شبکه‌های عصبی پیشرو

dayche.com . گروه دایکه

دای‌کِه DAYCHE

شبکه‌های عصبی پیشرو

- مدل‌های یادگیری ماشین
  - تقریب تابع

$$y = f(x) \quad \Rightarrow \quad \hat{y} = \hat{f}(x;\theta)$$

- شبکه‌های عصبی پیشرو
  - ابزاری قدرتمند به منظور توسعه مدل‌های یادگیری ماشین
  - نحوه آموزش شبکه‌ها و مشکلات

تولید محتوا: وحید محمدزاده ایوقی

شبکه‌های عصبی پیشرو

- شبکه‌های عصبی پیشرو
  - پیشرو – استنتاج تنها به متغیر ورودی بستگی دارد و هیچ ارتباط فیدبکی موثر نیست.

$$y = f(x_t, x_{t-1}, \dots, x_0; \theta)$$

  - شبکه – خروجی در اثر اعمال توابع غیرخطی پی در پی حاصل می‌شود. یک گراف جهت‌دار بدون دور

$$y = \underbrace{f^3}_{\text{لایه خروجی}}\left(f^2\left(f^1(x)\right)\right)$$

هر تابع معرف یک لایه است

  - عمق و عرض شبکه

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شبکه‌های عصبی پیشرو

- شبکه‌های عصبی پیشرو
  - عصبی – الهام گرفته شده از ساختار عصبی مغز از جهت اینکه هر نرون متغیر ورودی خود را از سایر نرون‌ها می‌گیرد و پردازش می‌کند.

- انتقال اطلاعات از یک نرون به نرون دیگر بر مبنای مقایسه بین سیگنال الکتریکی تولید شده توسط نرون و یک سطح آستانه صورت می‌پذیرد.
- اطلاعات منتقل شده به یک نرون، توسط مجموعه‌ای از نرون‌ها فراهم می‌شود که میزان مشارکت هر نرون از این مجموعه با یک وزن مشخص می‌شود.
- فرآیند یادگیری به تقویت وزن‌های ارتباطی می‌پردازد.

تولید محتوا: وحید محمدزاده ایوقی

شبکه‌های عصبی پیشرو

- شبکه‌های عصبی پیشرو
  - هر نمونه از متغیرهای ورودی مقدار مشخصی را در لایه‌های خروجی نتیجه می‌دهد – لایه‌های مشاهده پذیر
  - هر نمونه از متغیرهای ورودی مقدار نامشخصی را در لایه‌های میانی نتیجه می‌دهد – لایه‌های مخفی
  - فرآیند آموزش یک شبکه عصبی، تغییر خروجی لایه‌های مخفی به شکل مطلوب است.

- فرآیند آموزش شبکه عصبی پیشرو
  - مشابه سایر مدل‌های یادگیر نیاز است تا فرم توابع، روش بهینه‌سازی و تابع هزینه مشخص شده باشد.

فلسفه استفاده از شبکه‌های عصبی پیشرو

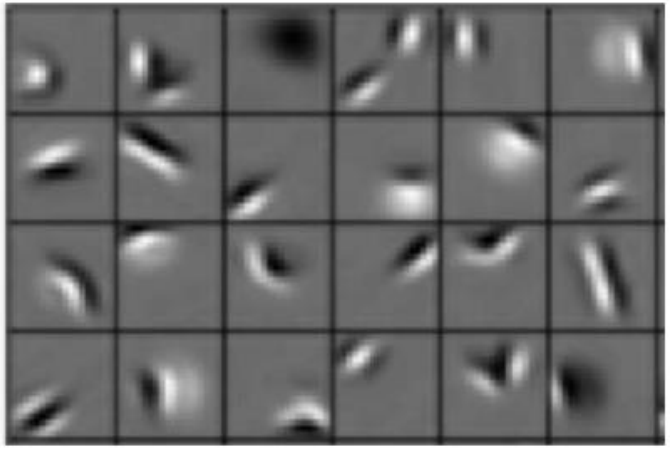
- تقریب تابع
  - هدف مدل کردن دقیق ساختار مغز نیست و این شبکه‌ها الگویی مشابه ساختار مغز برای تقریب توابع غیرخطی ارائه می‌کنند.
- مدل‌های خطی – رگرسیون خطی و رگرسیون لوجستیک
  - ارتباط غیرخطی بین ورودی و خروجی

$$y = \phi(x;\theta)W \quad \Rightarrow \quad \phi(x;\theta) = ?$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مثال‌هایی از شبکه‌های عصبی پیشرو

