$n := |D_{train}|$ $\quad \text{TrainLoss}(w) = \cancel{\frac{1}{n}} \sum_{i=1}^{n} (w \cdot \phi(x_i) - y_i)^2$

$\sum_{i=1}^{n} z_i^2 = \|z\|_2^2 \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$

$= z^T z$

$w \cdot \phi(x)$
$\Updownarrow$
$w^T \phi(x)$

$[w_1, \dots w_n] \begin{bmatrix} \phi(x)_1 \\ \vdots \\ \phi(x)_n \end{bmatrix}$

$\left\| \begin{bmatrix} w \cdot \phi(x_1) - y_1 \\ \vdots \\ w \cdot \phi(x_n) - y_n \end{bmatrix} \right\|_2^2 \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$

$X := \begin{bmatrix} \phi(x_1)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix} \in \mathbb{R}^{n \times d} \quad w \in \mathbb{R}^d$

$Xw \in \mathbb{R}^n$

$Xw - \vec{y} \in \mathbb{R}^n$

Matrix - Vector Product

Dotting rows of matrix w/ vector

$\text{TrainLoss}(w) = \|Xw - \vec{y}\|_2^2 = (Xw - \vec{y})^T (Xw - \vec{y})$

$= (w^T X^T - \vec{y}^T)(Xw - \vec{y})$ FOIL

$(AB)^T = B^T A^T$

$(A + C)^T = A^T + C^T$

$(\lambda A)^T = \lambda A^T$

$$= \underline{w^T X^T X w} - \underbrace{w^T X^T \vec{y}}_{(1)} \underset{=}{} - \underbrace{\vec{y}^T X w}_{(2)} + \underline{\vec{y}^T \vec{y}} = \text{TrainLoss}(w)$$

$\lambda^T = \lambda, \lambda \in \mathbb{R}$

(1) $w^T X^T \vec{y} = w^T (X^T \vec{y}) = (w^T (X^T \vec{y}))^T = (X^T \vec{y})^T w$

(2) $y^T X w = (y^T X) w = (X^T \vec{y})^T w$

$$\text{TrainLoss}(w) = w^T X^T X w - 2(X^T \vec{y})^T w + \vec{y}^T \vec{y}$$

Gradient Identities

(1) $\nabla_w (w^T A w) = Aw + A^T w, \quad A \in \mathbb{R}^{d \times d}$

$\Rightarrow$ (A symmetric) $= 2Aw$

$X^T X$ is symmetric!

(2) $\nabla_w (a^T w) = a, \quad a \in \mathbb{R}^d$

$$\nabla_w \text{TrainLoss}(w) = 2X^T X w - 2X^T \vec{y} := 0$$

$$X^T X w^* = X^T \vec{y} \quad \text{(Normal Equations)}$$

(1) $X^T X$ invertible $(X^T X)^{-1}$ exists

$$w^* = (X^T X)^{-1} X^T \vec{y} \quad \underline{\text{unique}}$$

(2) $X^T X$ not-invertible (pseudo-inverse) $\Rightarrow w^*$ (not unique)

