Session 1&2

Backpropagation in NN

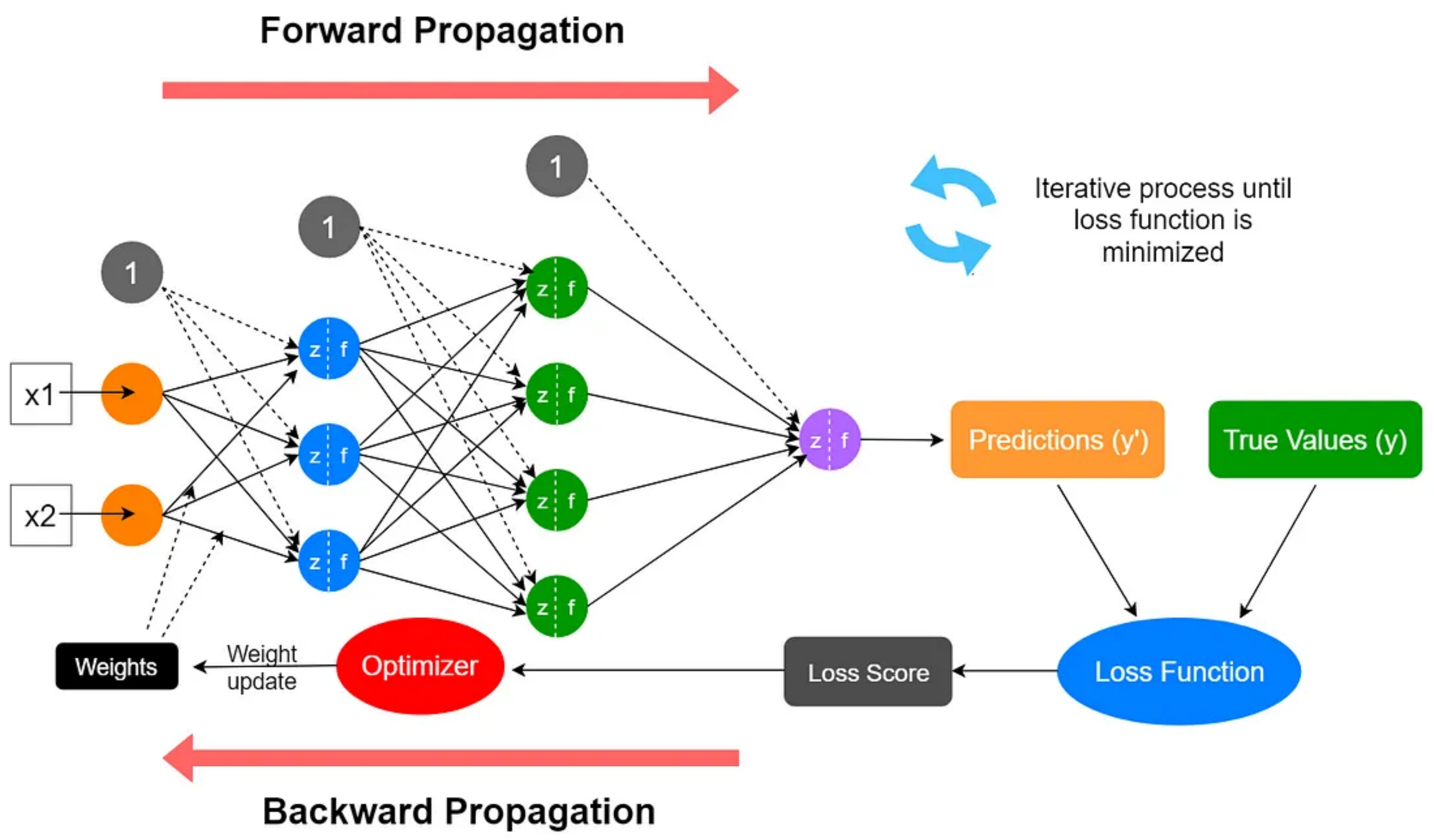
Deep Learning | Zahra Amini

Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

https://zil.ink/zahraamini

یادگیری در نورون مصنوعی چیست؟

یادگیری در نورون مصنوعی فرآیندی است که در آن مدل با تنظیم پارامترهای داخلی خود از طریق به‌روزرسانی وزن‌ها و بهینه‌سازی تابع خطا، توانایی تشخیص و پیش‌بینی الگوها در داده‌های ورودی را به‌دست می‌آورد