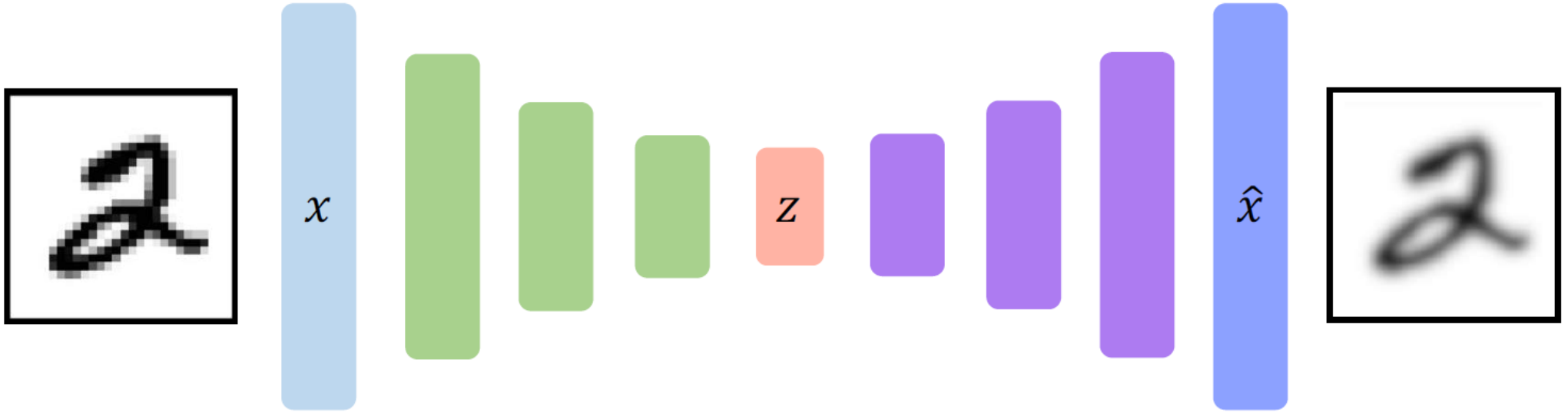
- شبکه‌های عصبی کانولوشنی



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

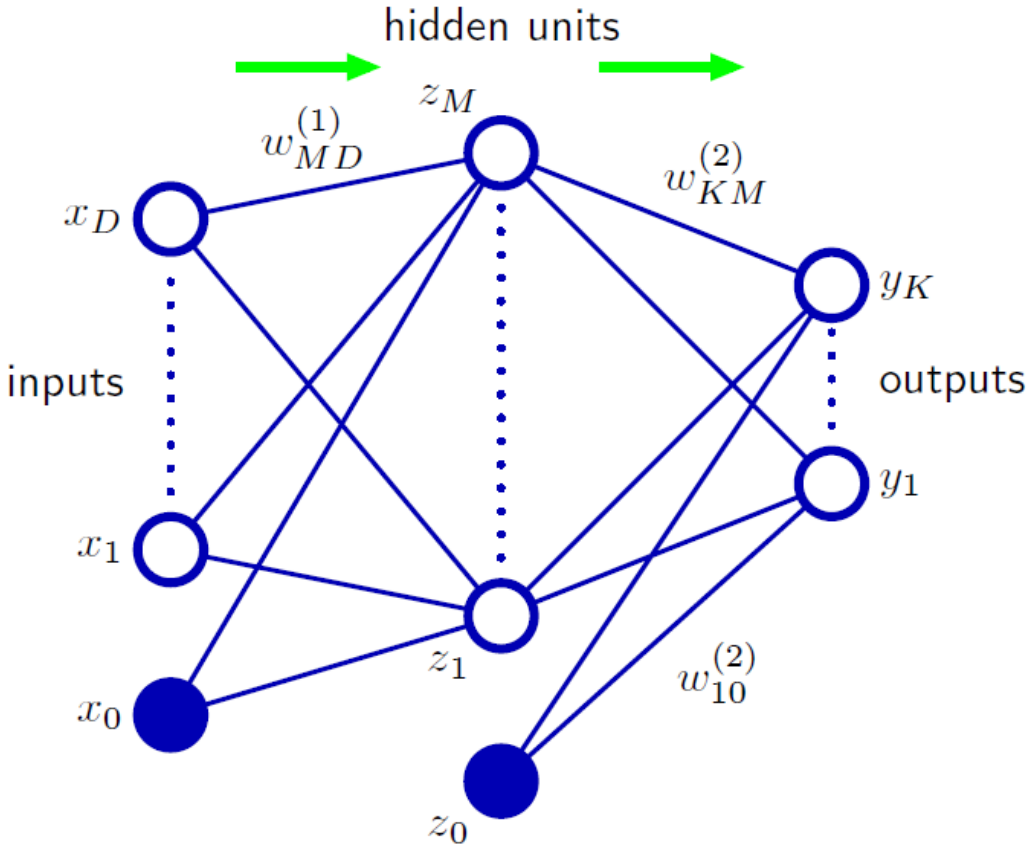
مثال‌هایی از شبکه‌های عصبی پیشرو

- شبکه‌های عصبی رمزگذار خودکار



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

ساختار شبکه عصبی



- ساختار شبکه عصبی

$$Z = f^1(XW^1 + b^1)$$

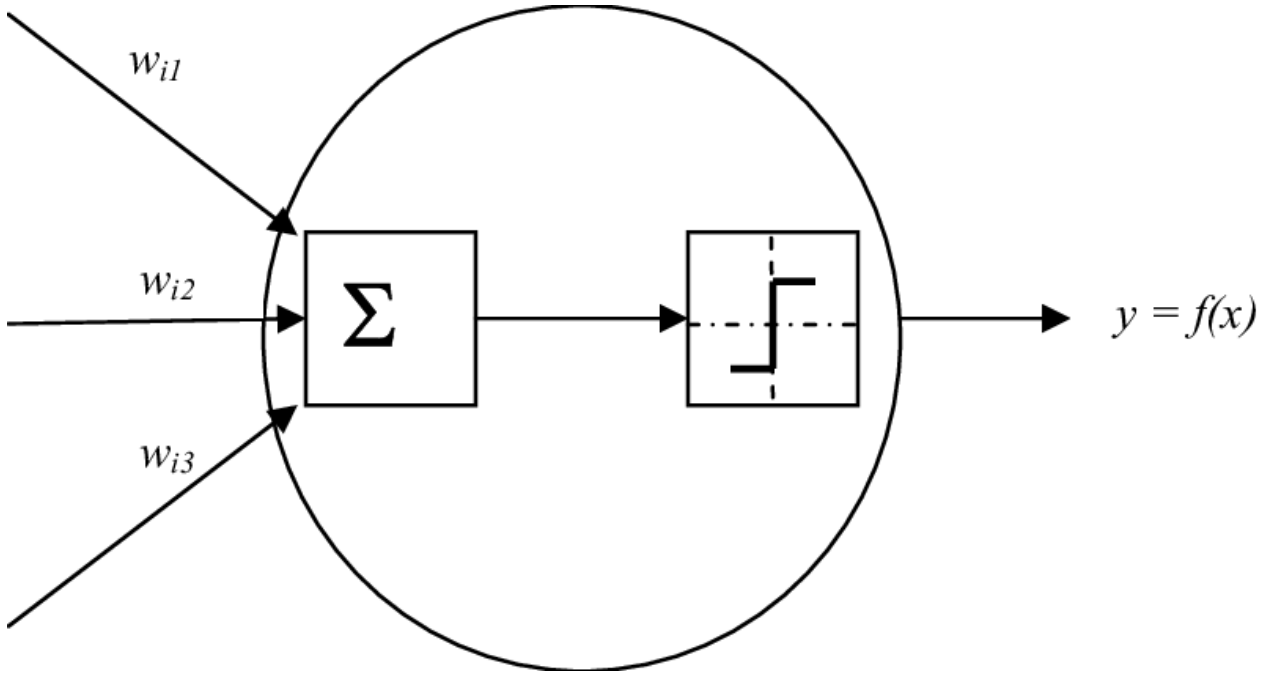
$$Y = f^2(ZW^2 + b^2)$$

- مدل نرون

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مدل نرون

- اولین مدل نرون – نرون McCulloch Pitts

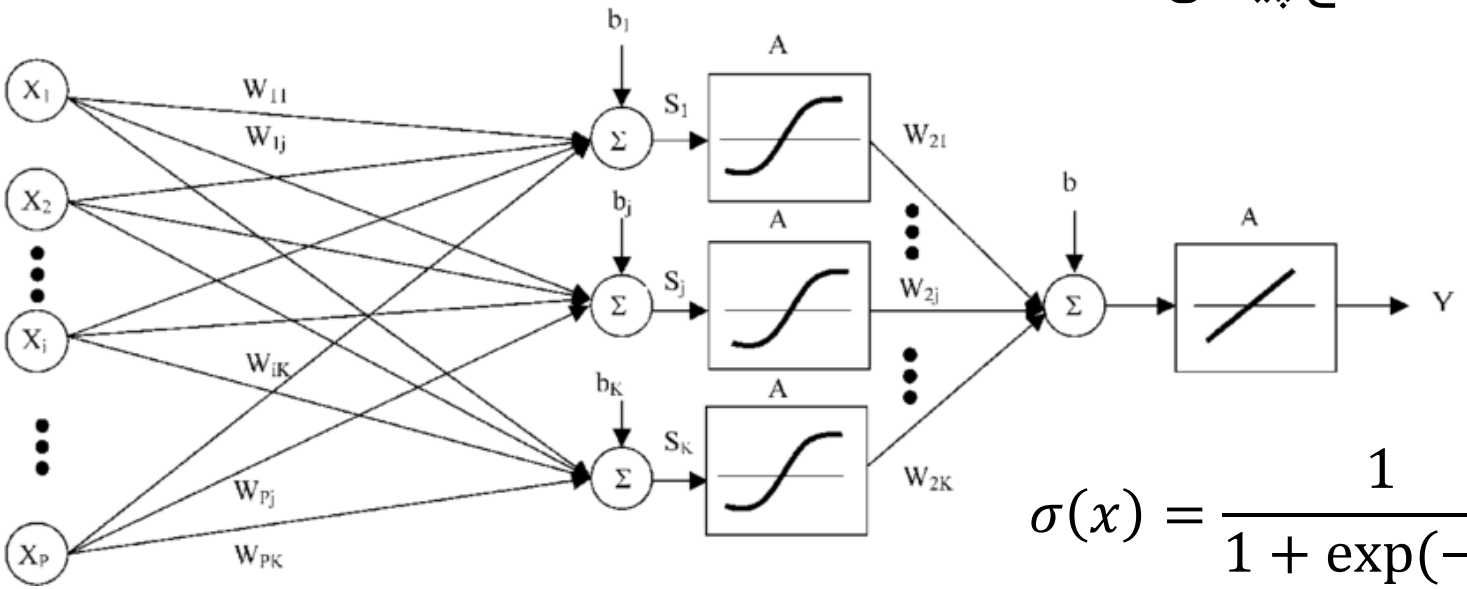


- ایرادات این مدل
  - مدل نرون مشتق‌پذیر نیست.

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مدل نرون

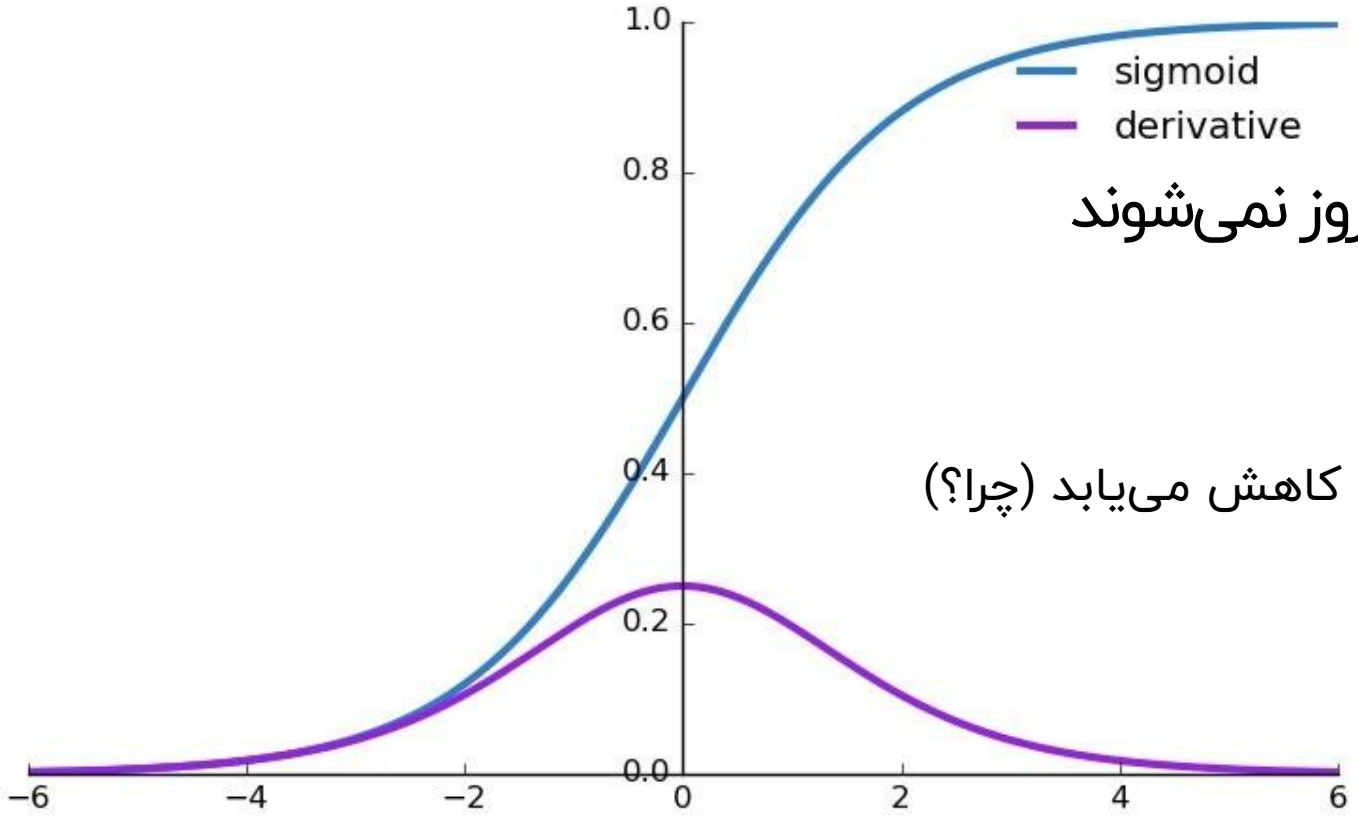
- مدل سیگموئید
- تقریبی هموار از مدل نرون مک‌کلاچ پیتس



$$\sigma(x) = \frac{1}{1+\exp(-x)}, \sigma'(x) = \sigma(x)\big(1-\sigma(x)\big)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مدل نرون



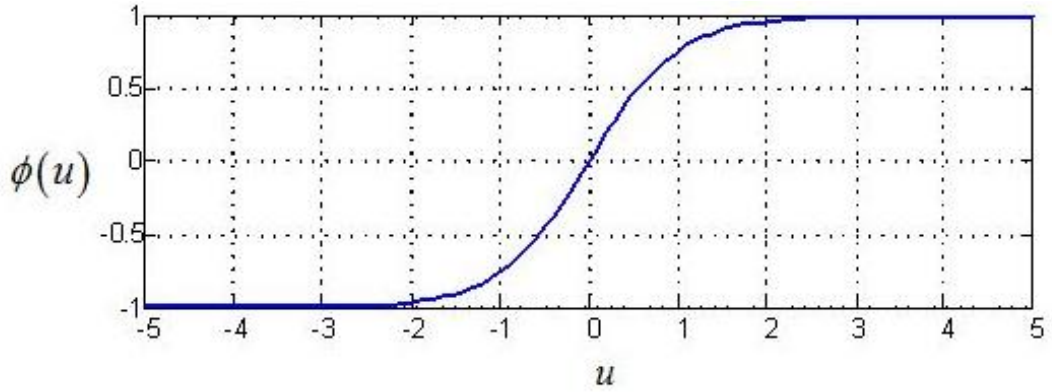
- مدل سیگموئید Sigmoid
- به ازای ورودی‌های زیادی پارامترها به روز نمی‌شوند

با افزایش تعداد لایه‌ها، سرعت آموزش وزن‌ها کاهش می‌یابد (چرا؟)

