Clip-stretch of labels for regularizing logistic regression

Manfred K. Warmuth

University of California - Santa Cruz

CMPS 242 - Lecture 8
(developed with lots of help from Dima Kuzmin)
Data: \((x_t, y_t)\), where \(y_t \in [0, 1]\)

Linear transformation of interval \([0,1]\) to subinterval \([a, b]\)

\[ y'_t := a + (b - a) y_t \]

i.e. \(0 \rightarrow a\) and \(1 \rightarrow b\)

Minimize logistic loss on training data \((x_t, y'_t)\)
Logistic loss when label $1 - > b$

\[ \sigma := x \mapsto \frac{e^x}{1+e^x} \]

\[ \text{loss} := (y, ah) \rightarrow y \ln \left( \frac{y}{\sigma(ah)} \right) + (1-y) \ln \left( \frac{1-y}{1-\sigma(ah)} \right) \]

```
> sigma := x -> exp(x) / (1+exp(x));

> loss := (y, ah) -> y*ln(y/sigma(ah)) + (1-y)*ln((1-y)/(1-sigma(ah)));

> plot([sigma(ah), .99999, loss(.99999, ah), .95, loss(.95, ah), .6, loss(.6, ah), .4,
```
Stretching the prediction

- A) Predict on instance $\mathbf{x}$ with $\hat{y} = \sigma(\mathbf{w} \cdot \mathbf{x})$
  where $\mathbf{w}$ is weight vector produced in training

- B) Predict with

\[
\hat{y} = \begin{cases} 
1 & \text{if } \sigma(\mathbf{w} \cdot \mathbf{x}) > b \\
0 & \text{if } \sigma(\mathbf{w} \cdot \mathbf{x}) < a \\
\frac{\sigma(\mathbf{w} \cdot \mathbf{x}) - a}{b - a} & \text{otherwise}
\end{cases}
\]

i.e. $\sigma(\mathbf{w} \cdot \mathbf{x}) \leq a$ mapped to 0 and $\sigma(\mathbf{w} \cdot \mathbf{x}) \geq b$ mapped to 1

- When there are lots of sparse features then
  B) prevents overfitting: worse average logistic loss on training set
  - better average logistic loss on test set
  A) does not prevent overfitting