Gradient

گرادیان در ریاضیات و به ویژه در زمینه بهینه‌سازی و یادگیری ماشین، نشان‌دهنده شیب یا نرخ تغییرات یک تابع در یک نقطه خاص است. به عبارت ساده‌تر، گرادیان برداری است که نشان می‌دهد چگونه مقدار تابع در یک نقطه خاص تغییر می‌کند و در چه جهتی این تغییرات بیشترین مقدار را دارد

Gradiant $\longrightarrow$ $\nabla f(x,y) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y}$

$$f(x,y) = x^2 + y^2 \longrightarrow \nabla f = 2x\hat{x} + 2y\hat{y}$$

مشتق جزئی $\longrightarrow$ $f(x,y) = x^2 + y^2$ $\longrightarrow$ $\frac{\partial f}{\partial x} = 2x$

$\longrightarrow$ $\frac{\partial f}{\partial y} = 2y$

$$\begin{cases} y = 8z^3 + z^2 + 5 \\ z = x^5 + 9 \end{cases} \longrightarrow \frac{dy}{dx} = ? \longrightarrow$$ مشتق زنجیره‌ای

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z}\,\frac{\partial z}{\partial x}$$

Gradient Descent

1. مقداردهی اولیه پارامترها ← W ها را به صورت رندوم مقداردهی می‌کنیم

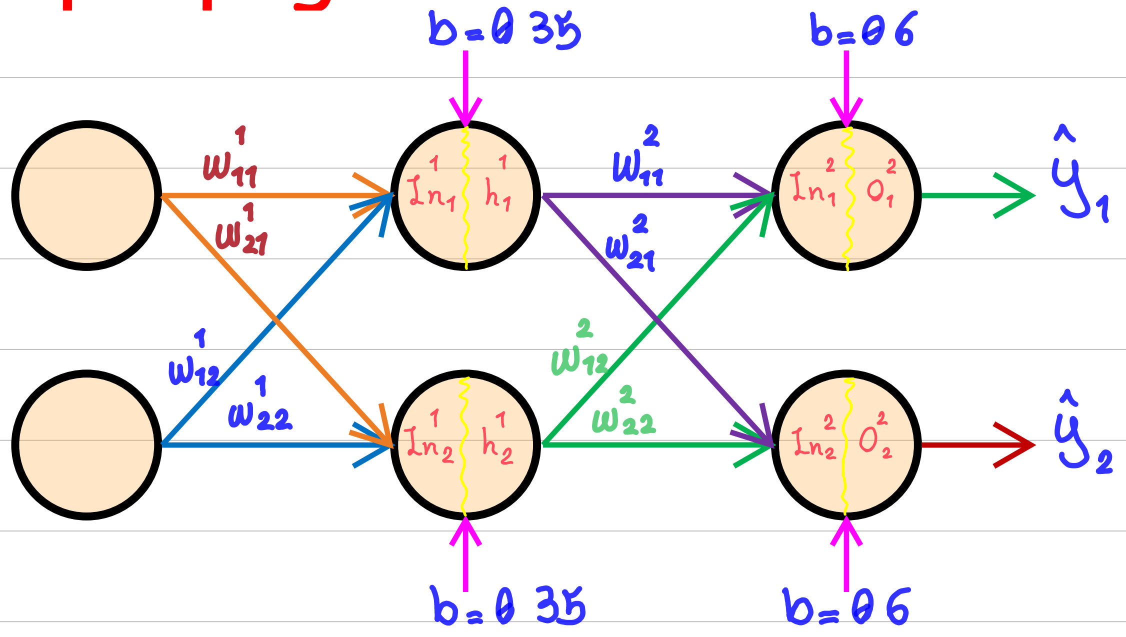
2. محاسبه گرادیان ← از تابع Loss نسبت به W مشتق می‌گیریم.