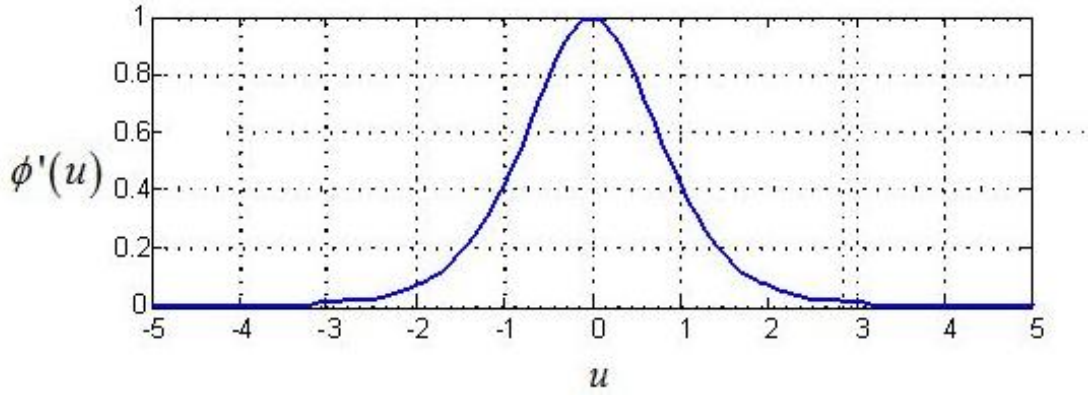
تولید محتوا: وحید محمدزاده ایوقی

مدل نرون

- مدل تانژانت هایپربولیک Tansig
- به ازای ورودی‌های زیادی پارامترها به روز نمی‌شوند



$$t(x) = 2\sigma(x) - 1$$



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

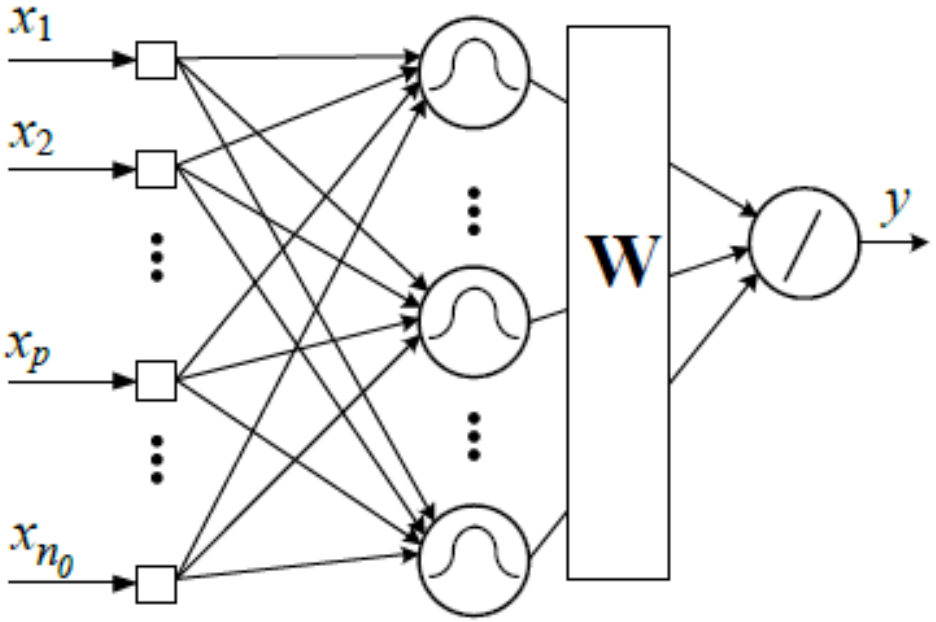
مدل نرون

- مدل تابع پایه‌ای شعاعی RBF

$$y = W\phi(x) + b$$

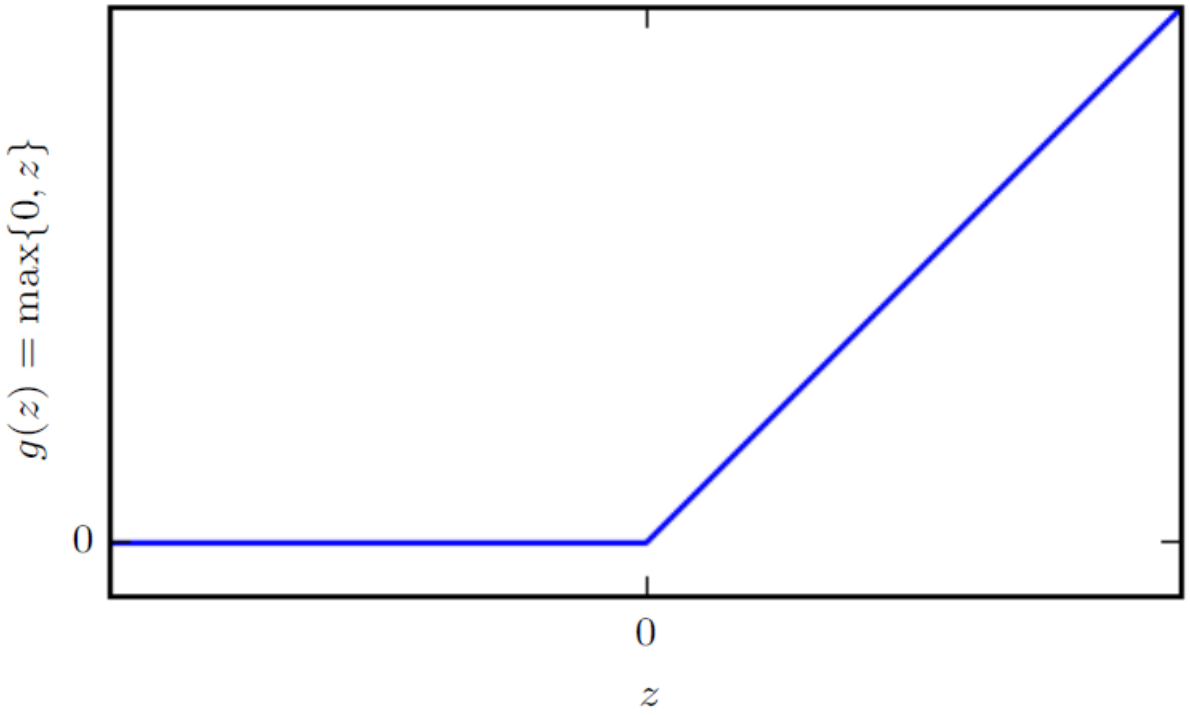
- پارامترهای هر تابع نرون قابل آموزش دیدن هستند
  - مشکل این نرون‌ها در تنظیم پارامترهای نرون است.

این توابع اشباع نمی‌شوند و رفتار محلی دارند
برای مسئله رگرسیون عملکرد مناسبی دارند



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مدل نرون



- مدل خطی یکطرفه ReLU
  - سرعت همگرایی را افزایش می‌دهد
  - به ازای ورودی‌های زیادی پارامترها به روز نمی‌شوند
  - مقدار اولیه بایاس اگر مقدار مثبتی داشته باشد به احتمال زیاد نرون فعال خواهد شد

$$g(z_i) = \max(0, z_i) + \alpha_i \min(0, z_i)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

آموزش شبکه‌های عصبی

- آموزش شبکه‌های عصبی – تعیین وزن‌ها
  - آموزش شبکه‌های عصبی، به عنوان یک ابزار یادگیری ماشین، تفاوت چندانی با سایر مدل‌های یادگیری ماشین ندارد.

- آموزش یک مدل یادگیر
  - معیار
  - الگوریتم بهینه‌سازی

$$\hat{y} = f^3\left(f^2\left(f^1(x)\right)\right)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

معیار – تابع هزینه

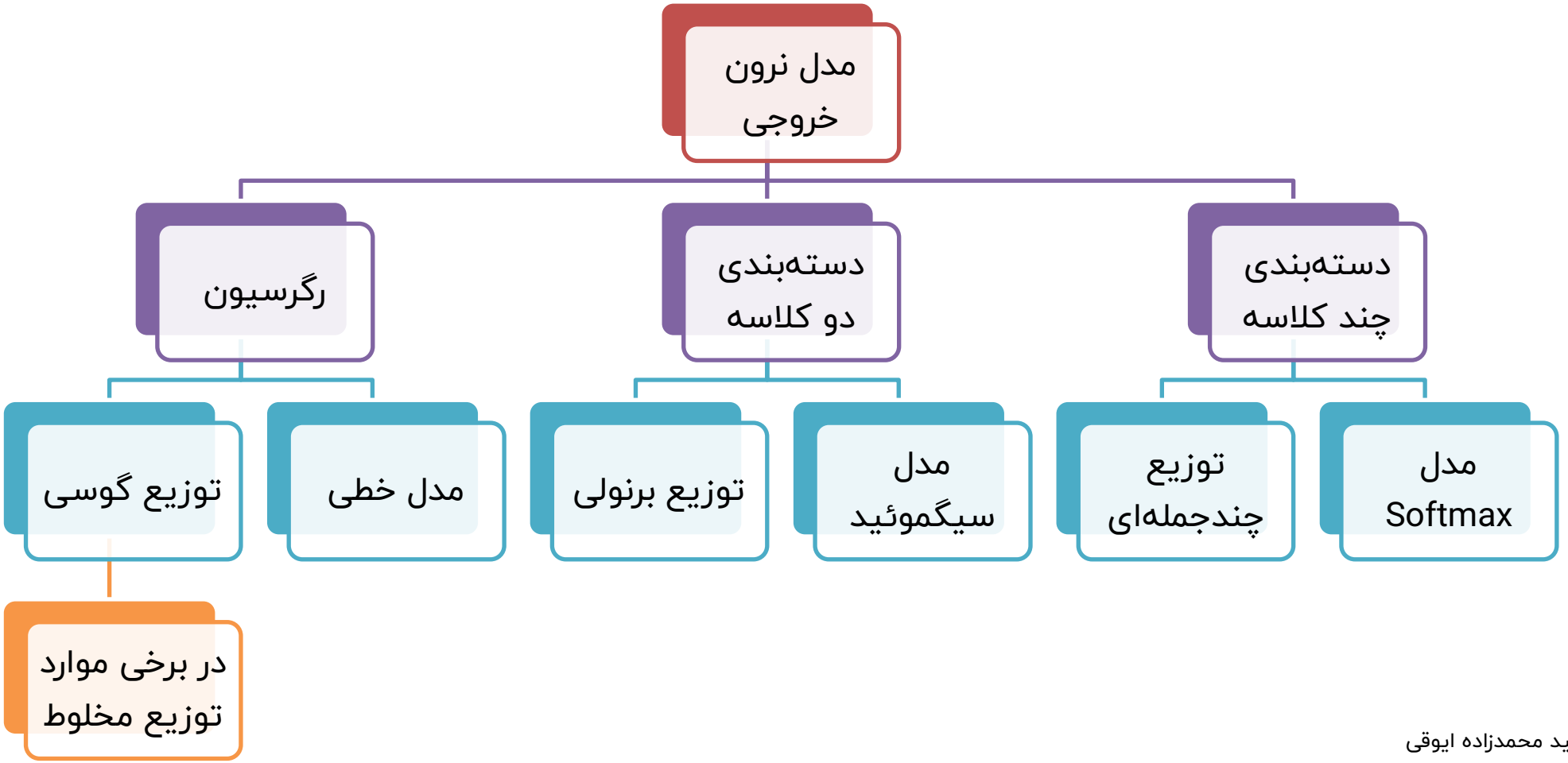
- تابع هزینه
  - برای هر مدل یادگیری ماشین
    - تخمین خود توزیع
    - تخمین پارامترهای آماری توزیع
- فرم تابع هزینه
  - به نوع تسک بستگی دارد.
  - نوع تسک را نوع نرون خروجی تعیین می‌کند.

