primal. Neural Computation 19(5), 1155-1178 (2007)
- Crammer, K., Singer, Y.: On the algorithmic implementation of multiclass kernel-based vector machines. The Journal of Machine Learning Research 2, 265-292(2002)
- Lee, Y., Lin, Y., Wahba, G.: Multicategory support vector machines: Theory andapplication to the classication of microarray data and satellite radiance data. Journal of the American Statistical Association 99(465), 67-81 (2004)
- Masnadi-Shirazi, H., Mahadevan, V., Vasconcelos, N.: On the design of robust classiers for computer vision. In: CVPR. pp. 779-786 (2010)
- Vapnik, V.N., Vapnik, V.: Statistical learning theory, vol. 1. Wiley New York (1998)