

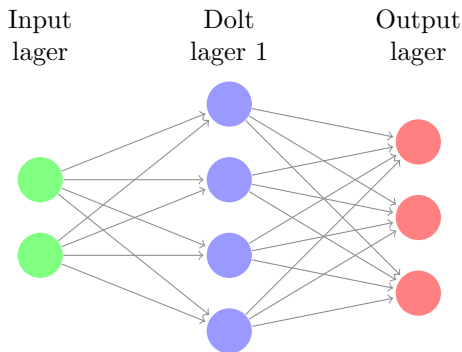
[www.math-stockholm.se/cirkel](http://www.math-stockholm.se/cirkel)

10 mars 2022



# Övning 6.1

Rita en bild av ett djupt neuralt nätverk med ett input lager med 2 noder, ett dolt lager med 4 noder och ett output lager med 3 noder.



Figur: Djupt neuralt nätverk

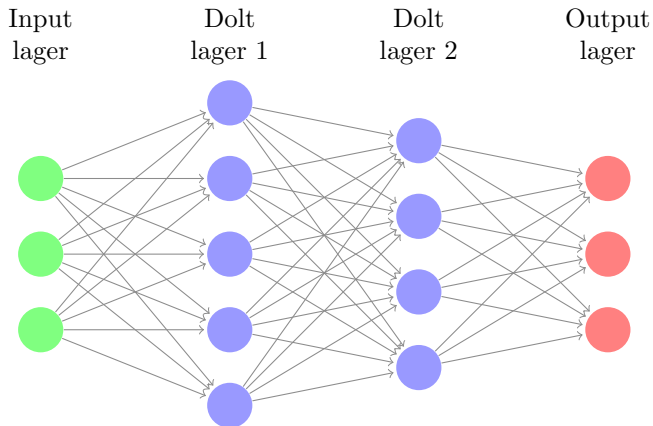
## Övning 6.2

Rita en bild av följande neurala nätverk. Nätverket har

- ett input lager
- 2 dolda lager
- ett output lager.

Nätverket tar in 3 ingångsnoder och ger 3 utgångsnoder. Det finns 5 noder i det första dolda lagret och 4 noder i det andra dolda lagret.

# Övning 6.2



Figur: Djupt neuralt nätverk

# Övning 6.3

Vi har

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} = \begin{bmatrix} 10 & 20 & 30 \\ -10 & -1 & -18 \\ 40 & 20 & 50 \\ -5 & -10 & -3 \end{bmatrix}, \quad \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} b_1^{(0)} \\ b_2^{(0)} \\ b_3^{(0)} \\ b_4^{(0)} \end{bmatrix} = \begin{bmatrix} -3 \\ -10 \\ 1 \\ -4 \end{bmatrix}$$

och aktiveringsfunktionen  $\sigma(x) = \frac{1}{1+e^{-x}}$ . Beräkna  $\left[ a_1^{(1)} a_2^{(1)} a_3^{(1)} a_4^{(1)} \right]^T$ .

# Övning 6.3

Vi gör så här. Först, beräknar vi

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix}$$

som

$$\begin{bmatrix} 10 & 20 & 30 \\ -10 & -1 & -18 \\ 40 & 20 & 50 \\ -5 & -10 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 20 \\ -13 \\ 90 \\ -22 \end{bmatrix}.$$

## Övning 6.3

Därefter beräknar vi

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix} + \begin{bmatrix} b_1^{(0)} \\ b_2^{(0)} \\ b_3^{(0)} \\ b_4^{(0)} \end{bmatrix}$$

som

$$\begin{bmatrix} 20 \\ -13 \\ 90 \\ -22 \end{bmatrix} + \begin{bmatrix} -3 \\ -10 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 17 \\ -23 \\ 91 \\ -26 \end{bmatrix}.$$

## Övning 6.3

Sedan evaluerar vi aktiveringsfunktionen  $\sigma$  på alla 4 element ovanför som

$$\sigma \left( \begin{bmatrix} 17 \\ -23 \\ 91 \\ -26 \end{bmatrix} \right) \approx \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} .$$



