bias vs. variance

under-fitting $\Uparrow$ training

overfitting $\uparrow$ variance

validation error - training error

bias $\xrightarrow{H}$
- Layers $\uparrow$ neurons $\uparrow$
- NN arch.

$\downarrow$ L

variance $\xrightarrow{H}$
- Collect more data.
- $\boxed{\text{regularization.}}$

$$\mathcal{L}(w,b) = \sum_{i=1}^{m} \left(y^{(i)} - w^T x^{(i)} - b\right)^2$$

$$\tilde{w} = \begin{bmatrix} b \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \qquad \tilde{x}^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \\ \vdots \\ x_d^{(i)} \end{bmatrix}$$

$(\tilde{x}^{(1)}, y^{(1)})$

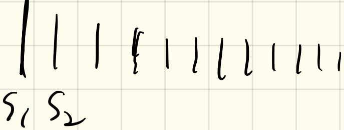
$(\tilde{x}^{(2)}, y^{(2)})$

$\vdots$

$(\tilde{x}^{(m)}, y^{(m)})$

$$\mathcal{L}(\tilde{w}) = \sum_{i=1}^{m} \left(y^{(i)} - \tilde{w}^T \tilde{x}^{(i)}\right)^2$$

Good.

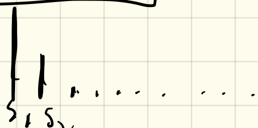


$$\|X^{-1}\|$$

$$\tilde{w}^* = \boxed{X^{-1}}\, Y$$

$$X = \sum_{i=1}^{m} \tilde{x}^{(i)} \tilde{x}^{(i)T}, \qquad Y = \sum_{i=1}^{m} y^{(i)} \tilde{x}^{(i)}$$

Bad.



$$\boxed{Y = \tilde{w}^T X + \varepsilon}$$

$$\tilde{w}^* = (X + \lambda I)^{-1} Y \quad \Leftarrow \quad X^{-1} Y$$

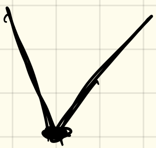
$$\|(X+\lambda I)^{-1}\|_2 \qquad\qquad \|X^{-1}\|_2$$

s_d : the least singular value.

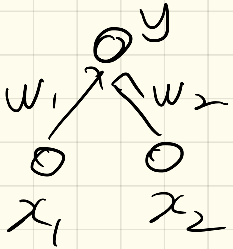
$$\|(X+\lambda I)^{-1}\|_2 = \frac{1}{s_d+\lambda} \qquad \|X^{-1}\|_2 = \frac{1}{s_d}$$

$$\mathcal{L}(\tilde{w}) = \frac{1}{m}\sum_{i=1}^{m} \left(Y^{(i)} - \tilde{w}^T x^{(i)}\right)^2 + \frac{\lambda}{2m} \underbrace{\tilde{w}^T \tilde{w}}_{\|\tilde{w}\|_2^2}$$

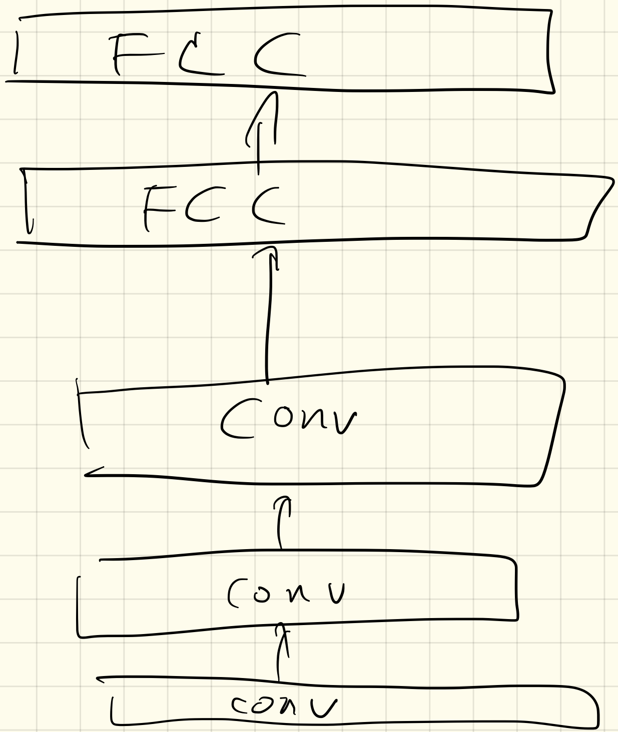
$$\boxed{l_2}$$



$\boxed{l_1}$: $\frac{\lambda}{2m} \cdot \|\tilde{w}\|_1 \Leftarrow$ LASSO



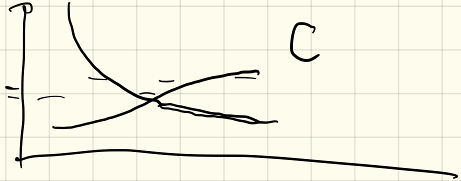
$$\hat{Y} = w_1 x_1 + w_2 x_2$$



- $\mathcal{L}(w,b) \to 0$

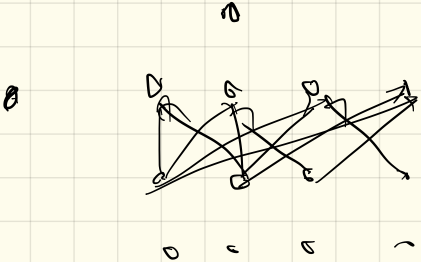
$+ \ C(w) \Leftarrow \|w\|_2^2, \ \|W\|_F^2, \ \|w\|_1$

$\mathcal{L}$

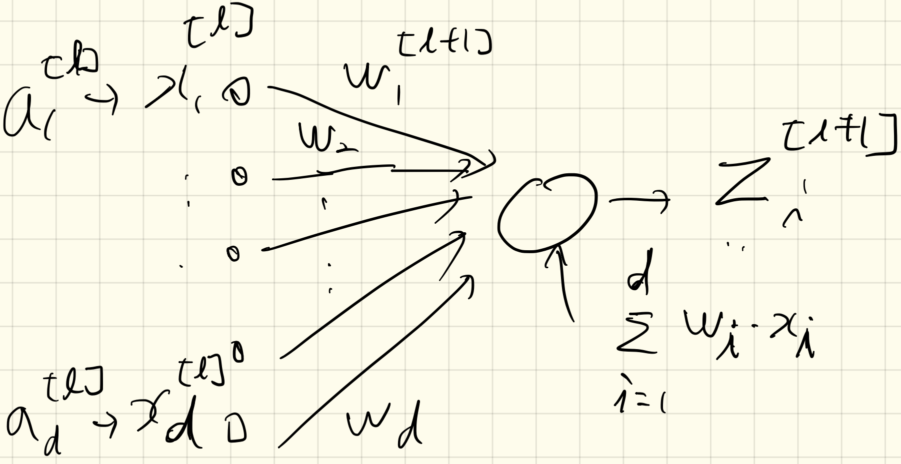


$\Downarrow$

weigh decay



dropout.



$$\sum_{i=1}^{d} w_i \cdot x_i$$

$$m_i = \begin{cases} 0 & \text{w.p. } 1-p \\ 1 & \text{w.p. } p \end{cases}$$

iid random variable

dropout : $\displaystyle\sum_{i=1}^{d} w_i \cdot x_i \cdot m_i$

test : $\displaystyle\mathbb{E}\left[\sum_{i=1}^{d} w_i x_i m_i\right]$

$$= p \cdot \sum w_i x_i$$