FoDA L25 : Linear Classifiers

Input $\quad X \subset \mathbb{R}^d$

$X = \{x_1, x_2, \dots x_n\}$

labels

$y \in \{-1, +1\}^n$

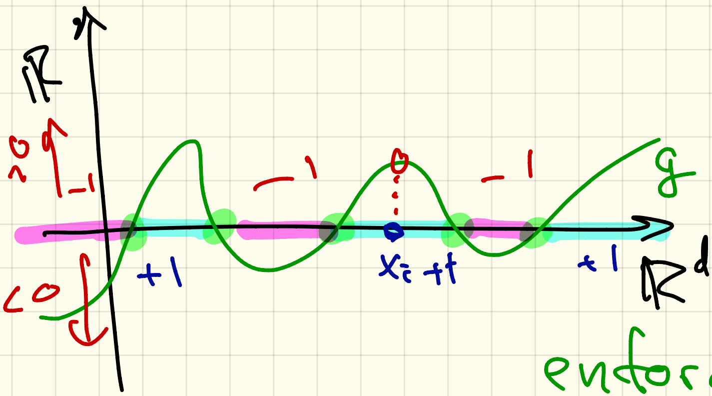
$y_i \in \{-1, +1\}$

Goal : function $\quad g : \mathbb{R}^d \to \mathbb{R}$

so. if x_i, y_i has $y = +1$

then $g(x_i) > 0$

o.w. $g(x_i) < 0$



enforce that g is linear

linear function $g : \mathbb{R}^d \to \mathbb{R}$

$x \in \mathbb{R}^d \quad x = (x^{(1)}, x^{(2)}, \dots x^{(d)})$

$$g(x) = \alpha_0 + \alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \dots + \alpha_d x^{(d)}$$

