Convolutional Neural Nets

Exploiting stationarity, locality, and compositionality of natural data

Input layer / samples

$$\mathcal{X} = \{\boldsymbol{x}^{(p)} \in \mathbb{R}^n \mid \boldsymbol{x}^{(p)} \text{ is a data sample}\}_{p=1}^{P} \quad \text{input samples}$$

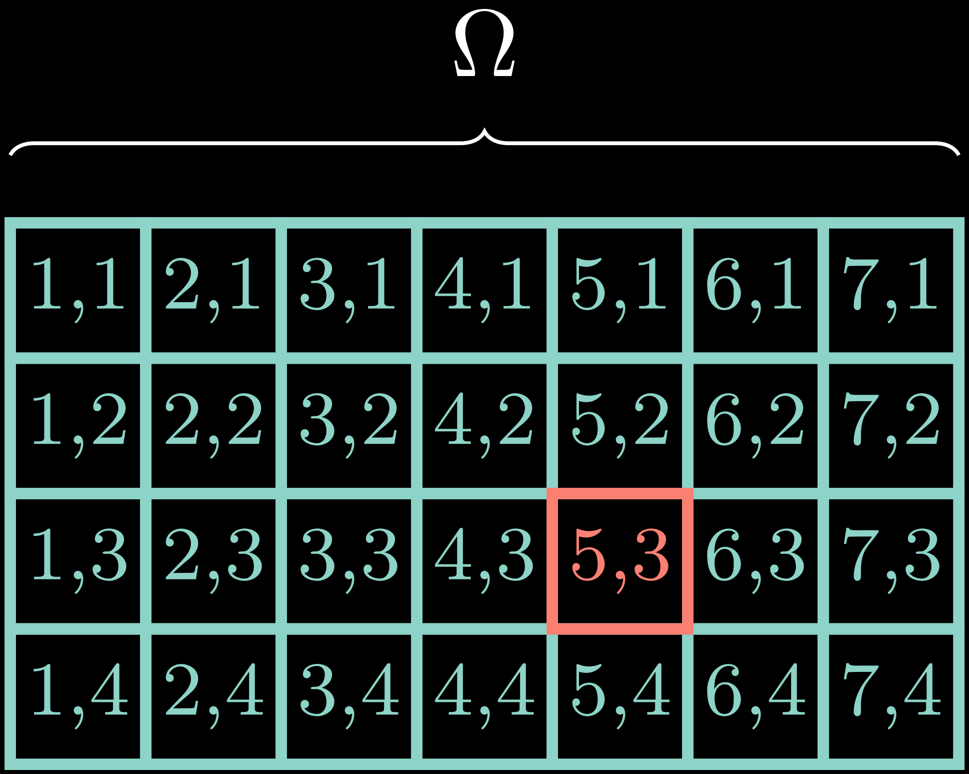
$$\mathcal{X} = \{\boldsymbol{x}^{(p)} : \overset{\text{domain}}{\Omega} \to \overset{\text{channels}}{\mathbb{R}^c}, \boldsymbol{\omega} \mapsto \boldsymbol{x}^{(p)}(\boldsymbol{\omega})\}_{p=1}^{P}$$

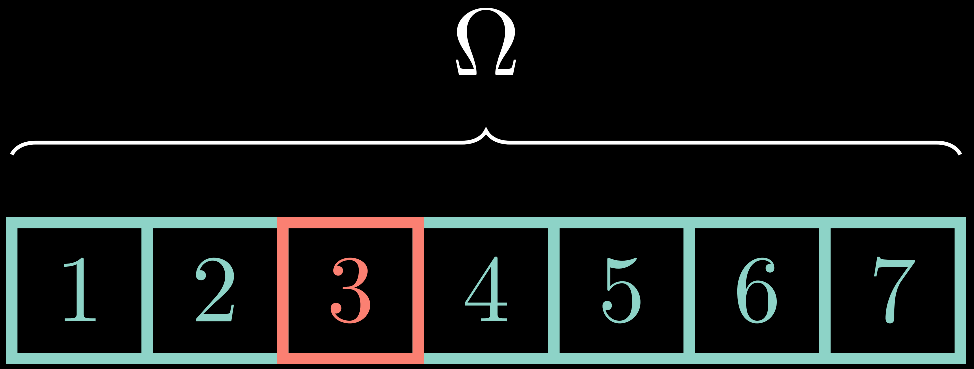
$$\Omega = \{1, 2, \cdots, \overset{\text{total time}}{\underset{\text{sampling interval}}{T/\Delta t}}\} \subset \mathbb{N}, \qquad c \in \{\underset{\text{mono}}{1}, \overset{\text{stereo}}{2}, \underset{\text{Dolby 5.1}}{5+1}, \cdots\}$$

