Loss-regularization family of classifiers

Note Title 14-08-2012

$K = 2$

$y_i \in \{+1, -1\}$

$f(x) = W \cdot x + w_0$ ← bias

$w, x \in \mathbb{R}^d$.

Classifiers
- Decision trees
- Probabilistic
-

Classifier: $\text{sign}(f(x))$

$$\text{Error} = 1 \quad \text{if } yf(x) < 0$$
$$= 0 \quad \text{otherwise}$$

Given training data

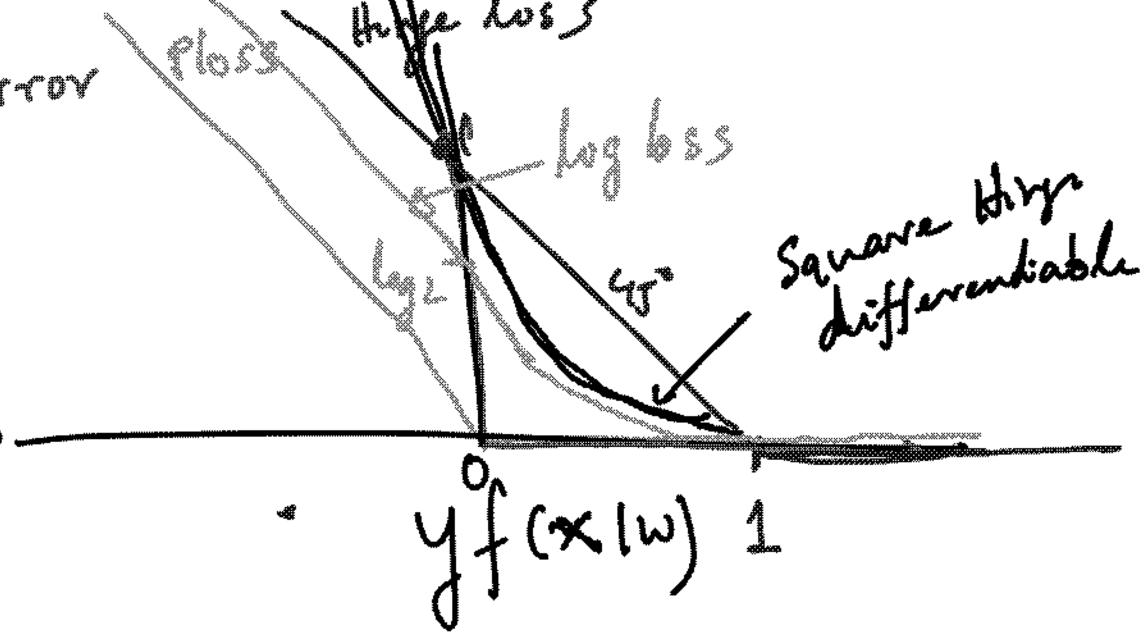
$$D = \{(x^i, y^i) : i = 1 \cdots N\}$$

Ideal value for W^{OPT}

$$\operatorname{argmin} \sum_{i=1}^{N} \mathrm{error}(y^i f(x^i | w))$$

$$= \operatorname{argmin} \sum_{i=1}^{N} \underline{\mathrm{error}}\,(y^i (w \cdot x^i + w_0))$$