تخمین بیشینه شباهت

$$J = -\log P(y|x;\theta)$$

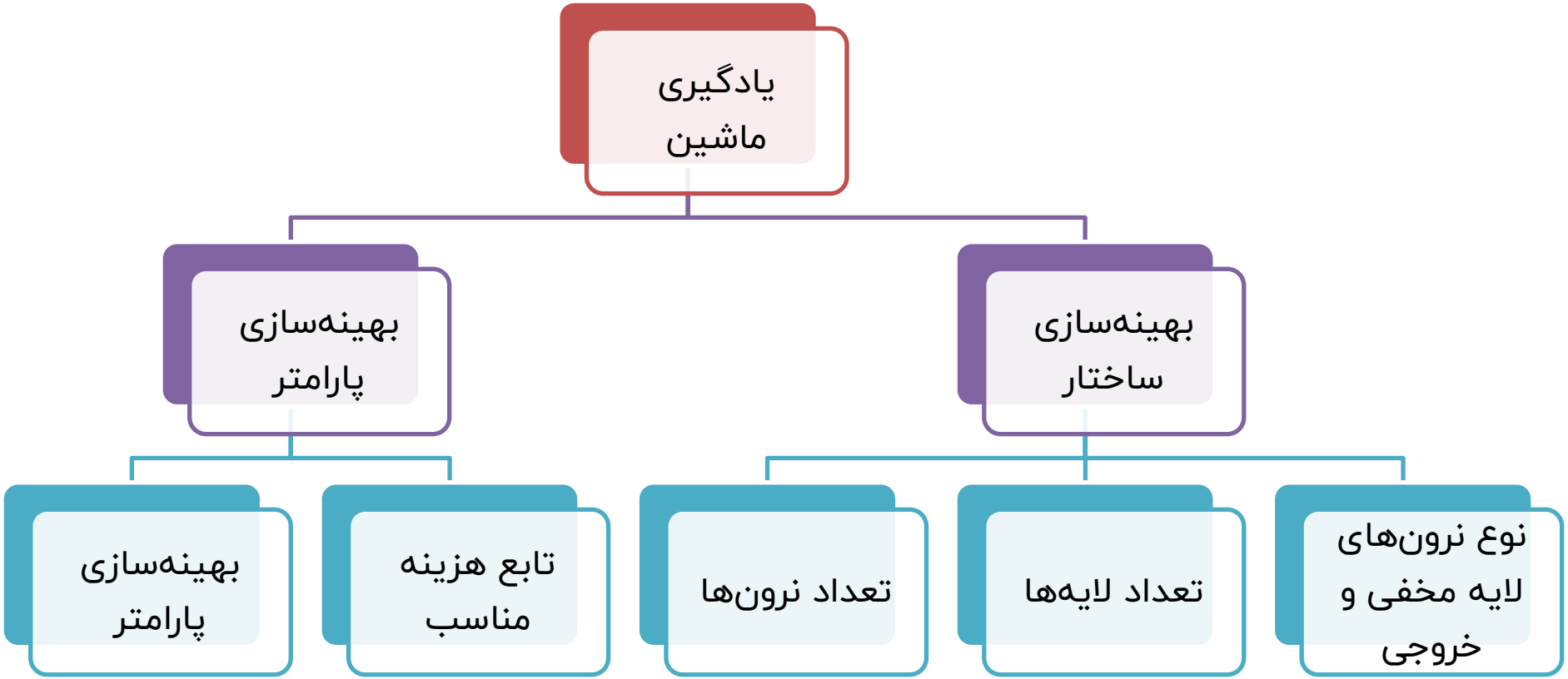
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مدل نرون خروجی



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

ساختار شبکه

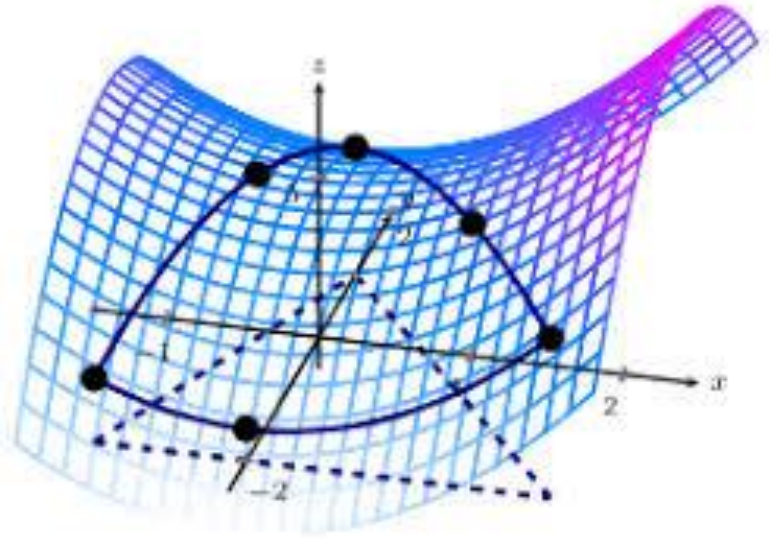


تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

بهینه‌سازی

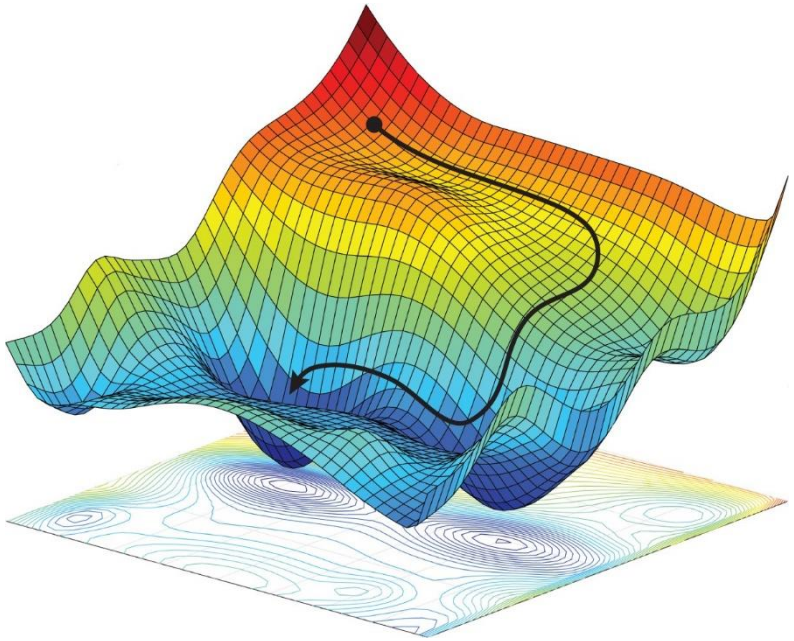
مسئله بهینه‌سازی مقید

فضای جستجوی پارامترها محدود به یک ناحیه خاص است



مسئله بهینه‌سازی نامقید

فضای جستجوی پارامترها محدود به یک ناحیه خاص نیست



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مسئله استاندارد بهینه‌سازی

- یک مسئله بهینه‌سازی استاندارد همواره به صورت زیر نوشته می‌شود.

$$x^* = \arg\min f(x)$$
$$h_j(x) = 0, j = 1, 2, \dots, k$$
$$g_i(x) < 0, i = 1, 2, \dots, m$$

$$x^* = \arg\max f(x) = -\arg\min f(x)$$

Linear programming

$$x^* = \arg\min C^T x$$
$$A_j x = 0, j = 1, 2, \dots, k$$
$$B_i x - b < 0, i = 1, 2, \dots, m$$

Quadratic programming

$$x^* = \arg\min x^T C x$$
$$A_j x = 0, j = 1, 2, \dots, k$$
$$B_i x - b < 0, i = 1, 2, \dots, m$$

شرط بهینه‌گی

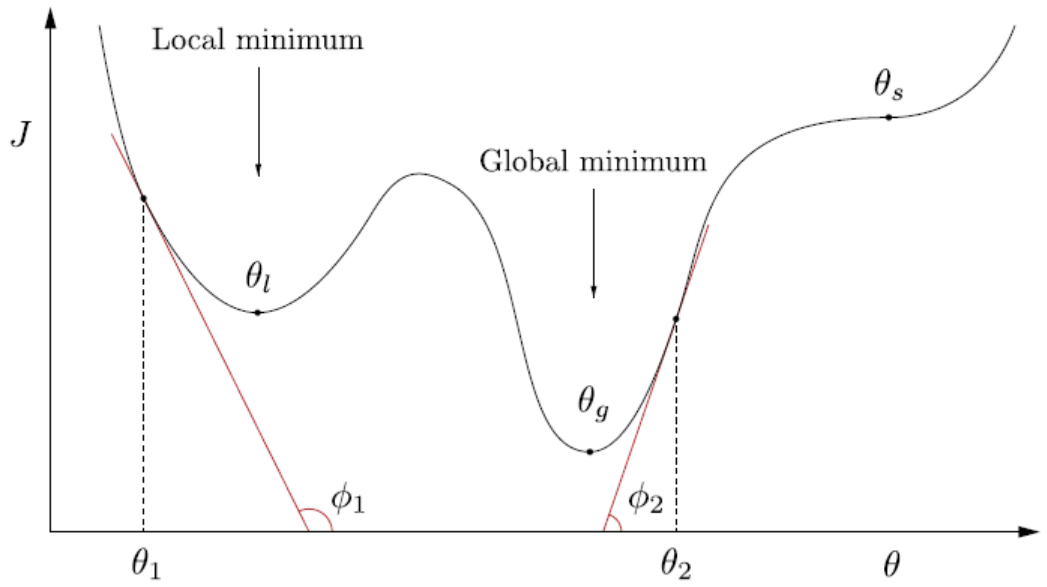
- تعریف بهینه‌گی

- مقدار یک پارامتر برای یک تابع هزینه بهینه است به چه معناست؟
- هر پارامتر بهینه باید یک سری از شرایط را ارضا کند.
- شرایط بهینه‌گی به عنوان نقطه خاتمه یک الگوریتم در نظر گرفته میشود.
- در یک الگوریتم گام به گام اگر به نقطه‌ای رسیدیم که شرایط داده شده را ارضا کند، نقطه مد نظ، یک نقطه بهینه خواهد بود.

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شرط بهینه‌گی

$$x^* = \arg\min f(x)$$



- بهینه‌سازی نامقید
  - اگر نقطه x^* یک نقطه مینیم محلی باشد باید گرادیان تابع در این نقطه برابر با صفر باشد (شرط لازم)
  - برای کلیه نقاط مشخص شده در منحنی روبه‌رو شرط لازم برقرار است.
    - ماتریس هسیان (مشتق مرتبه دوم) تابع بهینه‌سازی باید مثبت نیمه‌معین باشد تا از حداقل بودن نقطه مطمئن شویم (شرط کافی)

بهینه‌سازی نامقید

- مثال

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6 \rightarrow x_1^*, x_2^* = ?$$

شرط لازم: $\nabla f = 0 \rightarrow \begin{pmatrix} 3x_1^2 + 4x_1 \\ 3x_2^2 + 8x_2 \end{pmatrix} = 0 \rightarrow \begin{cases} x_1 = 0, & x_1 = -\frac{4}{3} \\ x_2 = 0, & x_2 = -\frac{8}{3} \end{cases}$

شرط کافی: $H = \begin{pmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{pmatrix}$

- اگر ماتریس هسیان در یک نقطه:
- مثبت معین باشد، نقطه مورد نظر مینیمم محلی است
- منفی معین باشد، نقطه مورد نظر ماکزیمم محلی است
- نه منفی معین و نه مثبت معین باشد، نقطه مورد نظر زینی است.

