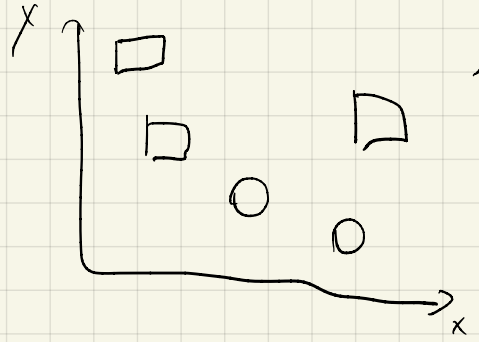
28.9

- implement fairness: keep FCNN as is, increase #channels in CNN to match #params of FCNN

- fairness v2 (baseline learning task):
  - fix FCNN architecture
  - increase #channels of CNN until same test accuracy as FCNN can be reached
  - plot $\to d$

e.g.: linear classification $\sum_{i=1}^{d/2} x_i \leq \sum_{i=d/2+1}^{d} x_i$

"sanity check":



$\text{label} = \mathbb{1}\{\#\square > \#\bigcirc\}$

→ should be much easier for CNN!

labeling fct. h, input dist D_x: $\text{dist.} = h \circ D_x$

draw once 50k, 10k test "images", fix them

$D_x \sim \text{Ber}(0.5)^d$ $(\mathcal{N}(0, I_d))$

Meta algorithm

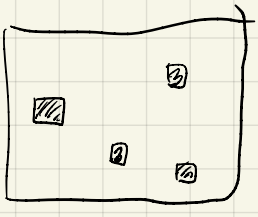
init $h_0 = (1, \dots, 1)$ (d) → fix D_x

train FCNN (S_{h_t}) → bisection till 0.7 test acc → $n_{t,\text{FCNN}} = ?$

-"- CNN (S_{h_t}) -"- → $n_{t,\text{CNN}} = ?$

sep. $s_t(h_t) = \frac{n_{t,\text{FCNN}}}{n_{t,\text{CNN}}}$

$h_{t+1} = h_t \oplus \operatorname{argmax}_h s_t(h_t \oplus h)$ ("GD-like")



need something smarter