# Övning 6.4

Vi har

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} = \begin{bmatrix} -10 & -20 & -30 \\ 10 & 1 & 18 \\ 40 & 20 & 50 \\ 5 & 10 & 3 \end{bmatrix}, \quad \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}$$
$$\begin{bmatrix} b_1^{(0)} \\ b_2^{(0)} \\ b_3^{(0)} \\ b_4^{(0)} \end{bmatrix} = \begin{bmatrix} -3 \\ 10 \\ 1 \\ -4 \end{bmatrix}$$

och aktiveringsfunktionen  $\sigma(x) = \frac{1}{1+e^{-x}}$ . Beräkna  $\begin{bmatrix} a_1^{(1)} & a_2^{(1)} & a_3^{(1)} & a_4^{(1)} \end{bmatrix}^T$ .

# Övning 6.4

Först, beräknar vi

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix}$$

som

$$\begin{bmatrix} -10 & -20 & -30 \\ 10 & 1 & 18 \\ 40 & 20 & 50 \\ 5 & 10 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} -80 \\ 16 \\ 150 \\ 52 \end{bmatrix}.$$

# Övning 6.4

Därefter beräknar vi

$$\begin{bmatrix} w_{1,1}^{(0)} & w_{1,2}^{(0)} & w_{1,3}^{(0)} \\ w_{2,1}^{(0)} & w_{2,2}^{(0)} & w_{2,3}^{(0)} \\ w_{3,1}^{(0)} & w_{3,2}^{(0)} & w_{3,3}^{(0)} \\ w_{4,1}^{(0)} & w_{4,2}^{(0)} & w_{4,3}^{(0)} \end{bmatrix} \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix} + \begin{bmatrix} b_1^{(0)} \\ b_2^{(0)} \\ b_3^{(0)} \\ b_4^{(0)} \end{bmatrix}$$

som

$$\begin{bmatrix} -80 \\ 16 \\ 150 \\ 52 \end{bmatrix} + \begin{bmatrix} -3 \\ 10 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} -83 \\ 26 \\ 151 \\ 48 \end{bmatrix}.$$

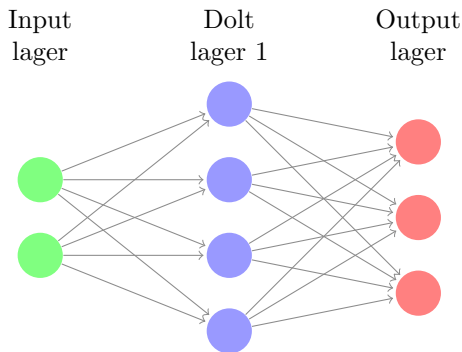
## Övning 6.4

Sedan evaluerar vi aktiveringsfunktionen  $\sigma$  på alla 4 element ovanför som

$$\sigma \left( \begin{bmatrix} -83 \\ 26 \\ 151 \\ 48 \end{bmatrix} \right) \approx \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} .$$

# Övning 6.5

Hur många parametrar måste vi beräkna för att träna modellen i Övning 6.1? Förklara ditt svar.



Figur: Djupt neuralt nätverk

## Övning 6.5

Vi måste beräkna

$$2 \times 4 + 4 = 12,$$

parametrar mellan inputlagret och det dolda lagret,

$$4 \times 3 + 3 = 15$$

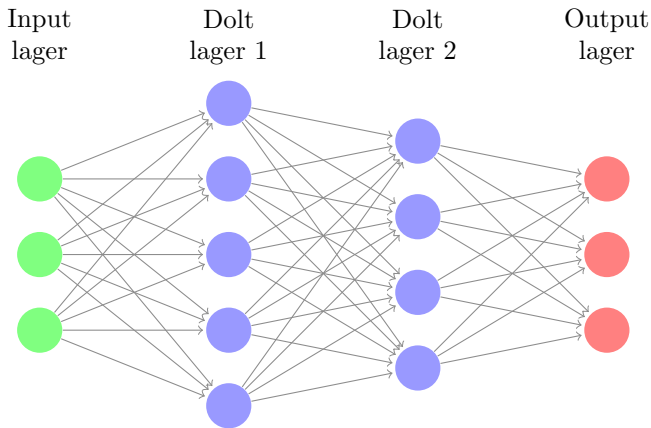
mellan det dolda lagret och outputlagret. Totalt blir det då

$$12 + 15 = 27$$

parametrar för att träna nätverket.