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

بهینه‌سازی نامقید

- مثال

$$H = \begin{pmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{pmatrix}$$

$x_1 = 0, x_2 = 0 \rightarrow H = \begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix}$ مثبت معین است

$x_1 = 0, x_2 = -\frac{8}{3} \rightarrow H = \begin{pmatrix} 4 & 0 \\ 0 & -8 \end{pmatrix}$ نه مثبت معین است نه منفی معین

$x_1 = -\frac{4}{3}, x_2 = 0 \rightarrow H = \begin{pmatrix} -4 & 0 \\ 0 & 8 \end{pmatrix}$ نه مثبت معین است نه منفی معین

$x_1 = -\frac{4}{3}, x_2 = -\frac{8}{3} \rightarrow H = \begin{pmatrix} -4 & 0 \\ 0 & -8 \end{pmatrix}$ منفی معین است

بهینه‌سازی نامقید

- مثال

$$\omega^* = \arg\min -\sum_{n=1}^{N}\{\sigma(\omega^T\phi(x_n))\ln t_n + (1-\sigma(\omega^T\phi(x_n))\ln(1-t_n)\}, \sigma\left(\omega^T\phi(x_n)\right) = \frac{1}{1+\exp\left(-\omega^T\phi(x_n)\right)}$$

شرط لازم: $\nabla f = 0 \rightarrow -\sum_{n=1}^{N}(\sigma(\omega^T\phi(x_n) - t_n)\phi(x_n) = 0$

حل بسته ریاضی برای این مسئله وجود ندارد! راه حل چیست؟

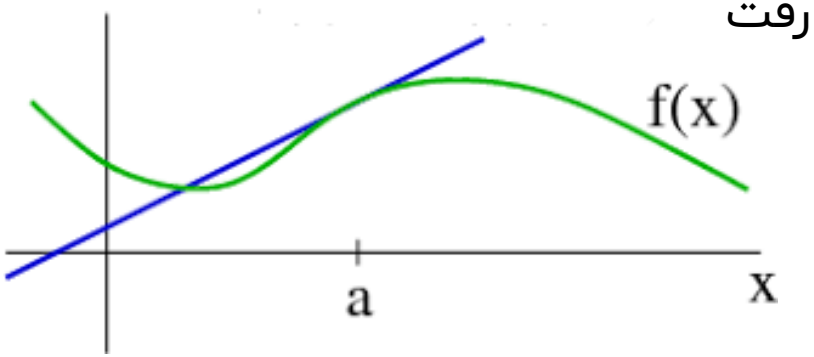
- روش‌های گام به گام
  - از یک نقطه اولیه شروع می‌کنیم و به صورت گام به گام پاسخ بدست آمده را در هر مرحله به روز می‌کنیم تا شرط بهینه‌گی مهیا شود.

بهینه‌سازی نامقید

- الگوریتم گرادیان نزولی
- بر اساس بسط تیلور مرتبه اول

$$f(x+\Delta x) \cong f(x) + \Delta x^T \nabla f \quad \Rightarrow \quad \Delta x = ? \rightarrow f(x+\Delta x) \leq f(x)$$

در خلاف جهت شیب منحنی اگر حرکت کنیم به سمت نقطه مینیمم خواهیم رفت



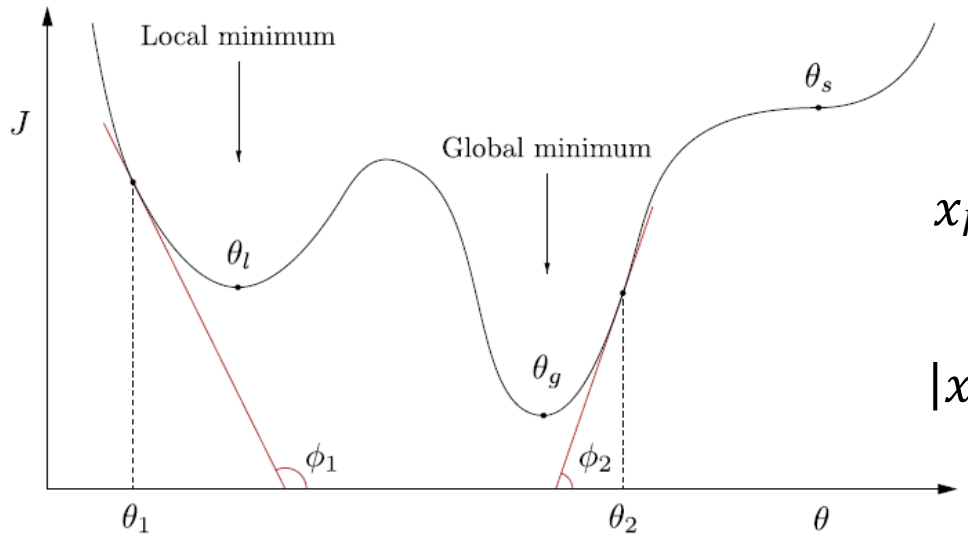
$$\Delta x = -\epsilon \nabla f$$

$$\boxed{x_{k+1} = x_k - \epsilon \nabla f}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

بهینه‌سازی نامقید

نرخ یادگیری – موثر بر همگرایی و سرعت همگرایی
- ثابت
- محاسبه بهینه نرخ یادگیری از طریق روش جستجوی خط



- الگوریتم گرادیان نزولی
  - انتخاب تصادفی مقدار اولیه برای پارامتر بهینه‌سازی
  - محاسبه مقدار تابع
  - محاسبه گرادیان
  - به روزرسانی مقدار پارامتر بهینه‌سازی $x_{k+1} = x_k - \epsilon \nabla f$
  - محاسبه مقدار تابع
  - بررسی شرط توقف $|x_{k+1} - x_k| < \epsilon, |f(x_{k+1}) - f(x_k)| < \epsilon$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

بهینه‌سازی نامقید

- محاسبه نرخ یادگیری بر اساس روش جستجوی خط

$$\epsilon^* = \arg\min(f\left(x_k - \epsilon\nabla f(x_k)\right)$$

$$f(x_1, x_2, x_3) = (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4 \rightarrow \nabla f = (4(x_1 - 4)^3 \quad 2(x_2 - 3) \quad 16(x_3 + 5)^3)^T$$

نقطه شروع اولیه: $x^0 = (4 \quad 2 \quad -1)^T \rightarrow \nabla f(x^0) = (0 \quad -2 \quad 1024)^T$

$$\epsilon^{0^*} - \arg\min f(x^0 - \epsilon\nabla f(x^0)) = f\left((4 \quad 2 - 2\epsilon \quad -1 - 1024\epsilon)\right) \rightarrow \epsilon^{0*} = 0.004$$

$$x^1 = x^0 - \epsilon^{0*}\nabla f = (0 \quad -1.98 \quad -0.003875)^T$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

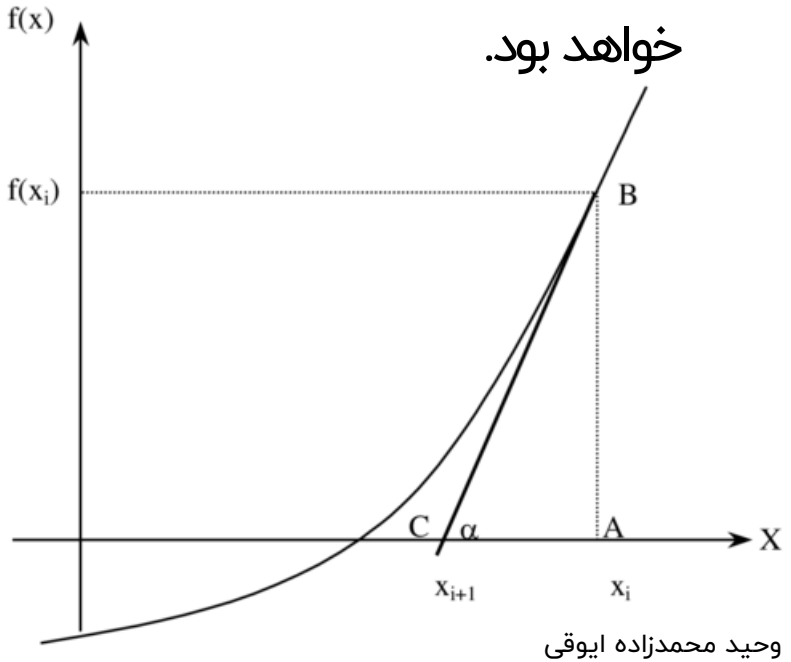
بهینه‌سازی نامقید

- روش حل نیوتن
- بسط تیلور بیان می‌کند هر قدر مشتقات مرتبه بالای یک تابع در ارزیابی تابع مشارکت داشته باشد، مقدار تابع دقیق‌تر خواهد بود.

$$f\left(x^k + \Delta x\right) = f\left(x^k\right) + \Delta x f'\left(x^k\right) = 0 \rightarrow \Delta x = -\frac{f\left(x^k\right)}{f'\left(x^k\right)}$$

$$x^{k+1} = x_k - \frac{f\left(x^k\right)}{f'\left(x^k\right)}$$ حالت اسکالر

$$x^{k+1} = x^k - \nabla f\left(x^k\right)^{-1} f\left(x^k\right)$$ حالت برداری



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

بهینه‌سازی نامقید

- روش حل نیوتن

$$\nabla f\left(x^k + \Delta x\right) = 0 \rightarrow \nabla f\left(x^k\right) + H(x^k)\Delta x = 0 \rightarrow \Delta x = -H\left(x^k\right)^{-1}\nabla f(x^k)$$

$$x^{k+1} = x_k - H\left(x^k\right)^{-1}\nabla f(x^k)$$

- ماتریس هسیان باید در هر نقطه مثبت معین باشد. (چرا؟)
- این امکان وجود دارد بدلیل وجود خطای محاسبات در محاسبه ماتریس هسیان، فرض مثبت معین بودن نقض شود.
- برای حل این مشکل چه راه حلی وجود دارد؟

بهینه‌سازی نامقید

- روش Levenberg - Marquadt

$$x^{k+1} = x_k - (H(x^k) + \lambda I)^{-1} \nabla f(x^k)$$

  - ترم افزوده شده نقش کنترلی دارد.
  - مقدار پارامتر را به نحوی تعیین میکنیم که ماتریس حاصل یک ماتریس مثبت معین باشد. آیا امکان تطبیقی کردن این ترم وجود دارد؟
  - این روش بسته به مقدار پارامتر کنترلی، عملکرد بین روش نیوتن و روش گرادیان نزولی دارد.
  - تمام روش‌های مبتنی بر مشتق مرتبه دوم بر روی روش نیوتن توسعه یافته‌اند.
  - در این روش فرض بر این است که تابع مورد نظر در دسترس است حال آنکه در کاربردهای عملی این فرض درست نیست
    - روش گوس – نیوتن
    - روش BFGS

تولید محتوا: وحید محمدزاده ایوقی

بهینه‌سازی نامقید

- محدودیت‌های روش مبتنی بر مشتق دوم
  - نیاز به محاسبه ماتریس هسیان و معکوس آن

- محدودیت روش‌های مبتنی بر مشتق اول
  - یکسان بودن نرخ یادگیری برای تمام داده‌های آموزش
  - کند بودن الگوریتم
  - احتمال گیر افتادن در نقطه زینی