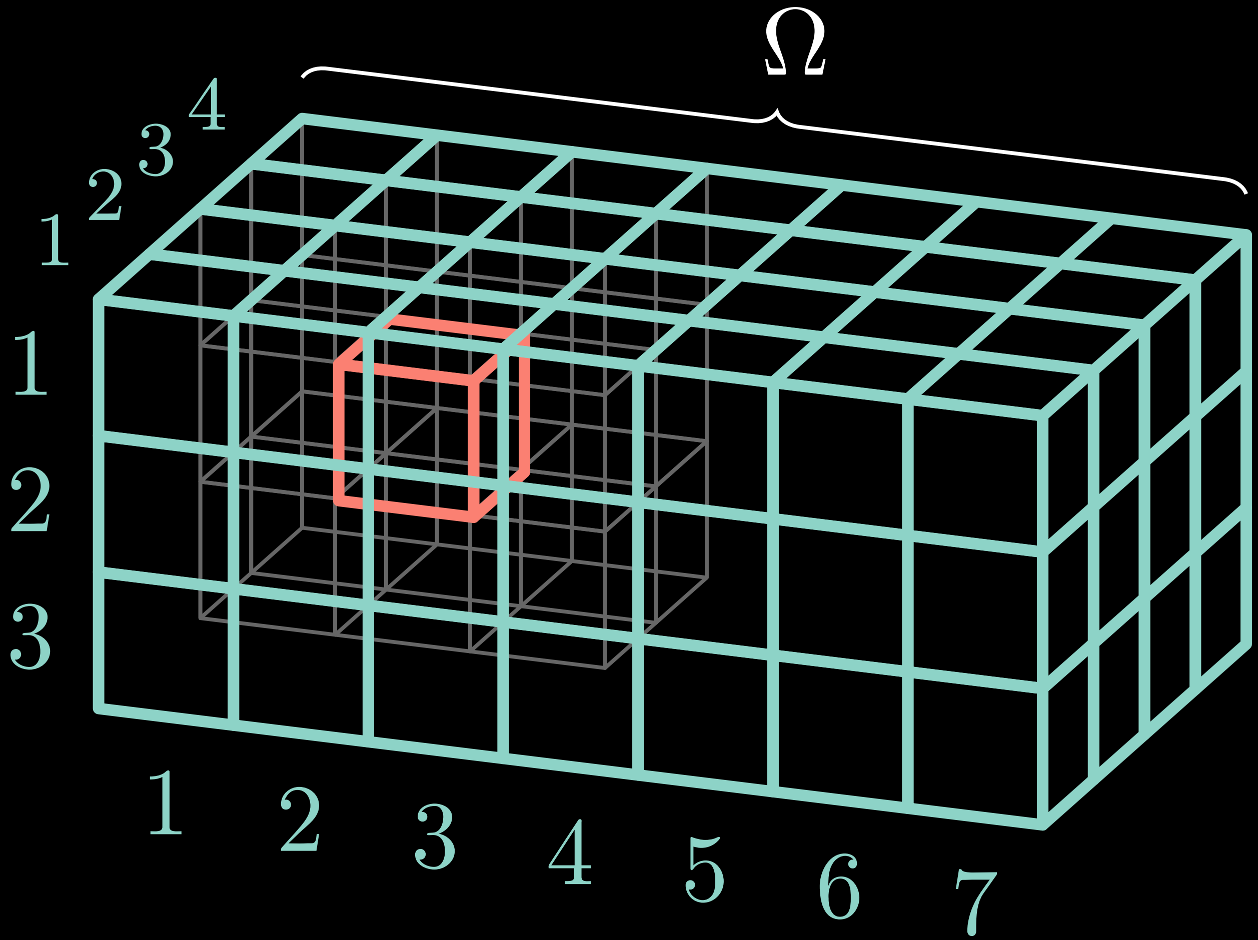
$$\Omega = \{1, \cdots, \overset{\text{height}}{h}\} \times \{1, \cdots, \overset{\text{width}}{w}\} \subset \mathbb{N}^2, \qquad c \in \{\overset{\text{grey scale}}{1}, \underset{\text{colour}}{3}, \underset{\text{hyperspectral}}{20}, \cdots\}$$

$$\Omega = \underset{\text{space-time}}{\mathbb{R}^4} \times \overset{\text{four-momentum}}{\mathbb{R}^4}, \qquad c = \underset{\text{Hamiltonian}}{1} \qquad \boldsymbol{x}(\omega_1, \omega_2) = \begin{pmatrix} r(\omega_1, \omega_2) \\ g(\omega_1, \omega_2) \\ b(\omega_1, \omega_2) \end{pmatrix}$$

Regular domains







Signals can be represented as vectors

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; \ldots \;\; x_t \;\; \ldots]^\top$$

x_t are waveform heights

$$\boldsymbol{x} = [x_{11} \;\; x_{12} \;\; \ldots \;\; x_{1n} \;\; x_{21} \;\; x_{22} \;\; \ldots]^\top$$