$$= \alpha_0 + \sum_{j=1}^{d} \alpha_j x^{(j)}$$

model $\quad \alpha = (\alpha_0 \dots \alpha_d) \in \mathbb{R}^{d+1}$

$b, w = b, w$

$w \in \mathbb{R}^d$

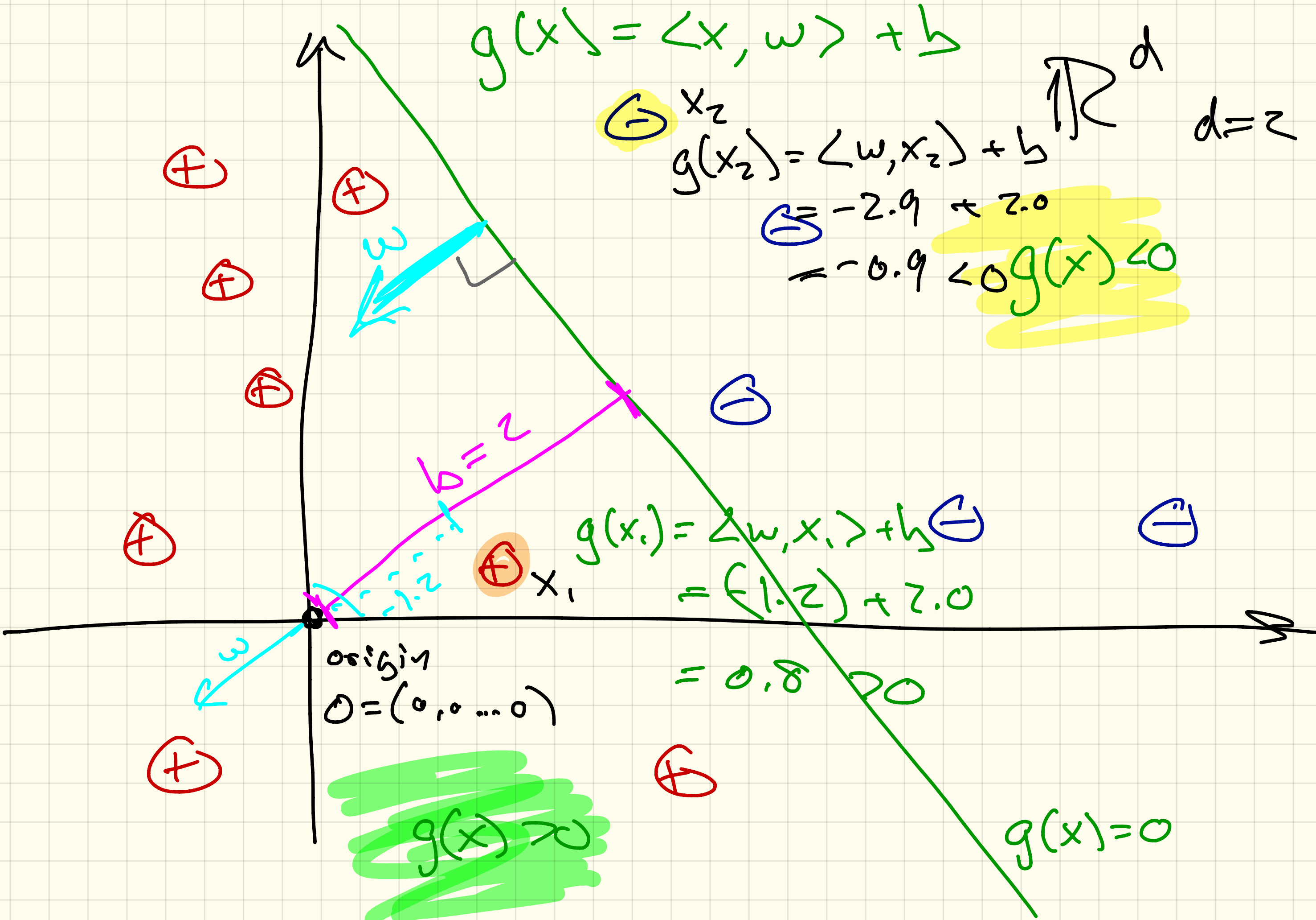
$$= b + \sum_{j=1}^{d} w_j x^{(j)}$$

$$= \langle w, x \rangle + b$$

offset: distance from origin 0 to classifier

normal of classifier

make unit vector


g(x) = <x, w> + b
R^d
d=2
x_2
g(x_2) = <w, x_2> + b
= -2.9 + 2.0
= -0.9 < 0
g(x) < 0
w
b = 2
-1.2
x_1
g(x_1) = <w, x_1> + b
= (-1.2) + 2.0
= 0.8 > 0
origin
O = (0, 0 ... 0)
w
g(x) > 0
g(x) = 0

