FoDA 225

Classification: Linear classifiers
loss functions

Nov 22, 2022

Classification

$X \in \mathbb{R}^{n \times d}$

Input $(X, y) \subset \mathbb{R}^d \times \{-1, +1\}$

$(x_i, y_i) \in (X, y)$

$x_i \in \mathbb{R}^d$ — stats

$y \in \{-1, +1\}$ — win / lose, bug / not bug

Goal on new data $x \in \mathbb{R}^d$

predict $g(x) \to \{-1, +1\}$

-1 if $g(x) \leq 0$

$+1$ if $g(x) > 0$

Supervised (data X, label y)

↳ cross-validation

Linear classifier $\quad (x, y) \in \mathbb{R}^d \times \{-1, +1\}$

$g : \mathbb{R}^d \to \mathbb{R} \qquad g(x) = \begin{cases} -1 & \text{if } < 0 \\ +1 & \text{if } > 0 \end{cases}$

$$g_\alpha(x) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_d x_d + \alpha_0$$

$$= \langle (\alpha_1, \dots, \alpha_d), x \rangle + \alpha_0$$

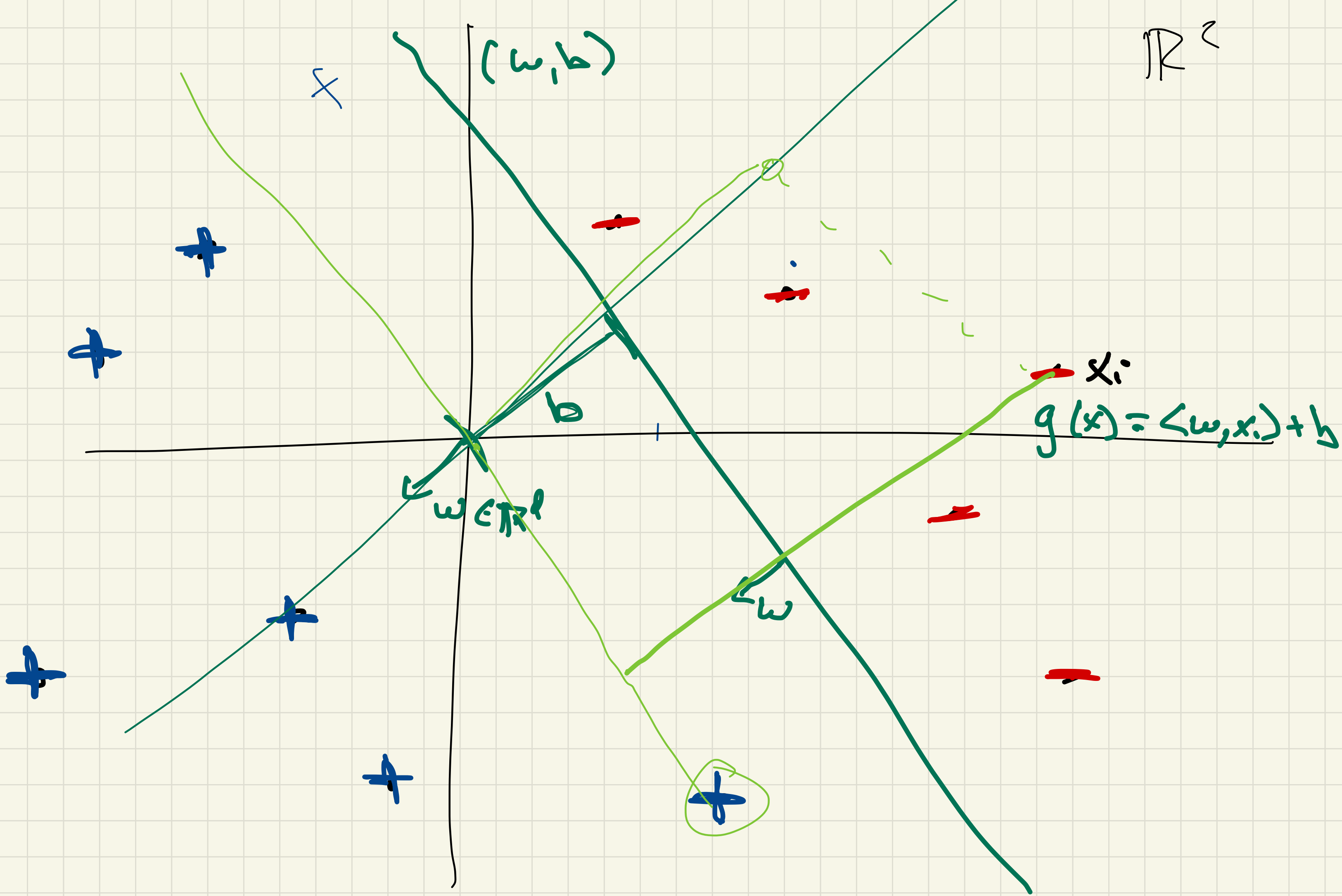
$w = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$

$b = \alpha_0$

$$g_{\alpha=(w,b)}(x) = \langle w, x \rangle + b$$

$$= \langle \alpha, (1; x) \rangle \qquad \alpha = (\alpha_0, \alpha_1 \dots \alpha_d) \in \mathbb{R}^{d+1}$$

$\|w\| = 1 \qquad b \in \mathbb{R}$


(w,b)
$\mathbb{R}^2$
x
b
$w \in \mathbb{R}^d$
x_i
$g(x) = \langle w, x_i \rangle + b$
w

How to formulate / solve for "best" $g_{w,b}$

?. Sum of squared errors?