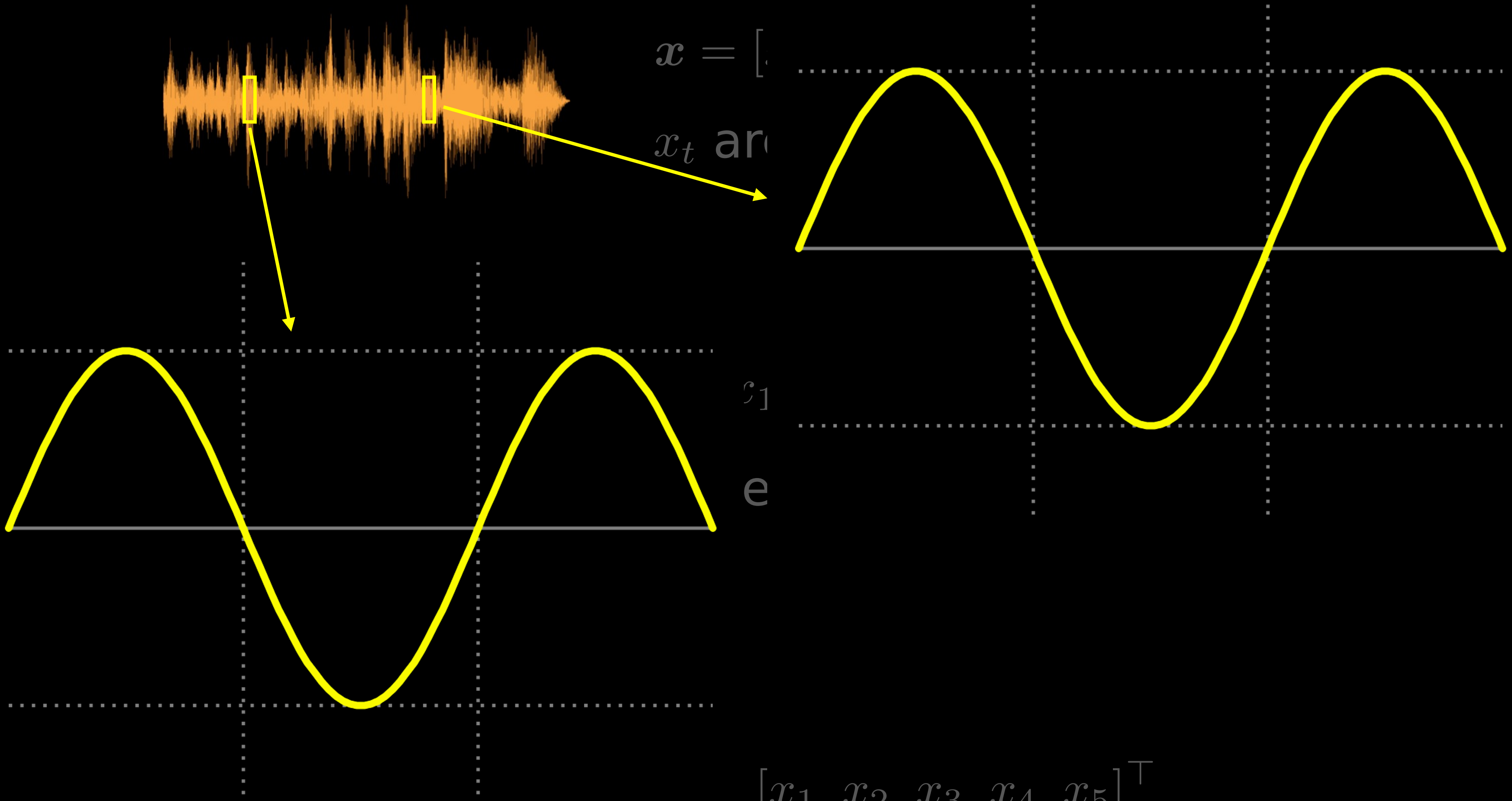
x_{ij} are pixel values

"John picked up the apple"

