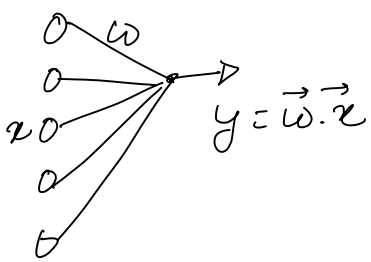
Pruning a neuron.

$y = (w, x)$

$w, x \in R^d$



Q. What should we try to preserve when pruning?

- Behavior of neuron.
- What function it was representing

Remark: While pruning at "Inference" time the considerations are ~~different~~ - we want to preserve the function itself as If we were thinking about storing a model ~~and its~~ where expressive power matters.

To model behavior: we want to say preserve the following under compression

If $\hat{w} \in R^d$ s.t $\|\hat{w}\|_1 \leq k$

we want to preserve the f^n

$\hat{w} \cdot x \simeq w \cdot x \quad \forall x$

Let X be the $\mathcal{F}$ representative data then

$$\hat{w} = \underset{\|\hat{w}\|_1 \leq k}{\operatorname{argmin}} \left[\|\hat{w} \cdot x - w \cdot x\|_1^2 \right]$$

Write this as,

$$\hat{w}, M = \underset{\substack{\hat{w} \in R^d \\ M \in [0,1]^d \\ \|M\|_1 \leq k}}{\operatorname{argmin}} \left[\|(\hat{w} \odot M) \cdot x - w \cdot x\|_2^2 \right]$$

Jointly optimising M & $\hat{w}$ is NP Hard (Blumensath & Davies 2008)

$\mathcal{A} \rightarrow$ Greedy Algorithm

Which element to prune first?

→ Magnitude? $|w_i|$ ↑ but $|x_i|$ very low.

→ $f(|w_i|, |x_i|)$? does not capture the correlations.

→

$y = w \cdot X$ ← true model

$\hat{y} = \hat{w} \cdot X$ ← approximated model.

$$\min \; \| \hat{w} \cdot X - w \cdot X \|_2^2$$

$\|\hat{w}\| \leq d-1$

$\hat{w}_i = 0$ for some i

$$\min_{w_1=0} \| \hat{w} \cdot X - w \cdot X \|_2^2$$

$$\min_{w_2=0} \| \hat{w} \cdot X - w \cdot X \|_2^2$$

$$\min_{w_d=0} \| \hat{w} \cdot X - w \cdot X \|_2^2$$

→ Whichever gives the best loss we choose that $\|\cdot\|$

Subproblem.

$$(i) \quad \hat{w}^{(i)} = \underbrace{\operatorname{argmin} \left\| X^{(i)} \cdot w^{(i)} - X \cdot w \right\|_2^2}_{\text{linear prob.}}$$

y

$$\left[X \,\middle|\, 0 \leftarrow i \right] = y$$

Normal eq. $\quad X_{-i} \cdot \hat{w} = y.$

$$\boxed{\hat{w} = \left(X_{-i}^T X_{-i} \right)^{-1} \cdot X_{-i}^T \cdot y}$$

loss: $\quad \hat{y} = X_{-i}^T \hat{w}$

$$l_i = \left(\| \hat{y} - y \|_2^2 \right) \leftarrow \text{choose that least}$$

(

① "Saliency"

② update $w_{-i} \leftarrow \hat{w}$

Pruning a Model