# Övning 6.6

Hur många parametrar måste vi beräkna för att träna modellen i Övning 6.2? Förklara ditt svar.



Figur: Djupt neuralt nätverk

## Övning 6.6

Vi måste beräkna

$$3 \times 5 + 5 = 20,$$

parametrar mellan inputlagret och det första dolda lagret,

$$5 \times 4 + 4 = 24$$

mellan det första dolda lagret och det andra dolda lagret och

$$4 \times 3 + 3 = 15$$

mellan det andra dolda lagret och outputlagret. Totalt blir det då

$$20 + 24 + 15 = 59$$

parametrar för att träna nätverket.



## Övning 6.7

Vi har en funktion  $f$  sådant att

$$f(x) = 2x^2 - x + 2.$$

Använd Algoritm 1 (gradientnedstigning) för att approximerar minimipunkten. Ta  $\gamma = 0.2$ ,  $m = 5$  och  $x^{(0)} = 2$ .

## Övning 6.7

Vi har att

$$\nabla f = 4x - 1.$$

Vi beräknar  $x^{(1)}$  som

$$\begin{aligned}x^{(1)} &= x^{(0)} - \gamma \nabla f(x^{(0)}) \\ &= 2 - .2 * 7 \\ &= .6\end{aligned}$$

Vi fortsätter på samma sätt och får

$$\begin{aligned}x^{(2)} &= x^{(1)} - \gamma \nabla f(x^{(1)}) \\ x^{(3)} &= x^{(2)} - \gamma \nabla f(x^{(2)}) \\ x^{(4)} &= x^{(3)} - \gamma \nabla f(x^{(3)}) \\ x^{(5)} &= x^{(4)} - \gamma \nabla f(x^{(4)})\end{aligned}$$

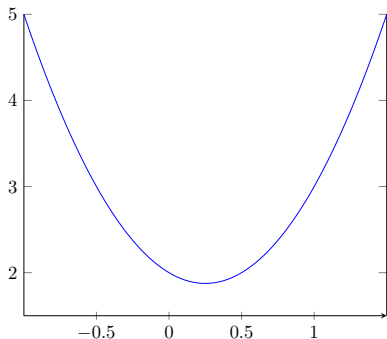
och att  $x^{(5)} = 0.25056$ .

## Övning 6.7

Då har vi att

$$(x^{(\min)}, f(x^{(\min)})) \approx (0.25056, 1.87500).$$

Vi ser en graf av  $f$  nedanför.



## Övning 6.8

Vi har en funktion  $f$  sådan att

$$f(x) = 4x^2 + 2x + 2.$$

Använd Algoritm 1 (gradientnedstigning) för att approximerar minimipunkten. Ta  $\gamma = 0.1$ ,  $m = 6$  och  $x^{(0)} = 0$ .

## Övning 6.8

Vi har att

$$\nabla f = 8x + 2.$$

Vi beräknar  $x^{(1)}$  som

$$\begin{aligned}x^{(1)} &= x^{(0)} - \gamma \nabla f(x^{(0)}) \\ &= 0 - .1 * 2 \\ &= -.2\end{aligned}$$

Vi fortsätter på samma sätt och får

$$\begin{aligned}x^{(2)} &= x^{(1)} - \gamma \nabla f(x^{(1)}) \\ x^{(3)} &= x^{(2)} - \gamma \nabla f(x^{(2)}) \\ x^{(4)} &= x^{(3)} - \gamma \nabla f(x^{(3)}) \\ x^{(5)} &= x^{(4)} - \gamma \nabla f(x^{(4)}) \\ x^{(6)} &= x^{(5)} - \gamma \nabla f(x^{(5)})\end{aligned}$$

## Övning 6.8

och att  $x^{(6)} \approx -0.249984$ . Då har vi att

$$(x^{(\min)}, f(x^{(\min)})) \approx (-0.249984, 1.750000).$$

Vi ser en graf av  $f$  nedanför.