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^\top$$

x_t are one-hot vectors

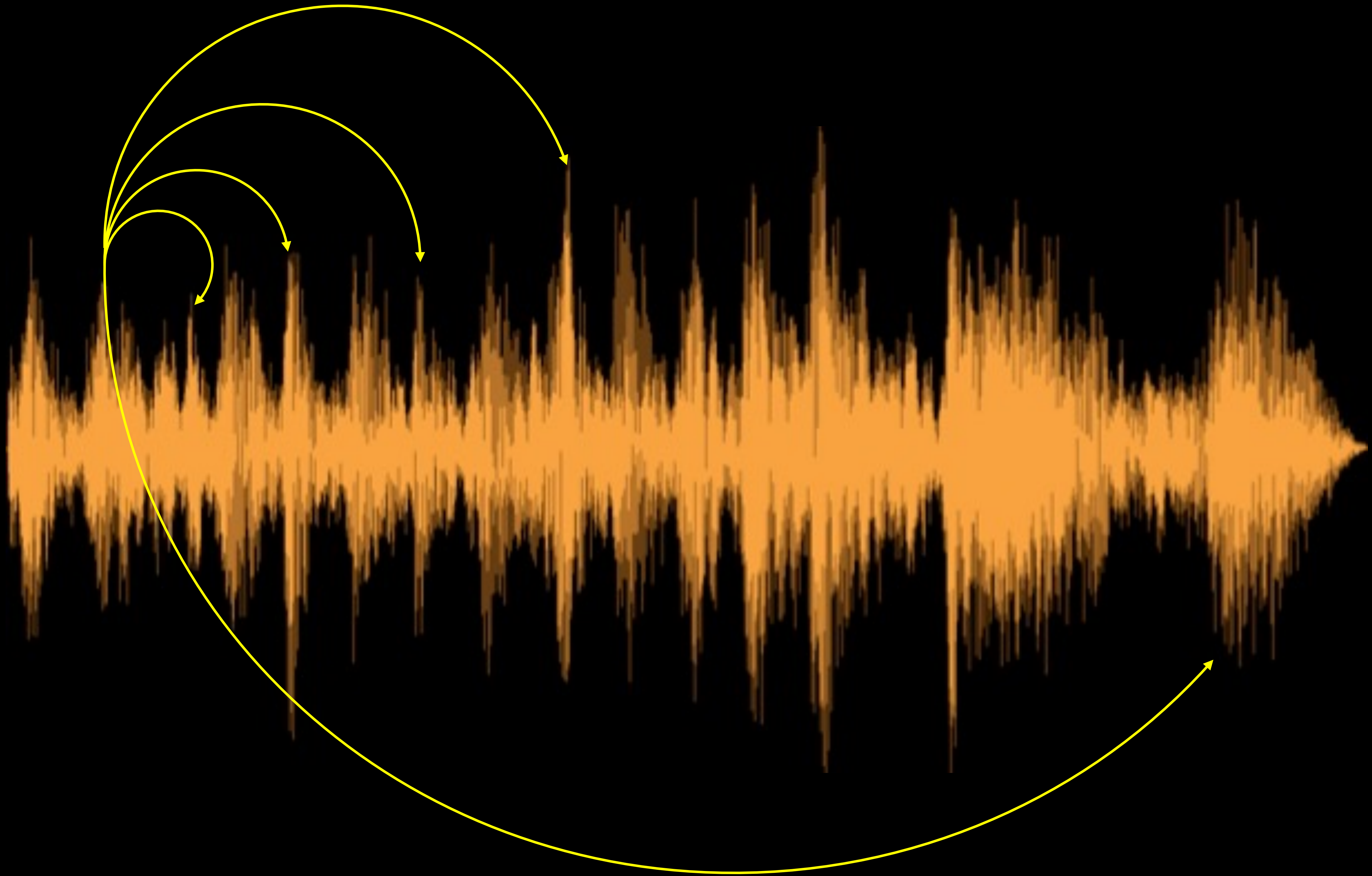
Signals can be represented as vectors



“John picked up the apple”

$$[x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^\top$$

x_t are one-hot vectors



Signals can be represented as vectors

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; \ldots \;\; x_t \;\; \ldots]^\top$$

x_t are waveform heights

$$\boldsymbol{x} = [x_{11} \;\; x_{12} \;\; \ldots \;\; x_{1n} \;\; x_{21} \;\; x_{22} \;\; \ldots]^\top$$

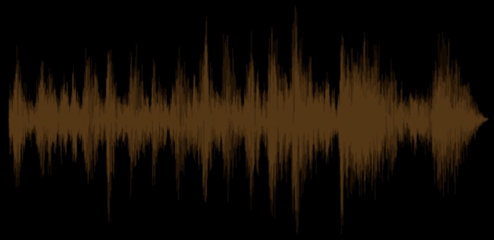
x_{ij} are pixel values

“John picked up the apple”

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^\top$$

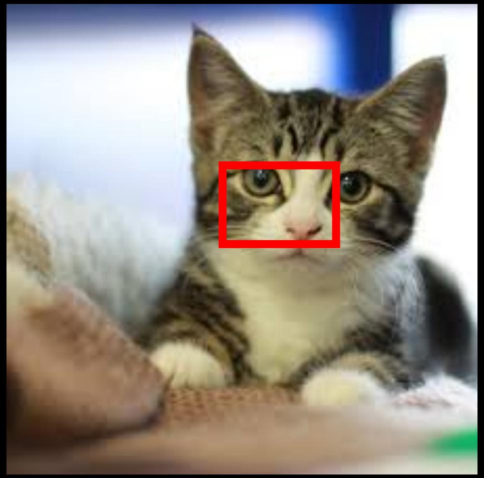
x_t are one-hot vectors

Signals can be represented as vectors



$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; \ldots \;\; x_t \;\; \ldots]^\top$$

x_t are waveform heights



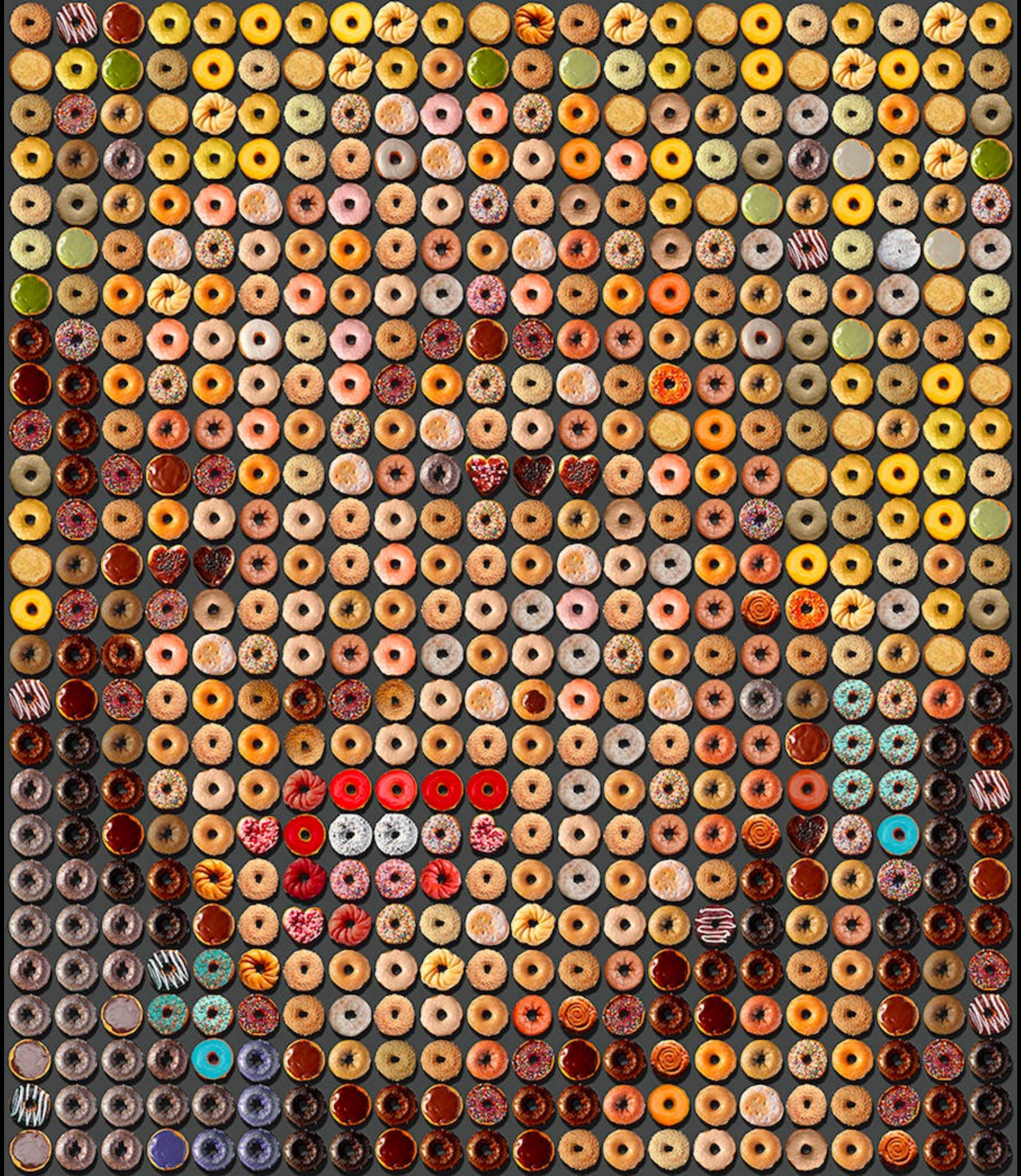
$$\boldsymbol{x} = [x_{11} \;\; x_{12} \;\; \ldots \;\; x_{1n} \;\; x_{21} \;\; x_{22} \;\; \ldots]^\top$$