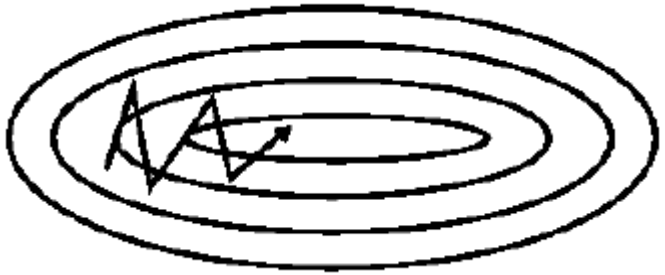
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

روش مومنتم

- کند بودن الگوریتم
  - احتمال گیر افتادن در نقطه زینی

تصحیح روش گرادیان نزولی

- ایجاد لختی در تغییرات
- تمایل به تغییرات شدید در هر جهت کاهش خواهد یافت



$$v_t = \mu v_{t-1} - \eta \nabla f(\theta_t)$$

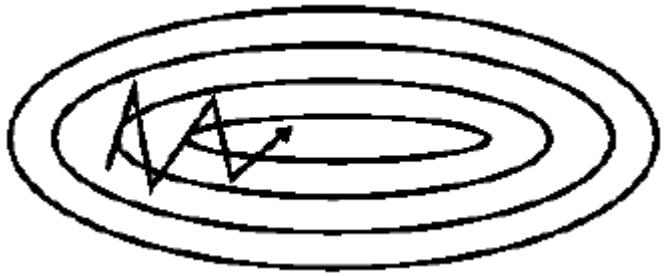
$$\theta_t = \theta_{t-1} + v_t$$

تولید محتوا: وحید محمدزاده ایوقی

روش NAG

- بر مبنای روش مومنتم

اگر بدانیم با به روزرسانی پارامترها گام بعدی کجاست، سرعت بهتر خواهد شد.



$$v_t = \mu v_{t-1} - \eta \nabla f(\theta_t)$$

$$\theta_t = \theta_{t-1} + v_t$$

$$v_t = \mu v_{t-1} - \eta \nabla f(\theta_{t-1} + \mu v_{t-1})$$

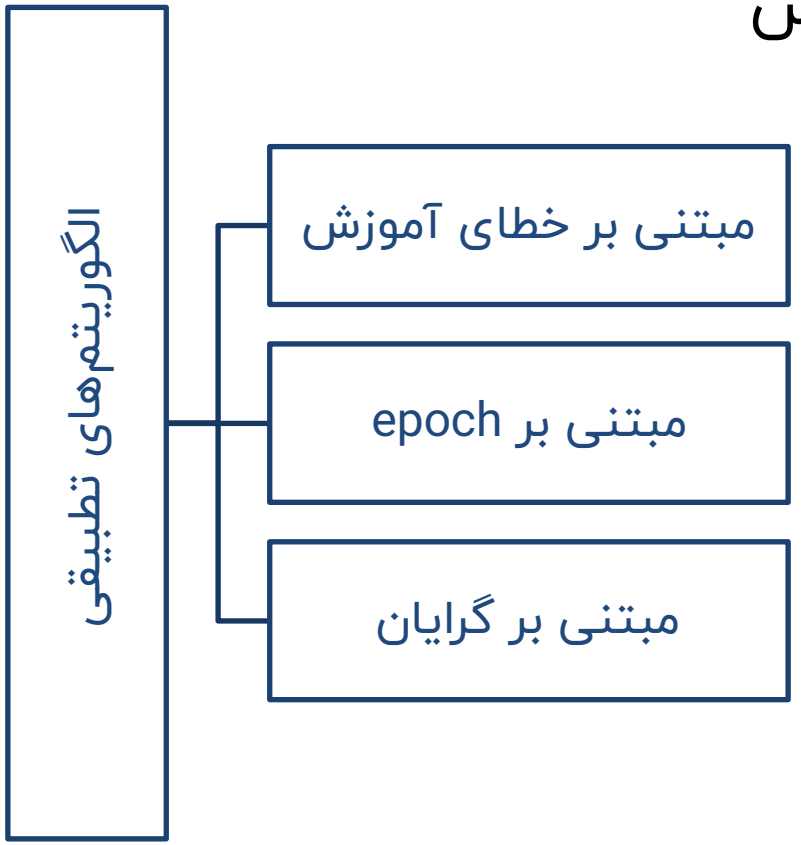
$$\theta_t = \theta_{t-1} + v_t$$

$$\phi_{t-1} = \theta_{t-1} + \mu v_{t-1} \rightarrow \begin{cases} v_t = \mu v_{t-1} - \epsilon \nabla f(\phi_{t-1}) \\ \phi_t = \phi_{t-1} - \mu v_{t-1} + (1+\mu) v_t \end{cases}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

الگوریتم بهینه‌سازی تطبیقی

- یکسان بودن نرخ یادگیری برای تمام داده‌های آموزش



روش Adagrad

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{G(t)} + \epsilon} \nabla f(\theta_{t-1})$$

$$G(t) = \sum_{i=1}^{t} \nabla f(\theta_i)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

الگوریتم بهینه‌سازی تطبیقی

- بهینه‌سازی به کمک روش RMS Prop

$$s_t = \gamma s_{t-1} + (1-\gamma)\nabla f(\theta_t)^2, \qquad \theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{s_t} + \epsilon}\nabla f(\theta_{t-1})$$

الگوریتم آداگراد رفته رفته نرخ یادگیری را کاهش می‌دهد

ارائه شده توسط Hinton

$$\gamma = 0.9, \qquad \eta \leq 0.001$$

الگوریتم بهینه‌سازی تطبیقی

- بهینه‌سازی به کمک روش Adadelta

$$g_t = \gamma \Delta\theta_{t-1}^2 + (1-\gamma)\Delta\theta_t^2, \qquad \Delta\theta_t = \frac{\sqrt{g_t} + \epsilon_1}{\sqrt{s_t} + \epsilon_2} \nabla f(\theta_{t-1})$$

$$s_t = \gamma s_{t-1} + (1-\gamma)\nabla f(\theta_t)^2$$

$$\theta_t = \theta_{t-1} - \Delta\theta_t$$

الگوریتم بهینه‌سازی تطبیقی

- بهینه‌سازی به کمک روش Adam

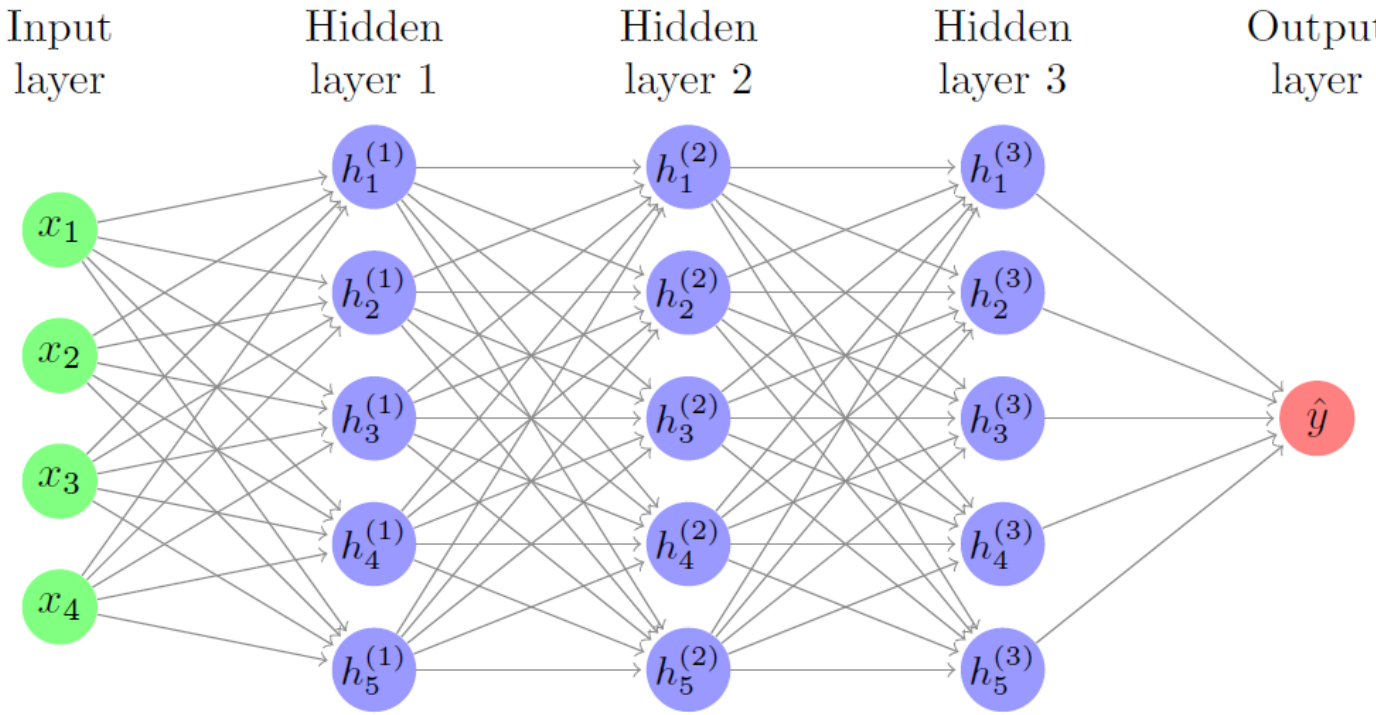
$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}(t)} + \epsilon}\hat{m}(t)$$

$$\hat{m}(t) = \frac{m(t)}{1+\beta_1}\,\hat{v}(t) = \frac{v(t)}{1+\beta_2} \rightarrow \begin{cases} m(t) = \beta_1 m(t-1) + (1-\beta_1)\nabla J \\ v(t) = \beta_2 v(t-1) + (1-\beta_2)\nabla J^2 \end{cases}$$

- از یک دیدگاه می‌توان گفت این الگوریتم بین بایاس و واریانس تخمین بردار گرادیان یک مصالحه ایجاد می‌کند