3. آپدیت وزنها ← $W_{new} = W_{old} - \alpha \dfrac{\partial L}{\partial W_{old}}$

4. شرط خاتمه ← مراحل 3,2 را تا رسیدن به Minimum یا Itration مشخص شده ادامه می‌دهیم

Backpropagation in MLP

:



$x_1 = 0.05, \quad x_2 = 0.1$

$$W^1 = \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} \qquad W^2 = \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix}$$

$b^1 = 0.35, \; b^2 = 0.6 \qquad y_1 = 0.01, \; y_2 = 0.99$

Activation Function $\quad f^1 = \text{ReLU} \quad f^2 = \text{Sigmoid}$

? Update W^1_{11} ⟿ SGD $\qquad \alpha = 0.1$

1. Feed Forward
2. Loss Function
3. Back Propagation

1. Feed Forward:

$$In^1_1 = \overset{0.15}{W^1_{11}} x_1 + \overset{0.2}{W^1_{12}} x_2 + \overset{0.35}{b^1} = 0.3375 \longrightarrow h^1_1 = \text{ReLU}(In^1_1) = 0.3375$$

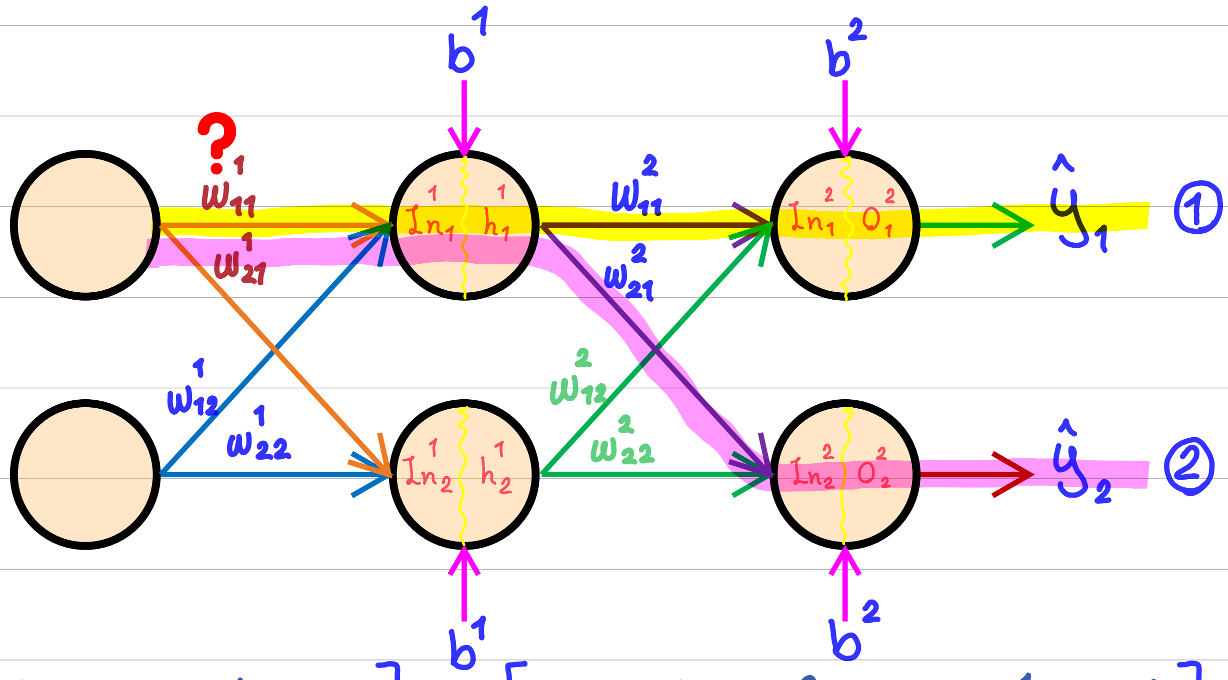
$$In^1_2 = \overset{0.25}{W^1_{21}} x_1 + \overset{0.3}{W^1_{22}} x_2 + \overset{0.35}{b^1} = 0.3925 \longrightarrow h^1_2 = \text{ReLU}(In^1_2) = 0.3925$$