x_{ij} are pixel values

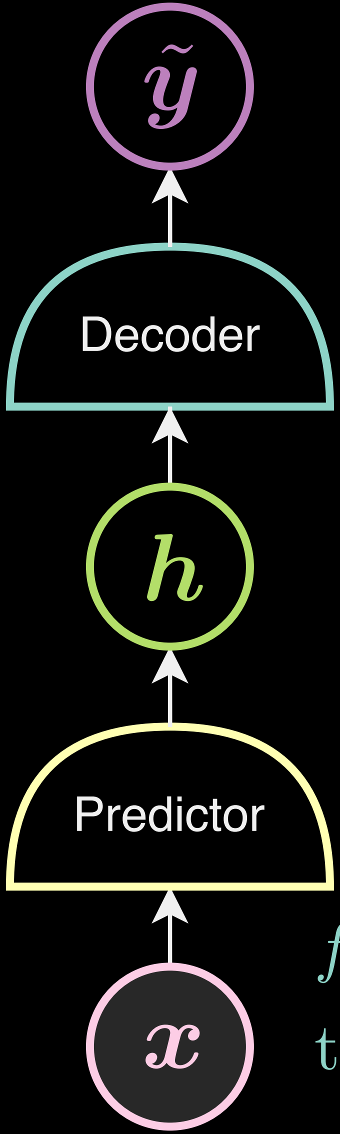
"John picked up the apple"