How do we find $\omega, b \Rightarrow g_{\omega,b}$

- By linear regression

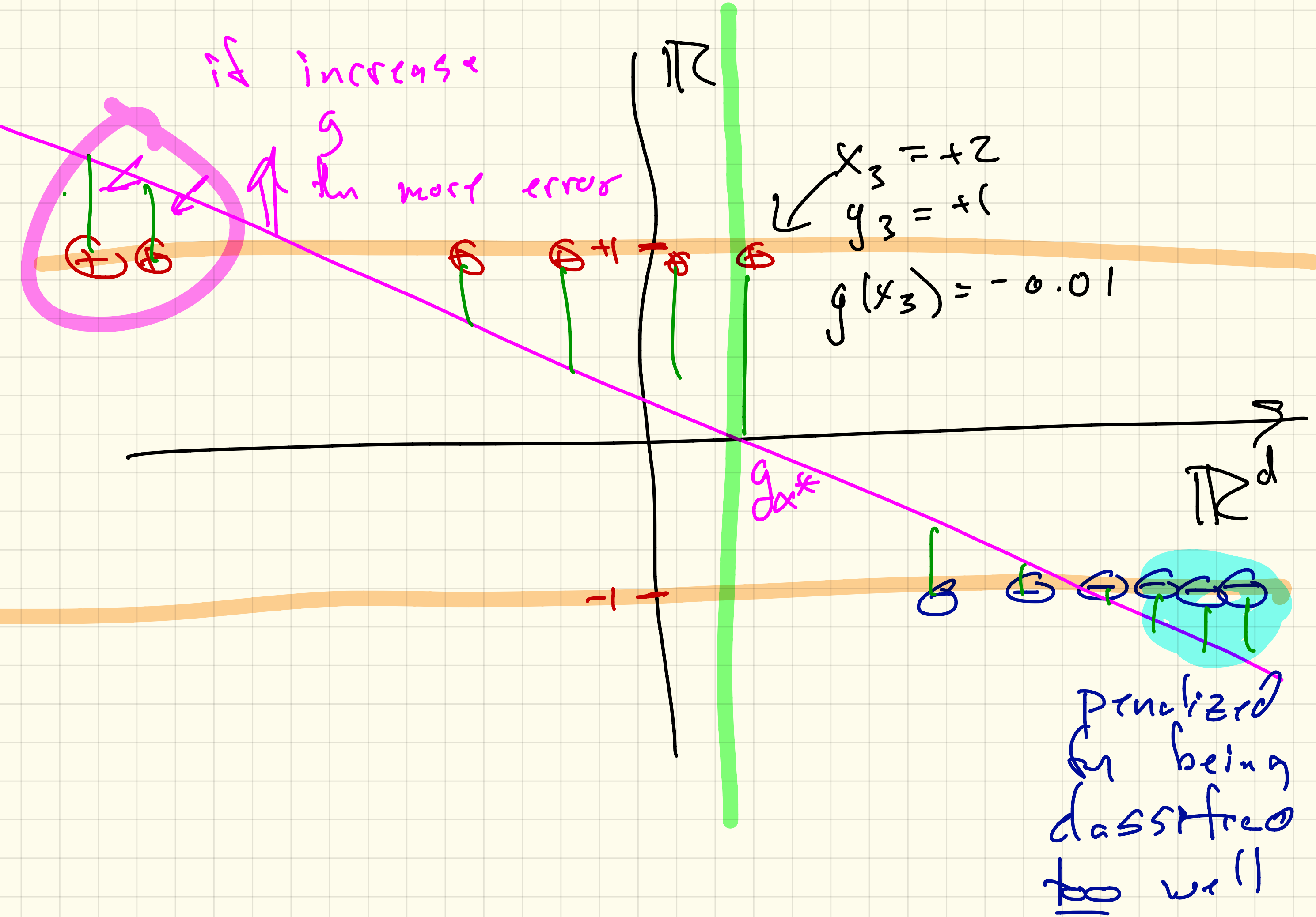
$$X \in \mathbb{R}^{n\times d} \rightarrow \tilde{X} \in \mathbb{R}^{n\times(d+1)} \qquad \tilde{x}_i := (1, x_i)$$

$$(b, \omega) = \alpha \in \mathbb{R}^{d+1} \qquad g_\alpha(x) = \langle \alpha, (1, x) \rangle$$

Solve w/ $\alpha^* = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$

This is optimizing (minimizing)

g so $\sum_{i=1}^{n} \left(g_\alpha(x_i) - y_i \right)^2$


if increase g the most error
ℝ
x3 = +2
y3 = +1
g(x3) = -0.01
+1
-1
gα*
ℝd
penalized for being classified too well

Δ cost function \Delta

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

find a proxy!

$\mathbb{1}$: identity function

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

$= \#$ of misclassified points.

Can I solve w/ gradient descent?

NO: not convex

no gradient.

Loss Function proxy for Δ

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

$$= \sum_{i=1}^{n} \ell_\alpha(z_i) \qquad z_i = y_i g_\alpha(x_i)$$

can run SGD

$$= \sum_{i=1}^{n} f_i(\alpha) \qquad f_i(\alpha) = \ell(z_i = y_i g_\alpha(x_i))$$