$$\overset{=0.92}{In^2_1 = \overset{0.4}{W^2_{11}} h^1_1 + \overset{0.45}{W^2_{12}} h^1_2 + \overset{0.6}{b^2}} \longrightarrow \overset{=0.92}{O^2_1 = \text{Sigmoid}(In^2_1) = \hat{y}_1}$$

$$\underset{=1.004}{In^2_2 = \underset{0.5}{W^2_{21}} h^1_1 + \underset{0.55}{W^2_{22}} h^1_2 + b^2} \longrightarrow \underset{=1.004}{O^2_2 = \text{Sigmoid}(In^2_2) = \hat{y}_2}$$

② Loss Function $\longrightarrow$ $MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$ SGD ~ ازای هر Sample که وارد شبکه می شود، یکبار وزنها را آپدیت می کند

③ Back Propagation



$$\frac{\partial L}{\partial w^1_{11}} = ① + ② = \left[\frac{\partial L}{\partial \hat{y}_1}\,\frac{\partial \hat{y}_1}{\partial O^2_1}\,\frac{\partial O^2_1}{\partial In^2_1}\,\frac{\partial In^2_1}{\partial h^1_1}\,\frac{\partial h^1_1}{\partial In^1_1}\,\frac{\partial In^1_1}{\partial w^1_{11}}\right] + \left[\frac{\partial L}{\partial \hat{y}_2}\,\frac{\partial \hat{y}_2}{\partial O^2_2}\,\frac{\partial O^2_2}{\partial In^2_2}\,\frac{\partial In^2_2}{\partial h^1_1}\,\frac{\partial h^1_1}{\partial In^1_1}\,\frac{\partial In^1_1}{\partial w^1_{11}}\right]$$

$$\frac{\partial L}{\partial w^1_{11}} = \left[\frac{\partial L}{\partial \hat{y}_1}\,\frac{\partial \hat{y}_1}{\partial O^2_1}\,\frac{\partial O^2_1}{\partial In^2_1}\,\frac{\partial In^2_1}{\partial h^1_1} + \frac{\partial L}{\partial \hat{y}_2}\,\frac{\partial \hat{y}_2}{\partial O^2_2}\,\frac{\partial O^2_2}{\partial In^2_2}\,\frac{\partial In^2_2}{\partial h^1_1}\right]\left[\frac{\partial h^1_1}{\partial In^1_1}\,\frac{\partial In^1_1}{\partial w^1_{11}}\right]$$

$* \ (ab + b.c) = b(a+c)$

$$\frac{\partial L}{\partial W_{11}^{1}} = \left[\frac{\partial L}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial O_1^2}\frac{\partial O_1^2}{\partial In_1^2}\frac{\partial In_1^2}{\partial h_1^1} + \frac{\partial L}{\partial \hat{y}_2}\frac{\partial \hat{y}_2}{\partial O_2^2}\frac{\partial O_2^2}{\partial In_2^2}\frac{\partial In_2^2}{\partial h_1^1}\right]\left[\frac{\partial h_1^1}{\partial In_1^1}\frac{\partial In_1^1}{\partial W_{11}^1}\right]$$

① Feed Forward:

$In_1^1 = W_{11}^1 x_1 + W_{12}^1 x_2 + b^1 \rightarrow h_1^1 = ReLU(In_1^1)$

$In_2^1 = W_{21}^1 x_1 + W_{22}^1 x_2 + b^1 \rightarrow h_2^1 = ReLU(In_2^1)$

$In_1^2 = W_{11}^2 h_1^1 + W_{12}^2 h_2^1 + b^2 \rightarrow O_1^2 = Sigmoid(In_1^2) = \hat{y}_1$