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^\top$$

x_t are one-hot vectors



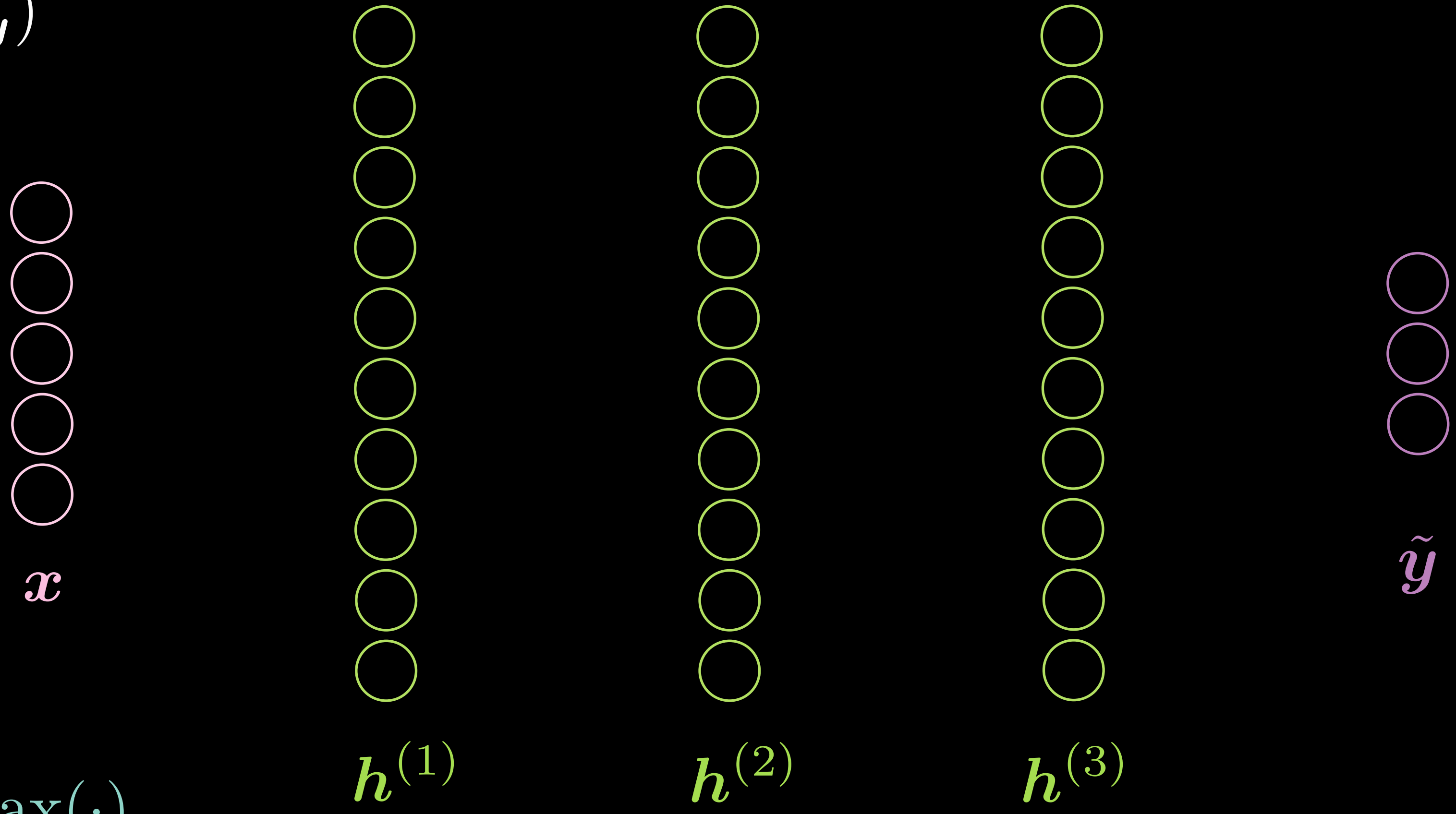
Fully connected (FC) layer



$$\boldsymbol{h} = f(\boldsymbol{W_h x} + \boldsymbol{b_h})$$

$$\tilde{\boldsymbol{y}} = g(\boldsymbol{W_y h} + \boldsymbol{b_y})$$

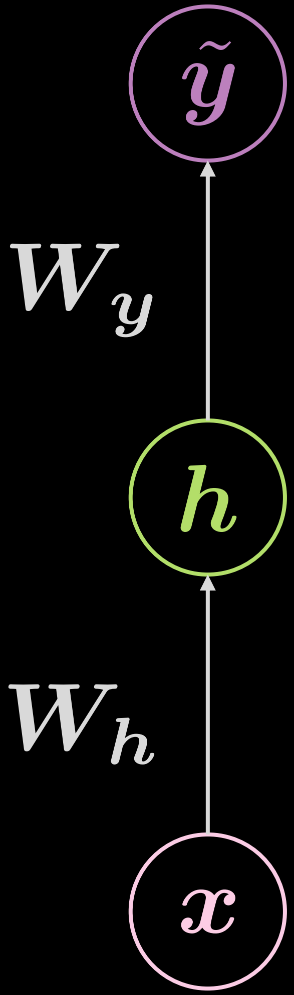
$f, g = (\cdot)^+, \sigma(\cdot),$
$\tanh(\cdot), \mathrm{soft}(arg)\mathrm{max}(\cdot)$



