Surrogate Loss

<u>def</u> Convexity ...

- $\ell \geq \ell_{0-1}$
- ℓ is convex

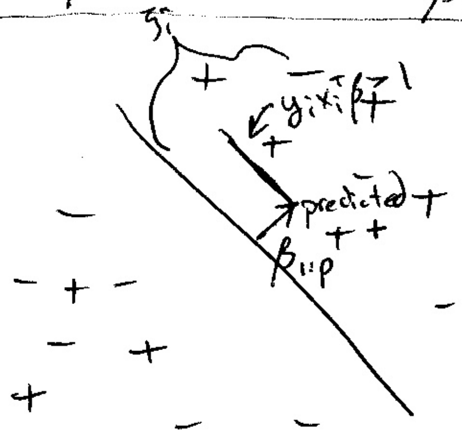
Hence, $R_{0-1}(g) \leq R(g)$ where R is calculated w/ surr. loss ℓ

Hinge loss and SVM

<u>def</u> <u>linear classifier</u> is $g_\beta(x) = \mathbb{1}\{\beta^T x > 0\}$

<u>note</u> $x_0 = 1 \Rightarrow \beta^T x = \beta_0 + \beta_{1:p}^T x_{1:p}$ (intercept term)

$$\beta^T x > 0 \Leftrightarrow \beta_{1:p}^T x_{1:p} > -\beta_0$$



Hard SVM:

$$\min \|\beta\|_2^2 \text{ s.t. }$$
$$y_i x_i^T \beta \geq 1$$

$$\alpha \left(\frac{\beta}{\|\beta\|}\right)^T \beta \geq 1 \Rightarrow \alpha = \frac{1}{\|\beta\|}$$

width of margin

<u>def</u> linearly separable if $\exists \beta$ s.t. $y_i x_i^T \beta \geq 1$

if not Soft-SVM includes slack variables

$$\min \frac{1}{n}\sum_i \xi_i + \lambda\|\beta\|_2^2 \quad \text{s.t.} \quad \xi_i \geq 0$$
$$\xi_i \geq 1 - y_i \beta^T x_i$$

$$= \min \frac{1}{n}\sum_i \max\{0, 1 - y_i \beta^T x_i\} + \lambda \|\beta\|_2^2$$