$In_2^2 = W_{21}^2 h_1^1 + W_{22}^2 h_2^1 + b^2 \rightarrow O_2^2 = Sigmoid(In_2^2) = \hat{y}_2$

$$\frac{\partial In_1^1}{\partial W_{11}^1} = \frac{\partial}{\partial W_{11}^1}\left(W_{11}^1 x_1 + \underbrace{\cancel{W_{12}^1 x_2 + b^1}}_{0}\right) = x_1$$

$$\frac{\partial In_1^2}{\partial h_1^1} = \frac{\partial}{\partial h_1^1}\left(W_{11}^2 h_1^1 + \underbrace{\cancel{W_{12}^2 h_2^1 + b^2}}_{0}\right) = W_{11}^2$$

$$\frac{\partial h_1^1}{\partial In_1^1} = \frac{\partial}{\partial In_1^1}\left(ReLU(In_1^1)\right) = \begin{cases} 1 & In_1^1 > 0 \\ 0 & In_1^1 < 0 \\ \text{Undefind} & In_1^1 = 0 \end{cases}$$

$$* \; ReLU(x) = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases}$$

$$\frac{\partial O_1^2}{\partial In_1^2} = \frac{\partial}{\partial In_1^2}\left(Sigmoid(In_1^2)\right) = \sigma(In_1^2) \times (1 - \sigma(In_1^2))$$

$$\frac{\partial In_2^2}{\partial h_1^1} = \frac{\partial}{\partial h_1^1}\left(W_{21}^2 h_1^1 + \underbrace{\cancel{W_{22}^2 h_2^1 + b^2}}_{0}\right) = W_{21}^2$$

$$\frac{\partial \hat{y}_1}{\partial O_1^2} = 1$$

$$\frac{\partial O_2^2}{\partial In_2^2} = \frac{\partial}{\partial In_2^2}\left(Sigmoid(In_2^2)\right) = \sigma(In_2^2) \times (1 - \sigma(In_2^2))$$

$$\frac{\partial L}{\partial \hat{y}_1} = \frac{\partial}{\partial \hat{y}_2}\left(\frac{1}{m}\sum (y - \hat{y})^2\right) = -\frac{2}{m}\sum (y - \hat{y})$$

$$\frac{\partial \hat{y}_2}{\partial O_2^2} = 1$$

$$\frac{\partial L}{\partial \hat{y}_2} = \frac{\partial}{\partial \hat{y}_2}\left(\frac{1}{m}\sum (y - \hat{y})^2\right) = -\frac{2}{m}\sum (y - \hat{y})$$

$* \; (f^2)' = f . f'$

$$\frac{\partial L}{\partial w_{11}^{1}} = \left[\underset{-0.249}{\frac{\partial L}{\partial \hat{y}_1}} \; \underset{1}{\frac{\partial \hat{y}_1}{\partial O_1^{2}}} \; \underset{0.187}{\frac{\partial O_1^{2}}{\partial In_1^{2}}} \; \underset{0.4}{\frac{\partial In_1^{2}}{\partial h_1^{1}}} + \underset{-0.217}{\frac{\partial L}{\partial \hat{y}_2}} \; \underset{1}{\frac{\partial \hat{y}_2}{\partial O_2^{2}}} \; \underset{0.174}{\frac{\partial O_2^{2}}{\partial In_2^{2}}} \; \underset{0.55}{\frac{\partial In_2^{2}}{\partial h_1^{1}}} \right] \left[\underset{0.3375}{\frac{\partial h_1^{1}}{\partial In_1^{1}}} \; \underset{0.05}{\frac{\partial In_1^{1}}{\partial w_{11}^{1}}} \right] \simeq -0.0006647$$

$$\frac{\partial In_1^{1}}{\partial w_{11}^{1}} = \frac{\partial}{\partial w_{11}^{1}}\left(w_{11}^{1} x_1 + \overset{0}{\cancel{w_{12}^{1} x_2 + b^{1}}}\right) = x_1 = 0.05$$

$$\frac{\partial In_1^{2}}{\partial h_1^{1}} = \frac{\partial}{\partial h_1^{1}}\left(w_{11}^{2} h_1^{1} + \overset{0}{\cancel{w_{12}^{2} h_2^{1} + b^{2}}}\right) = w_{11}^{2} = 0.4$$