Fully connected (FC) layer

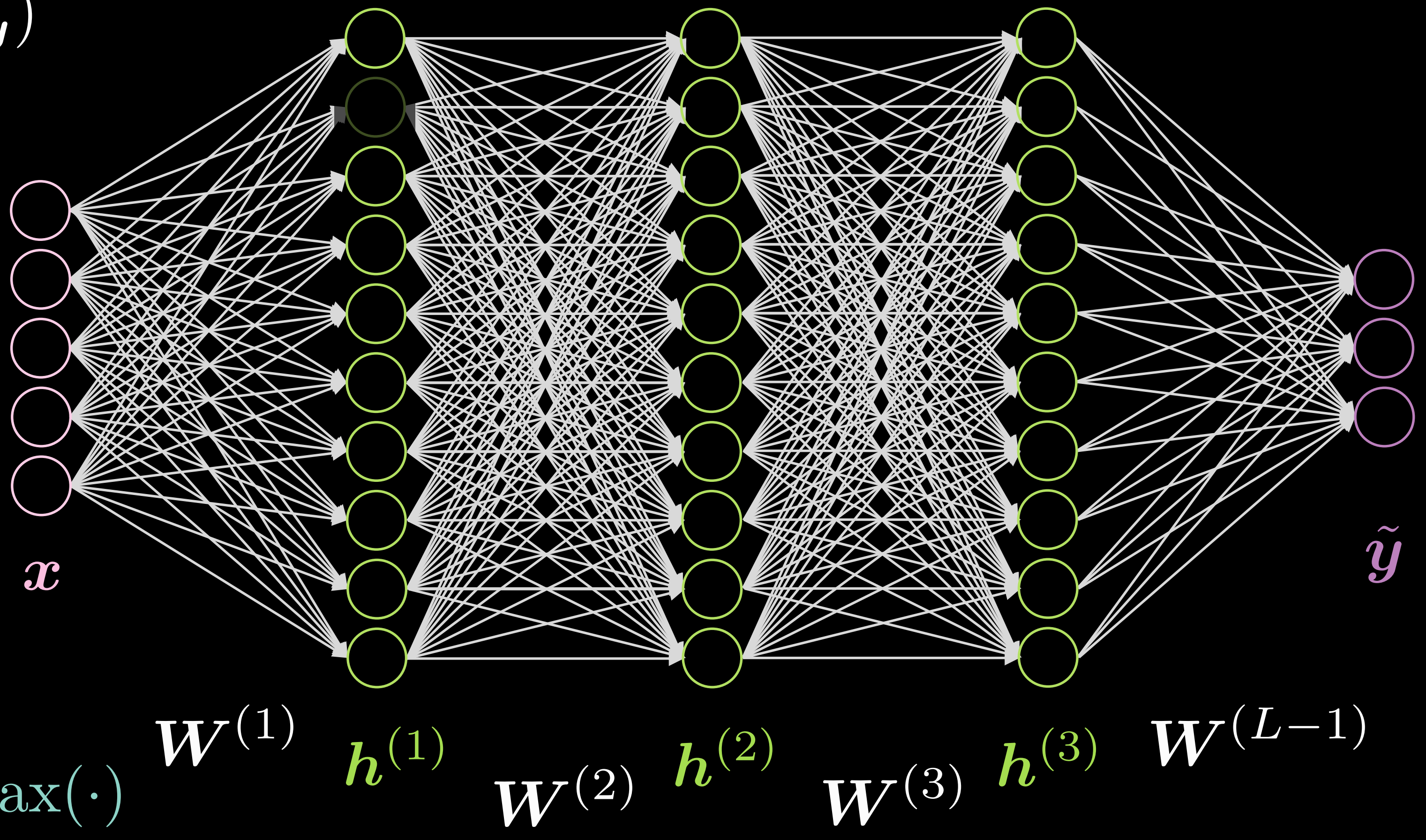
j-th row of $W^{(1)}$

$$h_j^{(1)} = f(\boxed{\boldsymbol{w}^{(j)}}\boldsymbol{x} + b_j) = f\left[\left(\sum_{i=1}^{n} w_i^{(j)} x_i\right) + b_j\right]$$



$$\boldsymbol{h} = f(\boldsymbol{W_h}\boldsymbol{x} + \boldsymbol{b_h})$$

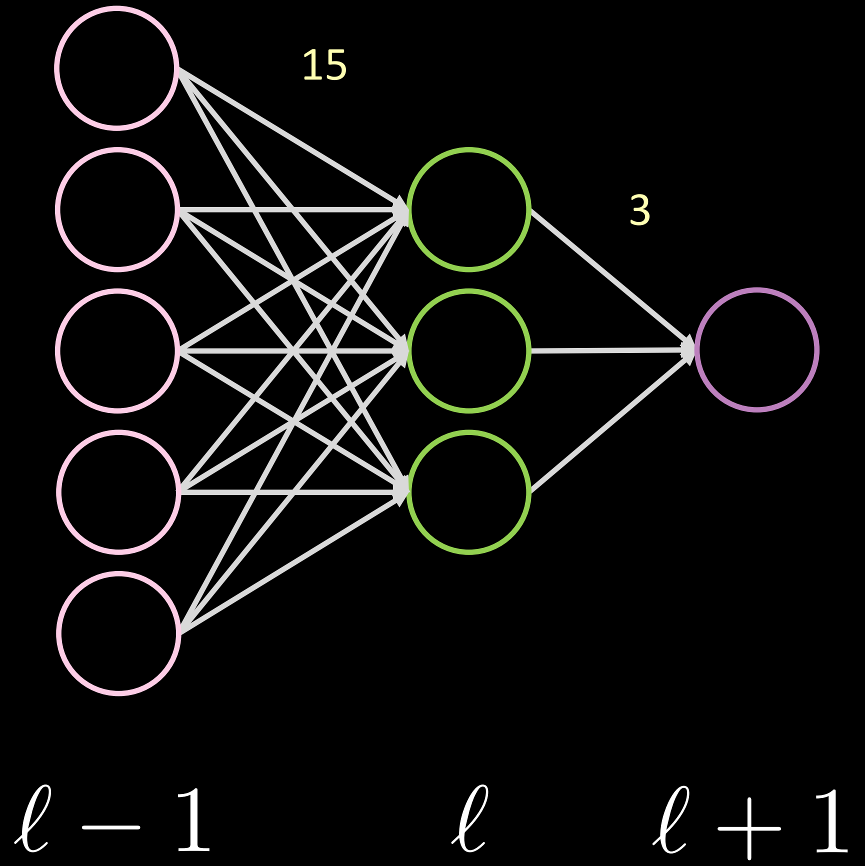
$$\tilde{\boldsymbol{y}} = g(\boldsymbol{W_y}\boldsymbol{h} + \boldsymbol{b_y})$$

