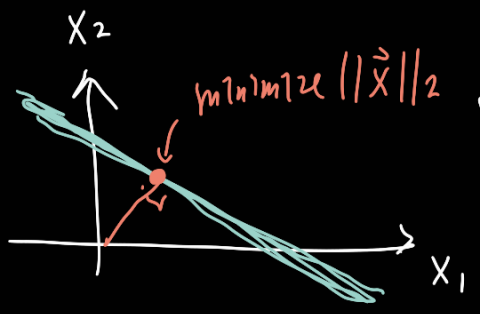
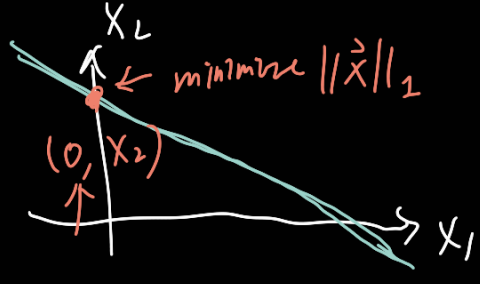
l_1 vs l_2 regularization

- Model $\vec{A}\vec{x} = \vec{b}$, where $\vec{A}, \vec{b}$ are known, find $\vec{x}$

eg $b = [a_1 \; a_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow b = a_1x_1 + a_2x_2$ — ①

1. 

$$L = \underbrace{||\vec{A}\vec{x} - \vec{b}||_2}_{\text{minimize get approx soln of } \vec{x} \text{ to ①}} + \underbrace{\lambda ||\vec{x}||_2}_{\text{minimized}} \qquad \sqrt{x_1^2 + x_2^2}$$

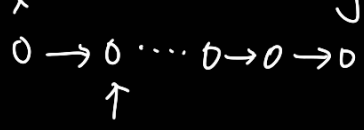
2. 

$$L = ||\vec{A}\vec{x} - \vec{b}||_2 + \lambda ||\vec{x}||_1 \rightarrow \text{sparse soln to } \vec{x} \qquad |x_1| + |x_2|$$

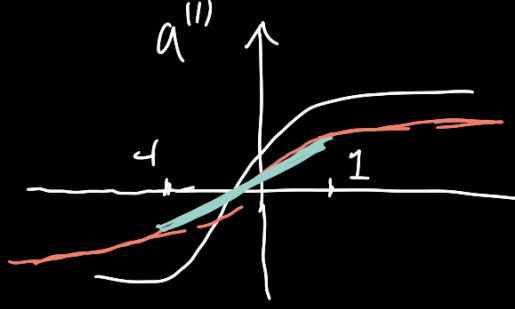
- NN model. $y = fn(w_i, b_i, \vec{x})$ (output, input)

What does simple models mean? 1. want as many $w_i = 0$ as possible

2. want smaller w_i → simpler linear model

Eg. 

$$a^{(1)} = \tanh(wx + b)$$



smaller $w \rightarrow a^{(1)} \sim x$ for 1 layer

for all layers small $w \rightarrow y = NN(x) \sim x$ linear.

give simpler models

1. $J = \underbrace{\frac{1}{m}\sum_{i}^{m} L(\hat{y}_i, y_i)}_{} + \boxed{\underbrace{\lambda \|\vec{w}\|_2^2}_{\hookrightarrow \text{give smaller weights.}}}$

2. $J = \frac{1}{m}\sum_{i}^{m} L(\hat{y}_i, y_i) + \boxed{\lambda \|\vec{w}\|_1}$ $\quad \vec{w} = (w_1, w_2, w_3 \ldots)$

want sparse solution to $\vec{w}$