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

if $y_i = -1$ then we want $g_i(x_i)$ small < 0

if $y_i = +1$ then we want $g_i(x)$ large > 0

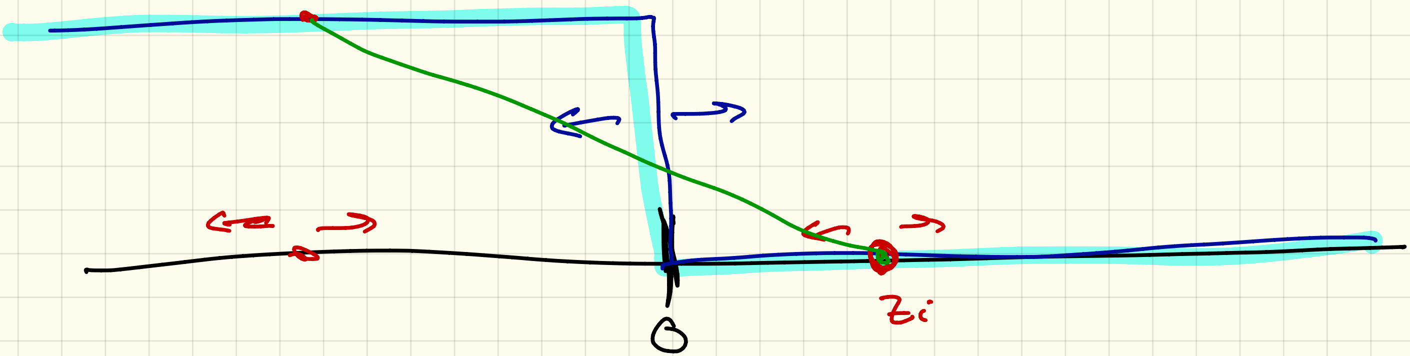
$\Rightarrow$ we want $z_i > 0$

Is (x_i, y_i) misclassified by g_α?