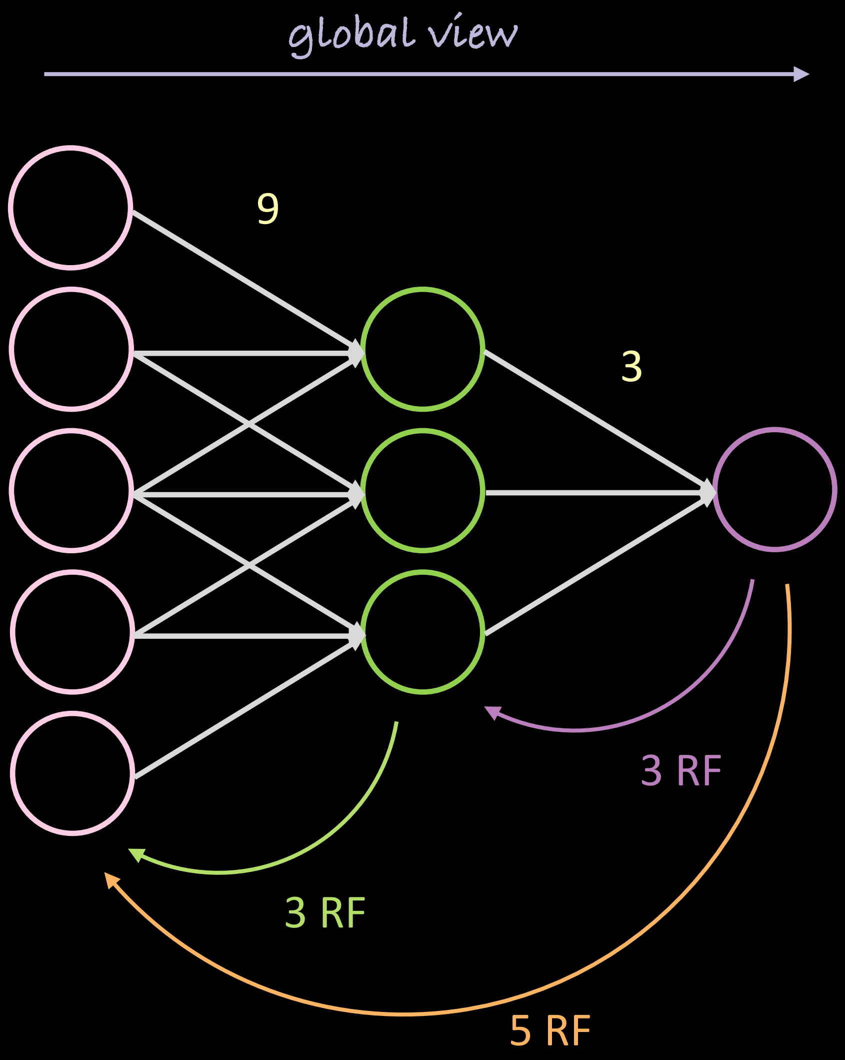


$f, g = (\cdot)^+, \sigma(\cdot),$
$\tanh(\cdot), \mathrm{soft}(arg)\max(\cdot)$

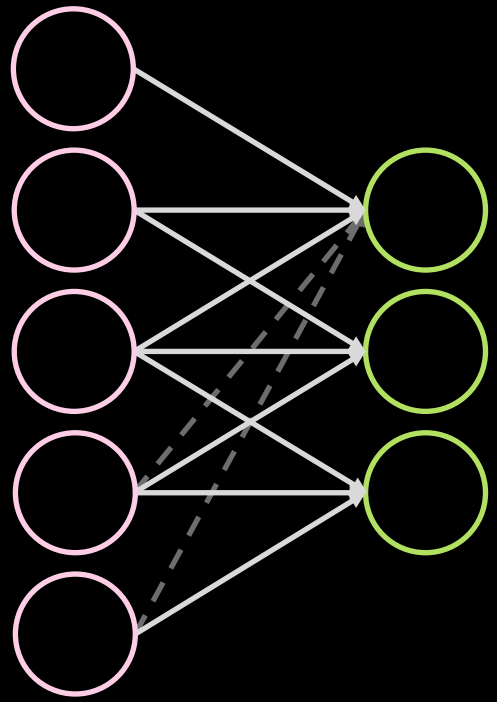
RF: receptive field

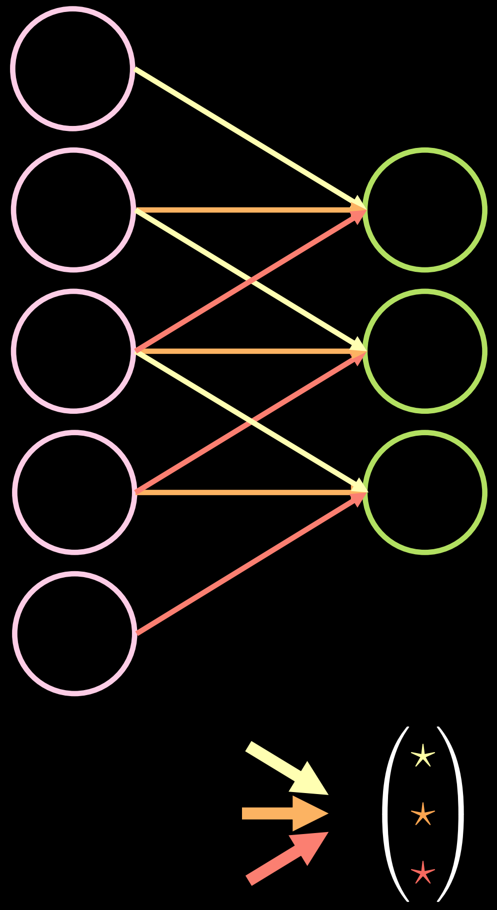
Locality ⇒ sparsity





Stationarity ⇒ parameters sharing





Parameters sharing

- faster convergence
- better generalisation
- not constrained to input size
- kernel independence
  ⇒ high parallelisation

Connection sparsity

- reduced amount of computation