$$\frac{\partial h_1^{1}}{\partial In_1^{1}} = \frac{\partial}{\partial In_1^{1}}\left(ReLU(In_1^{1})\right) = \begin{cases} 1 & In_1^{1} > 0 \\ 0 & In_1^{1} < 0 \\ \text{Undefind} & In_1^{1} = 0 \end{cases} = 0.3375$$

$$\frac{\partial O_1^{2}}{\partial In_1^{2}} = \frac{\partial}{\partial In_1^{2}}\left(Sigmoid(In_1^{2})\right) = \sigma(In_1^{2}) \times \left(1 - \sigma(In_1^{2})\right) = 0.187$$

$$\frac{\partial In_2^{2}}{\partial h_1^{1}} = \frac{\partial}{\partial h_1^{1}}\left(w_{21}^{2} h_1^{1} + \overset{0}{\cancel{w_{22}^{2} h_2^{1} + b^{2}}}\right) = w_{21}^{2} = 0.55$$

$$\frac{\partial \hat{y}_1}{\partial O_1^{2}} = 1$$

$$\frac{\partial O_2^{2}}{\partial In_2^{2}} = \frac{\partial}{\partial In_2^{2}}\left(Sigmoid(In_2^{2})\right) = \sigma(In_2^{2}) \times \left(1 - \sigma(In_2^{2})\right) = 0.174$$

$$\frac{\partial L}{\partial \hat{y}_1} = \frac{\partial}{\partial \hat{y}_2}\left(\frac{1}{m}\sum (y - \hat{y})^2\right) = -\frac{\overset{-1}{2}}{m}\sum (\overset{1}{y} - \overset{0.751}{\hat{y}}) = -0.249$$

$$\frac{\partial \hat{y}_2}{\partial O_2^{2}} = 1$$

$$\frac{\partial L}{\partial \hat{y}_2} = \frac{\partial}{\partial \hat{y}_2}\left(\frac{1}{m}\sum (y - \hat{y})^2\right) = -\frac{2}{m}\sum (y - \hat{y}) = -0.217$$

$$W_{new} = W_{old} - \alpha \frac{\partial L}{\partial W_{old}}$$

$$W_{11\,new}^{1} = 0.15 - \left(0.1 \times (-0.0006647)\right) = 0.15006647$$

پارامترها

پارامترها مقادیر داخلی یک مدل یادگیری ماشین هستند که در طی فرایند آموزش مدل از داده‌ها، یاد گرفته(تنظیم) می‌شوند. این پارامترها معمولاً وزن‌ها و بایاس‌ها در مدل‌های شبکه عصبی یا ضرایب در مدل‌های رگرسیون هستند. پارامترها مستقیماً از داده‌های آموزشی استخراج می‌شوند و مقادیرشان به طور خودکار تنظیم می‌شود تا مدل بتواند عملکرد بهتری روی داده‌های آموزشی داشته باشد

وزن‌ها در شبکه‌های عصبی (weights)

بایاس‌ها در شبکه‌های عصبی (biases)

هایپرپارامترها

هایپرپارامترها مقادیری هستند که قبل از فرایند آموزش مدل تعیین می‌شوند و بر روی فرایند آموزش و ساختار مدل تأثیر می‌گذارند. هایپرپارامترها توسط انسان یا به کمک روش‌های جستجوی هایپرپارامتر مانند گرید سرچ یا جستجوی تصادفی تنظیم می‌شوند. تنظیم درست هایپرپارامترها می‌تواند تأثیر بزرگی بر عملکرد نهایی مدل داشته باشد