Loss function ≥ 0



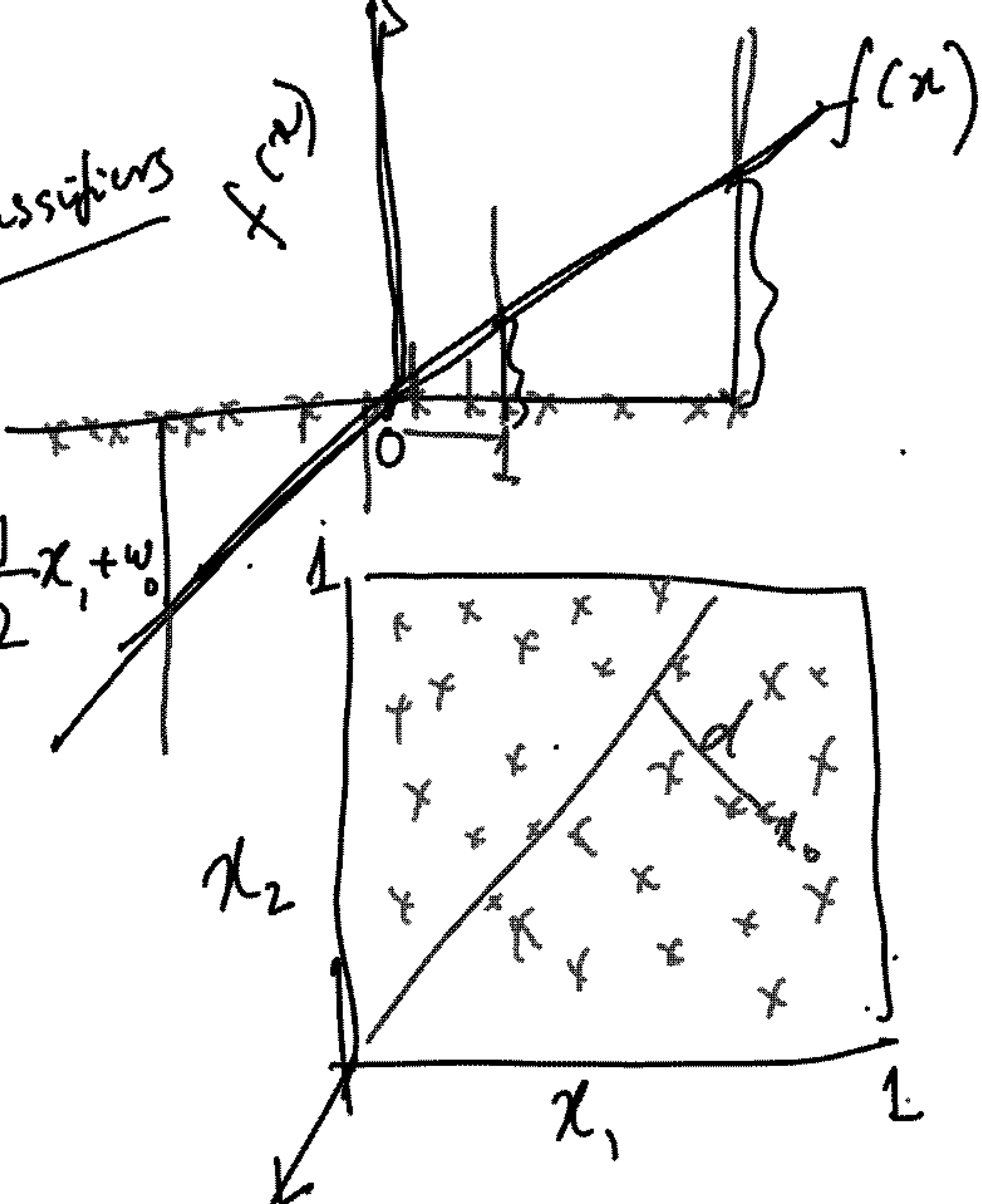
Perceptron loss

$$\text{ploss}(yf(x)) = 0 \quad \text{if } yf(x) \geq 0$$
$$= -yf(x) \quad \text{otherwise.}$$

Example $d=1$

Margin-based classifiers

$$f(x) = \frac{1}{2}x_1 + w_0$$



$d=2$

$$f(x) = w_1 x_1 + w_2 x_2 + w_0. \qquad \|w\| = 1.$$

$$f(x_0) = d\|w\|$$

Hinge Loss

$$\text{Hinge}(yf(x)) = 1 - yf(x) \quad \text{if } yf(x) \leq 1$$
$$= 0 \quad \text{otherwise}$$

Square Hinge loss

$$\text{Hinge}(y(f(x))^2$$

Logistic loss

$$\text{logloss}(yf(x)) = \log\left(1 + e^{-yf(x)}\right)$$

$yf(x) \to +\infty \quad \text{logloss} \to 0$

$yf(x) \to -\infty \quad \text{logloss} \to \infty$

$yf(x) = 0 \quad \text{logloss} = \log 2$

for large $-yf(x)$

$$\text{logloss} \approx -yf(x)$$

convex

differentiable