آموزش شبکه عصبی

- شبکه عصبی
  - متشکل از لایه ورودی، لایه‌های مخفی و لایه خروجی

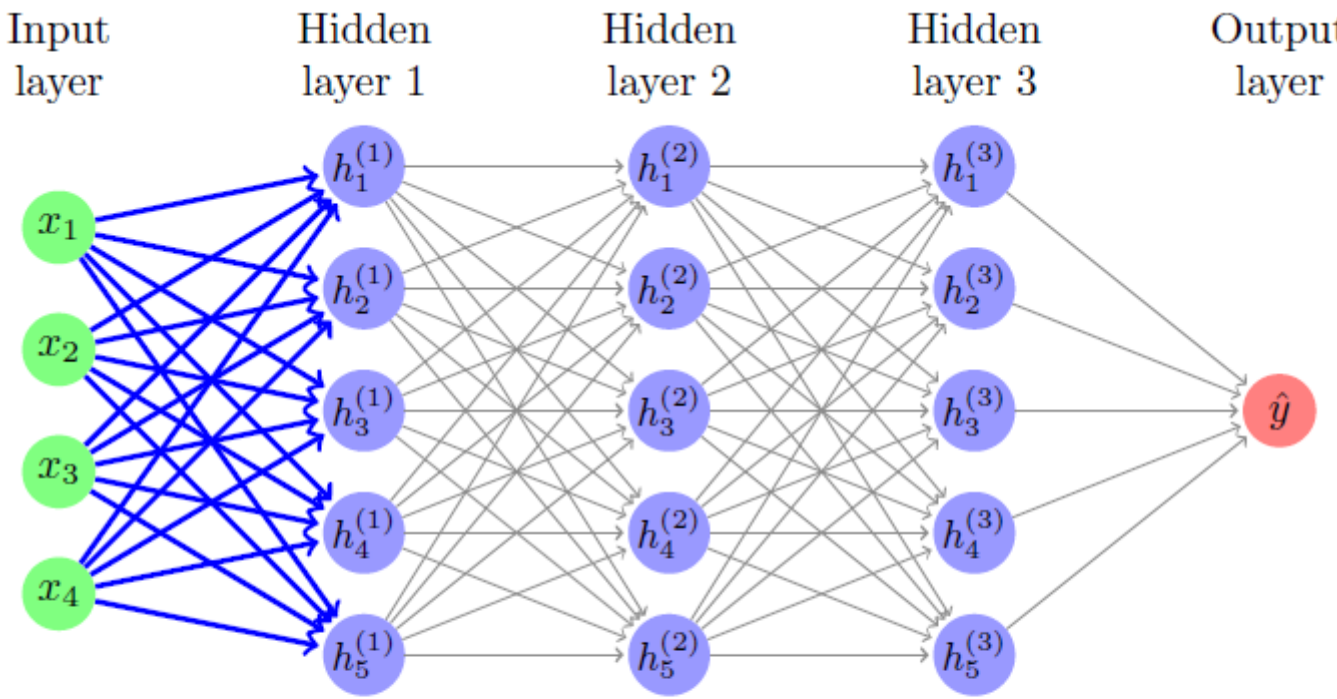


آموزش شبکه
- مسیر Forward
- مسیر Backward

$$\hat{y} = f\left(h_3(h_2(h_1(\underbrace{\mathbf{x}}_{\text{Input layer}}; \mathbf{W}_1); \mathbf{W}_2); \mathbf{W}_3); \mathbf{W}_4\right)$$

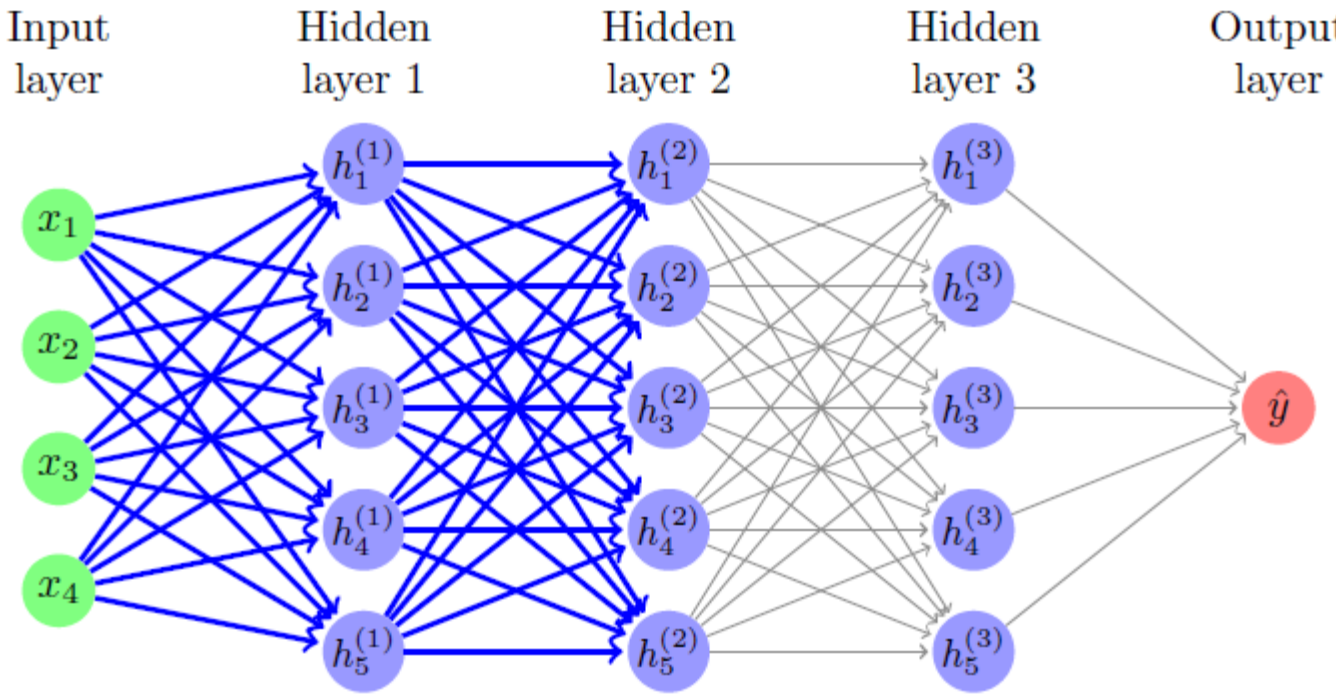
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

مسیر Forward



$$\hat{y} = f\left(h_3(h_2(\underbrace{h_1(\mathbf{x};\mathbf{W}_1)}_{\text{Hidden layer 1}};\mathbf{W}_2);\mathbf{W}_3);\mathbf{W}_4\right)$$

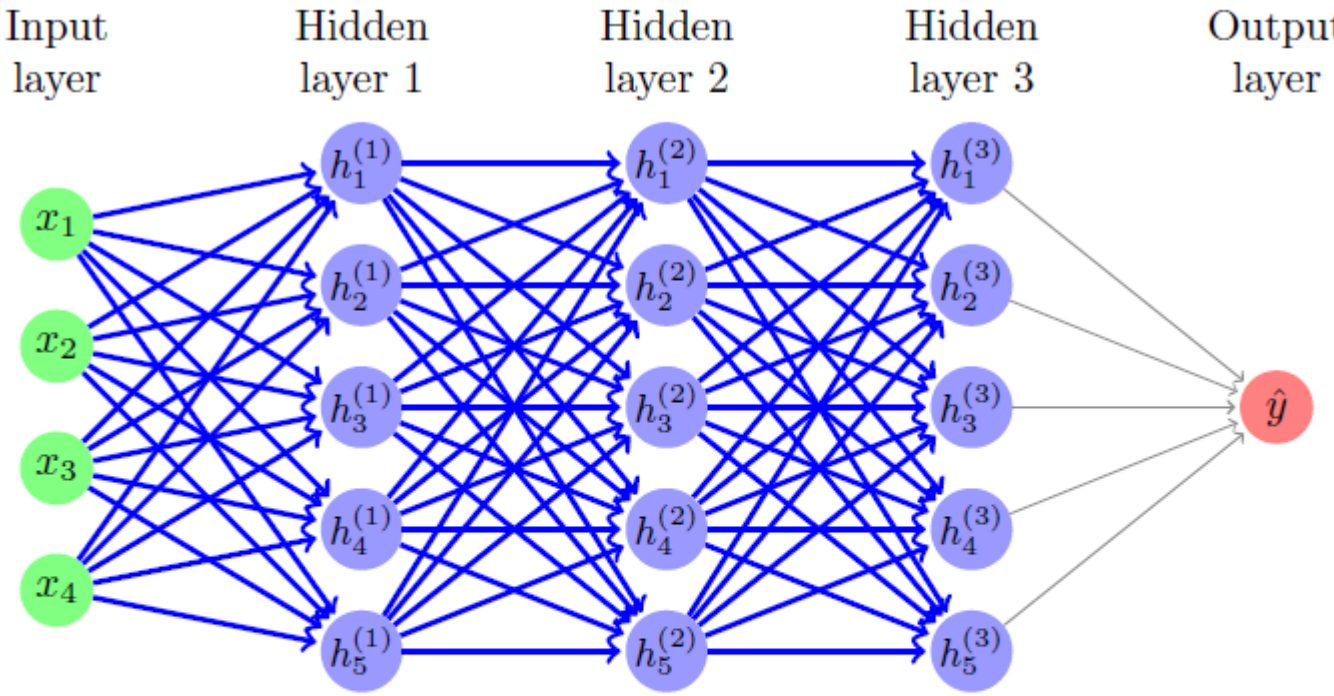
مسیر Forward



$$\hat{y} = f\left(h_3(\underbrace{h_2(h_1(\mathbf{x};\mathbf{W}_1);\mathbf{W}_2)}_{\text{Hidden layer 2}};\mathbf{W}_3);\mathbf{W}_4\right)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

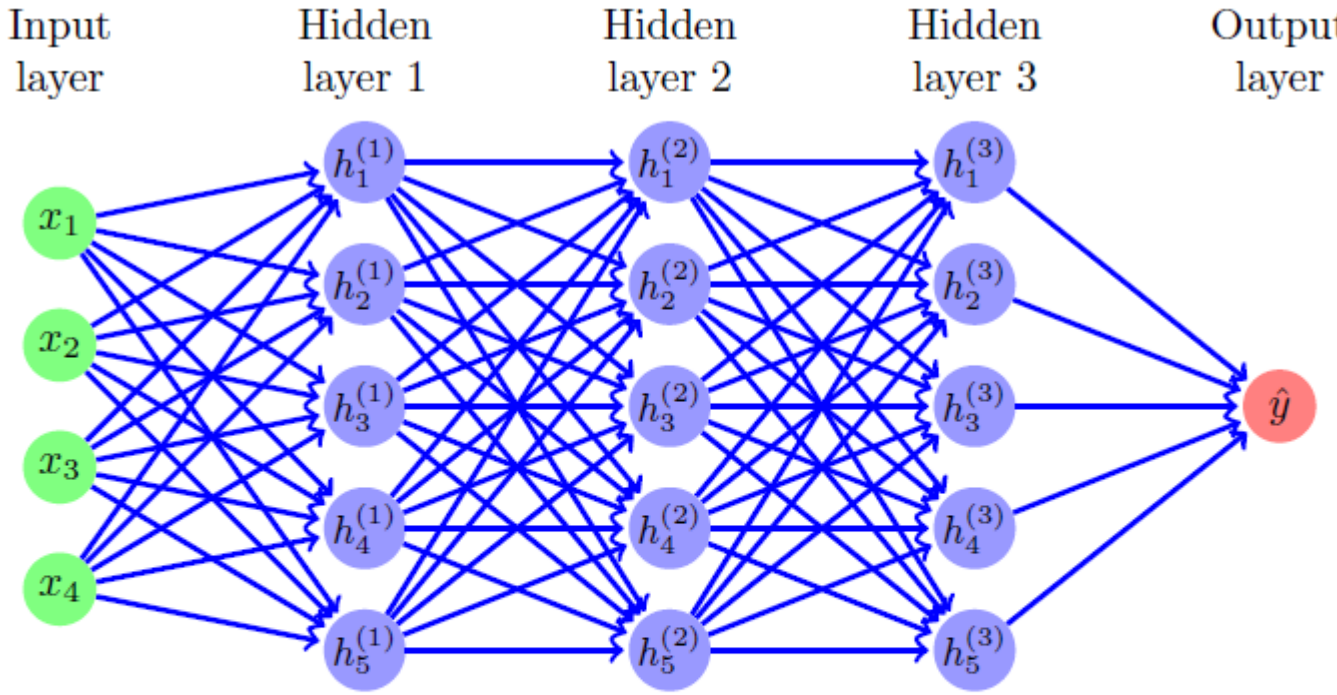
مسیر Forward



$$\hat{y} = f\left(\underbrace{h_3(h_2(h_1(\mathbf{x};\mathbf{W}_1);\mathbf{W}_2);\mathbf{W}_3)}_{\text{Hidden layer 3}};\mathbf{W}_4\right)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