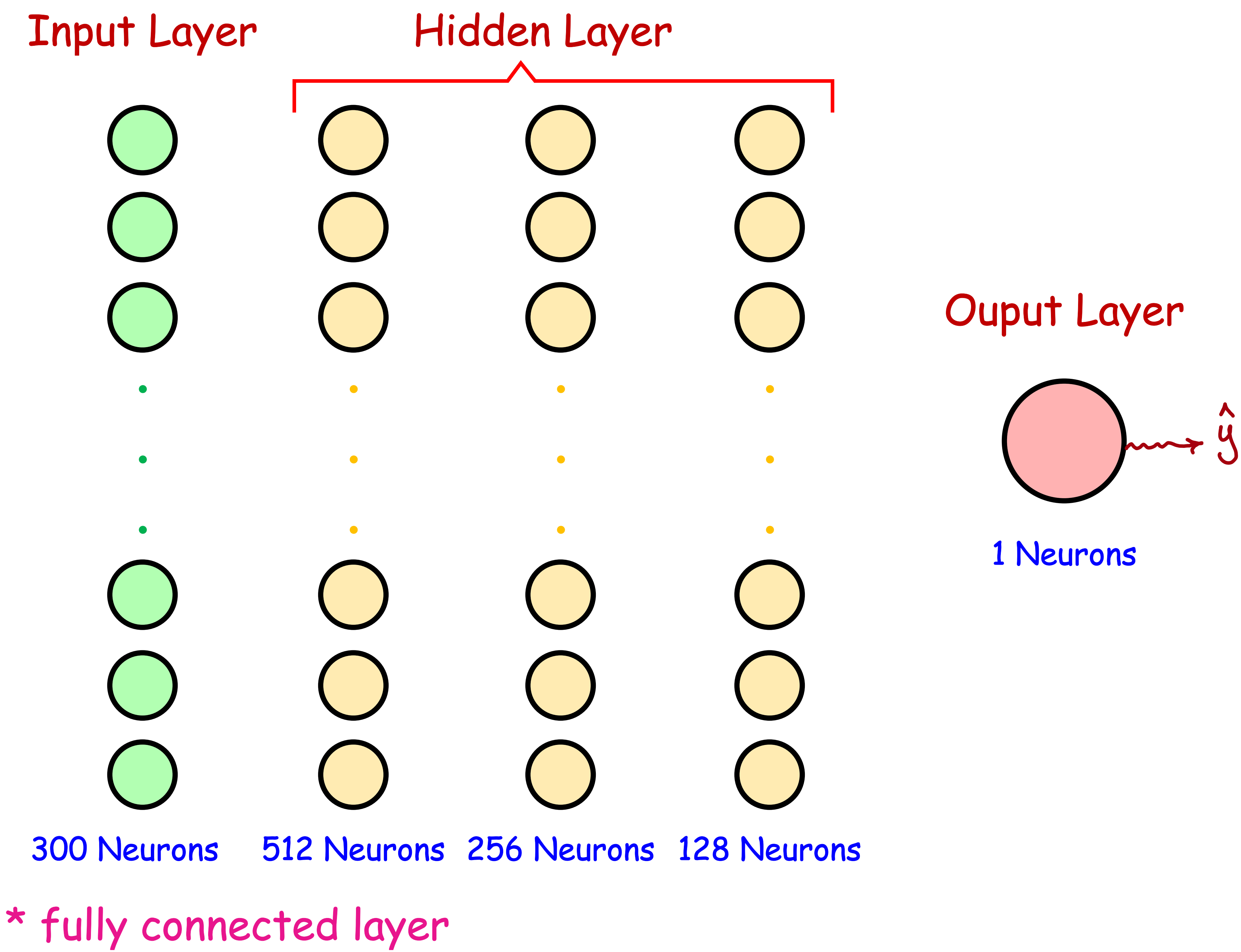
مثال‌هایی از هایپرپارامترها

نرخ یادگیری (learningrate)

تعداد لایه‌ها و نورون‌ها در هر لایه (numberofneuronsperlayer و numberoflayers)

اندازه بچ، تعداد نمونه‌هایی که قبل از به‌روزرسانی وزن‌ها در طول یک مرحله آموزش پردازش می‌شوند (batchsize)

تعداد دوره‌های آموزش، تعداد تکرارهای کامل روی داده‌های آموزشی در طول آموزش (numberofepochs)


Input Layer
Hidden Layer
Ouput Layer
$\hat{y}$
1 Neurons
300 Neurons
512 Neurons
256 Neurons
128 Neurons
* fully connected layer