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

$= y_i \langle \alpha, x_i \rangle$

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

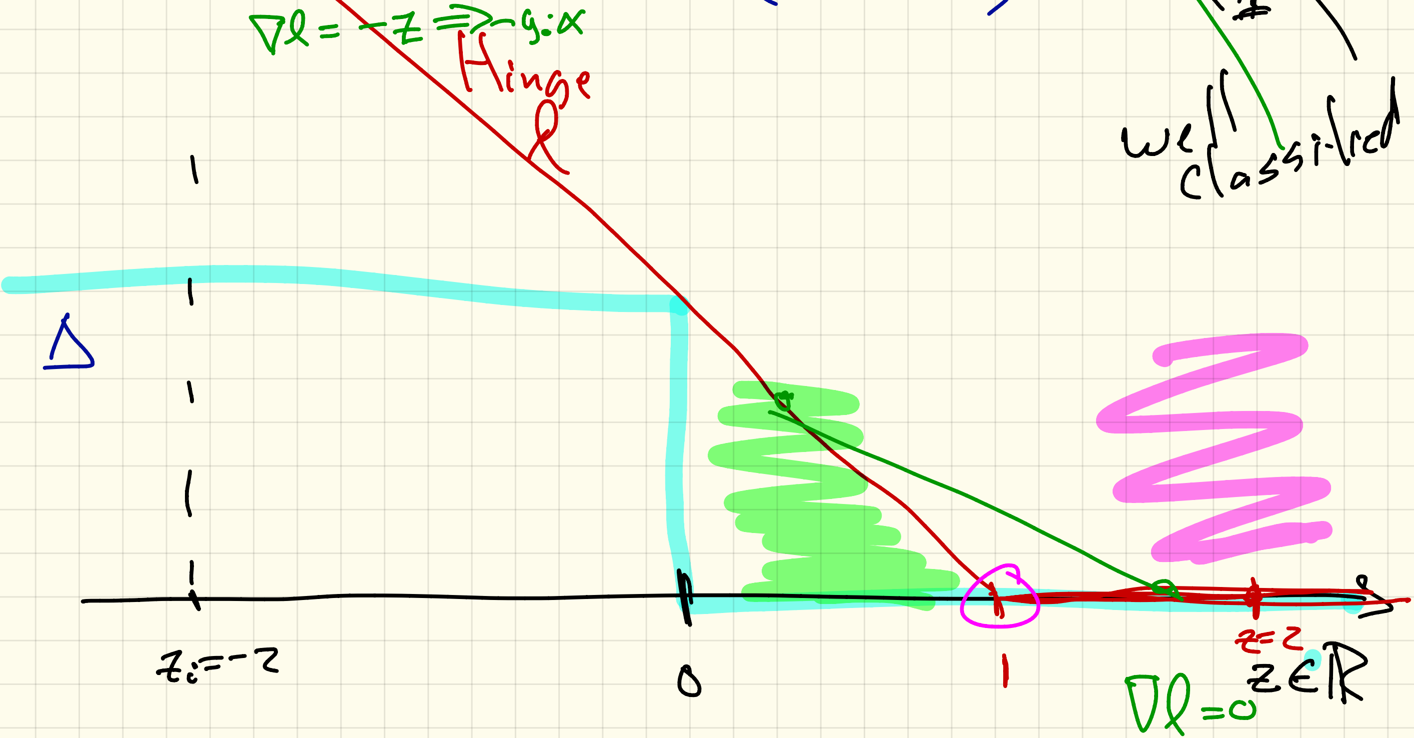


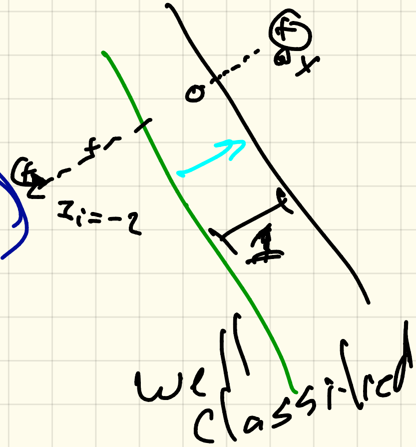
$z_i \in \mathbb{R}$

$y_i \, g_\alpha(x_0)$

Hinge Loss

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

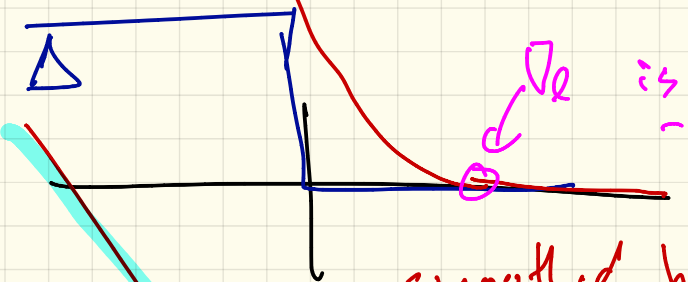




Other Loss Functions

squared hinge

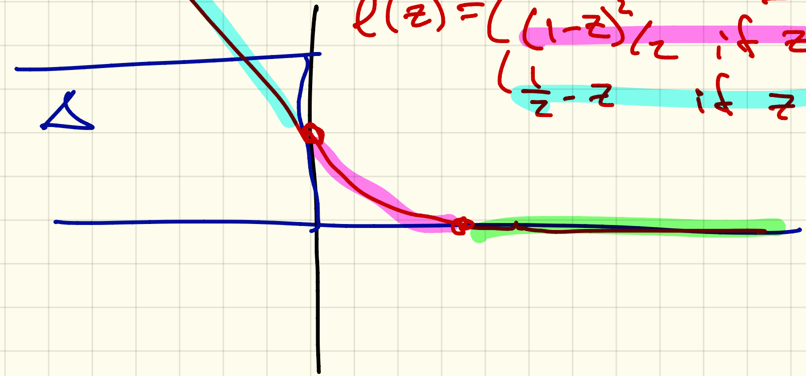
$\ell(z) = \max(0, 1-z)^2$



$\nabla \ell$ is well-defined

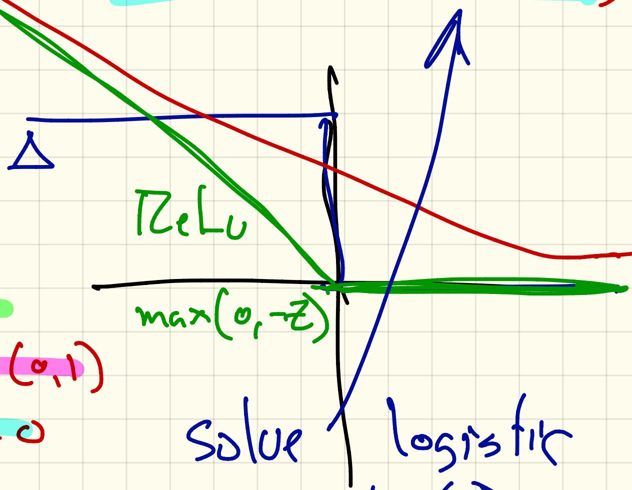
smoothed hinge

$$\ell(z) = \begin{cases} 0 & \text{if } z \ge 1 \\ (1-z)^2/2 & \text{if } z \in (0,1) \\ \frac{1}{2} - z & \text{if } z < 0 \end{cases}$$



logistic loss

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



Solve logistic loss w/ GD

↳ logistic regression