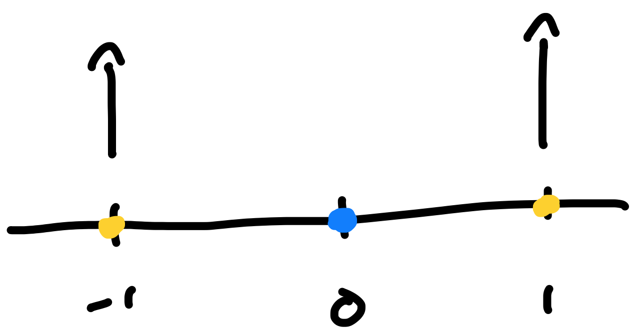
$D_1 = \{(-1,+1),(0,-1),(1,+1)\}$

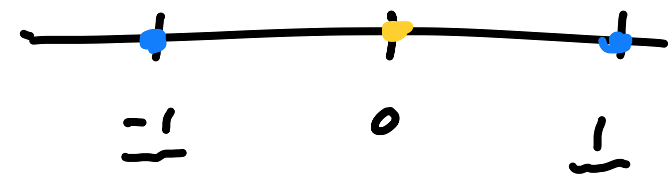
$D_2 = \{(-1,-1),(0,+1),(1,-1)\}$

$\phi(x) = (x, |x|)$

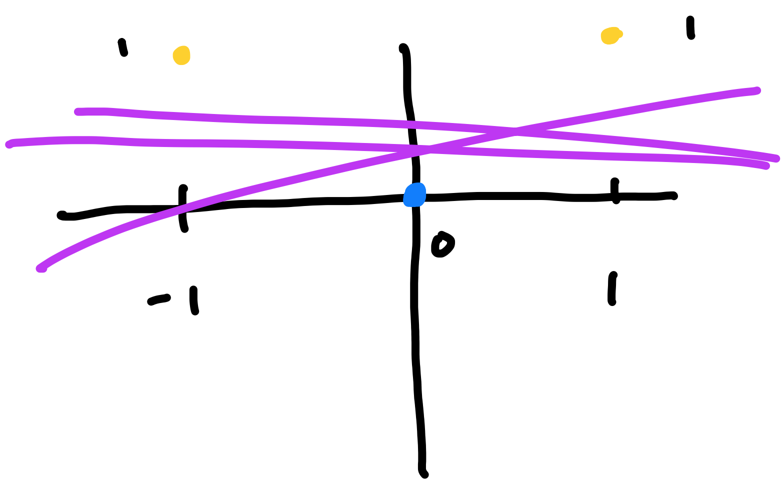
$\phi(-1) = (-1, 1)$

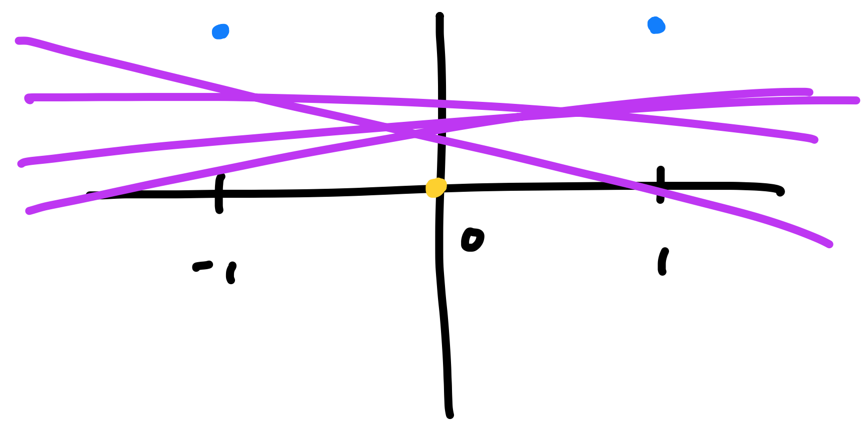
$\phi(1) = (1, 1)$





ϕ





$-1 \mapsto 1$

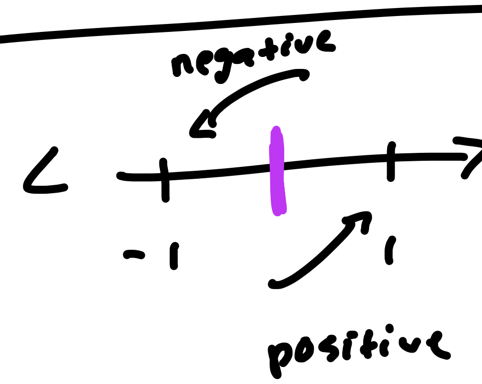
$0 \mapsto 0$

$1 \mapsto 1$

$\phi(x) = (x, |x|)$

$\phi(x_i) \mapsto y_i$

$i = 1, \dots, n$



ϕ is not applicable to new points!

ϕ, w essentially just memorization

$$\text{Loss}(x, y, z, w) = 2(xy + \max\{w, z\})$$

$$\nabla_x : 2 \times 1 \times y$$
$$= 2 \times 1 \times -4 = -8$$

$$\nabla_z : 2 \times 1 \times 1 = 2$$