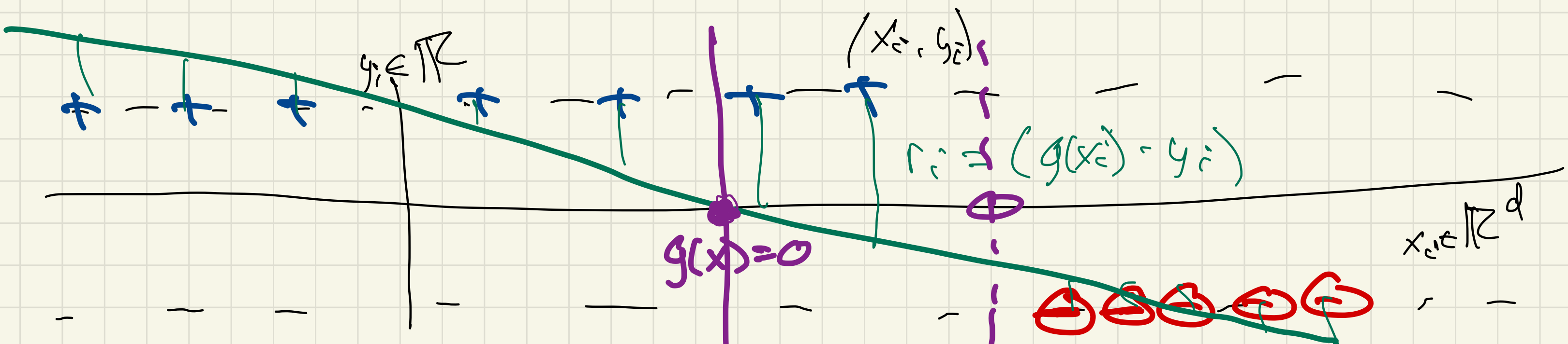
$$f_{X,y}(\alpha) = \sum_{x_i \in X} \left(g_\alpha(x_i) - y_i \right)^2$$

ideal

$g_\alpha(x_i) = +1$ if $y_i = +1$

$g_\alpha(x_i) = -1$ if $y_i = -1$

issues



Goal Minimize # misclassified points

$$f_{x,y}(\alpha) = \Delta\left(g_\alpha, (X, y)\right) = \sum_{i=1}^{n} \left[1 - \mathbb{1}\left(\text{sign}(y_i) = \text{sign}(g(x_i))\right) \right]$$

indicator

$\mathbb{1}$: True/False $\rightarrow$ 1, 0

$\mathbb{1}(\text{True}) = 1$

$\mathbb{1}(\text{False}) = 0$

Run GD on $f_{x,y}(\alpha)$?

- find gradient ?
  constant or undefined

- convex ? } $\rightarrow$ local min
  No

Loss Function

decomposable $\Rightarrow$ SGD

$$f(\alpha) = \mathcal{L}(g_\alpha, (x, y)) = \sum_{i=1}^{n} \ell(g_\alpha, (x_i, y_i))$$

$$= \sum_{i=1}^{n} \ell_\alpha(z_i)$$

$$= \sum_{i=1}^{n} f_i(\alpha)$$

$z_i = y_i \cdot g_\alpha(x_i)$

if $g_\alpha(x_i) > 0$ and $y_i > 0$

$\hookrightarrow z_i > 0$

if $g_\alpha(x_i) < 0$ and $y_i > 0$

$\hookrightarrow z_i < 0$

$z_i =$ if $z_i > 0$

$\hookrightarrow$ classified correctly

if $z_i < 0$

$\hookrightarrow$ misclassified.

$$f_i(\alpha) = \ell(z_i = y_i \, g_\alpha(x_i))$$

loss function $\ell(z_i)$

$$\Delta(z) = \begin{cases} 0 & \text{if } z \geq 0 \\ 1 & \text{if } z < 0 \end{cases}$$

hinge $\ell = \max(0, 1-z)$

squared hinge $\ell = \max(0, 1-z)^2$

smoothed hinge $= \ell = \begin{cases} 0 & z \geq 1 \\ \frac{(1-z)^2}{2} & 0 \leq z \leq 1 \\ \frac{1}{2} - z & \text{if } z < 0 \end{cases}$

logistic loss $\ell(z) = \ln(1 + \exp(-z))$

→ logistic regression