مسیر Forward



$$\hat{y} = f\left(\underbrace{h_3(h_2(h_1(\mathbf{x};\mathbf{W}_1);\mathbf{W}_2);\mathbf{W}_3)}_{\text{Input to output layer}};\mathbf{W}_4\right)$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

پس‌انتشار خطا

- آموزش شبکه عصبی

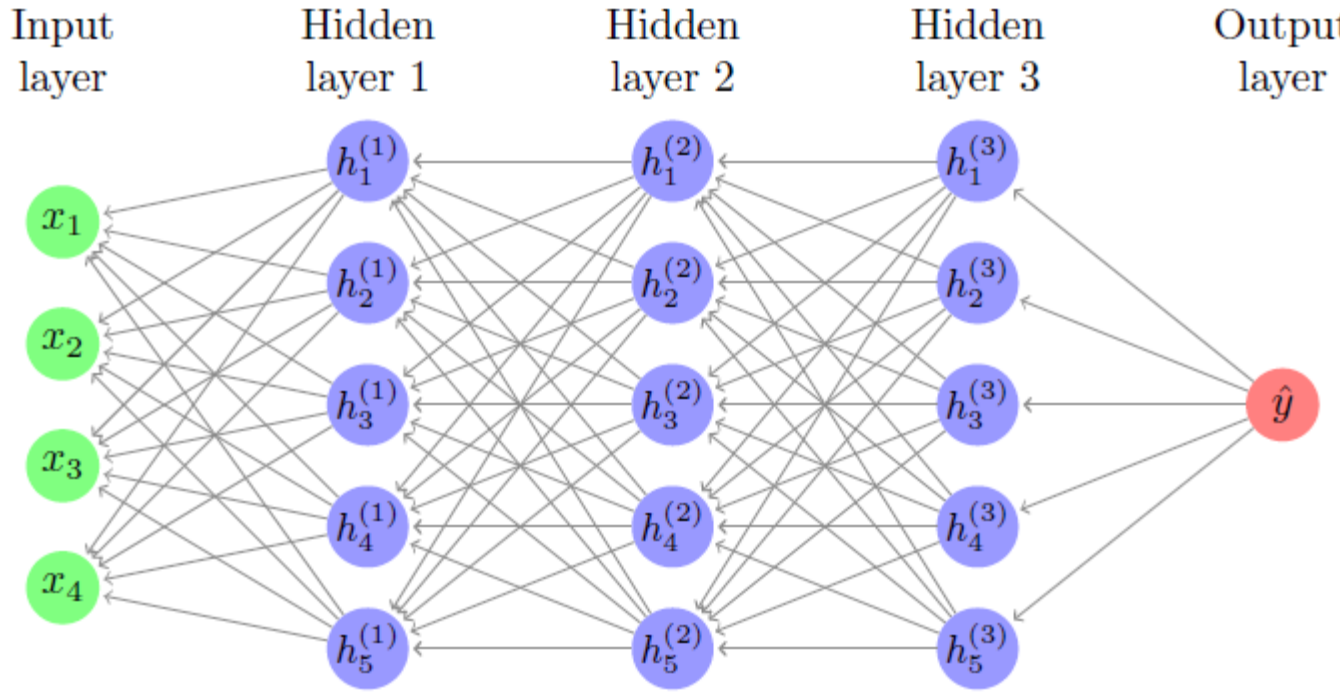
$$\hat{y} = f\left(h_3(h_2(h_1(\mathbf{x}; \underbrace{\mathbf{W}_1); \mathbf{W}_2); \mathbf{W}_3); \mathbf{W}_4}_{\text{Correspond to } \theta}\right)$$

$$L(y, \hat{y}) = -\log P(y|x, \theta) = \frac{1}{2}(y - \hat{y})^T(y - \hat{y}) = \frac{1}{2}E^T E$$

- مشتق زنجیره‌ای

$$y = f_1\left(f_2\left(f_3(\ldots(x))\right)\right) \rightarrow \frac{\partial y}{\partial x} = \frac{\partial y}{\partial f_1} \times \frac{\partial f_1}{\partial f_2} \ldots \frac{\partial f_n}{\partial x}$$

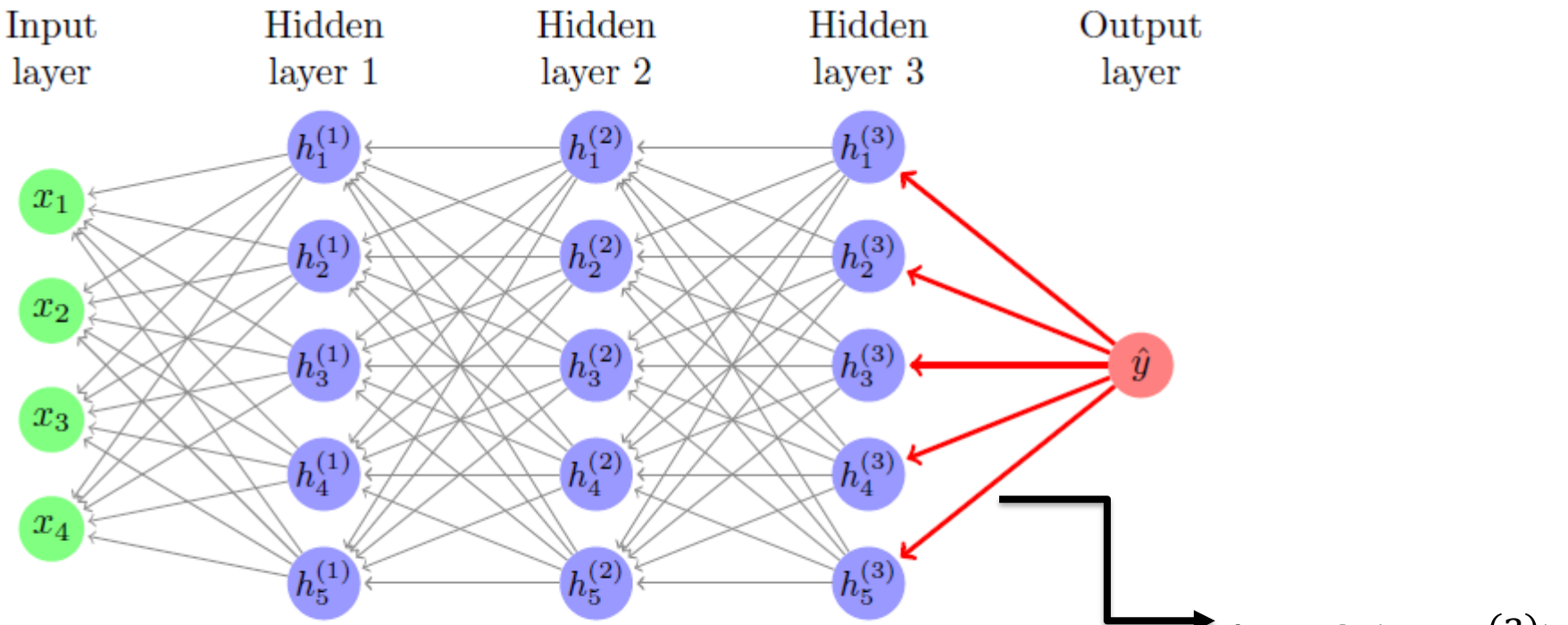
مسیر Backward



Loss function $L(y, \hat{y})$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

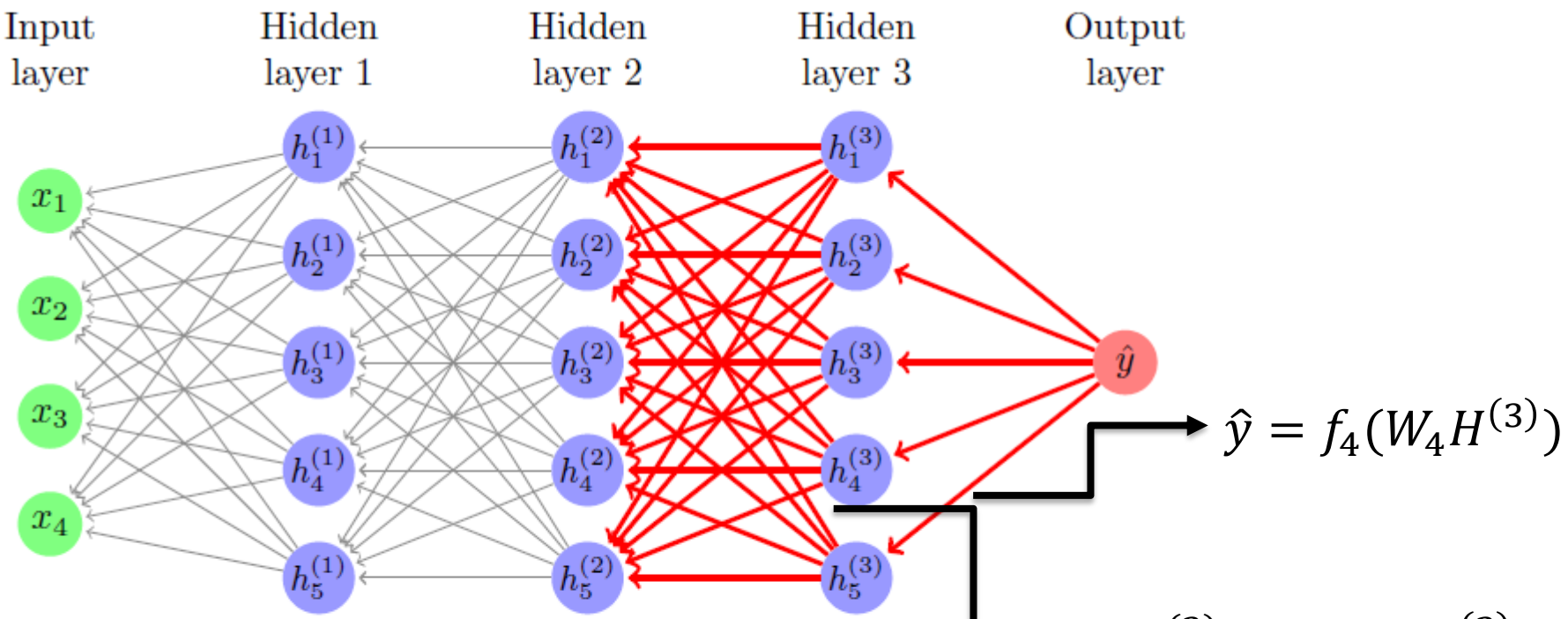
مسیر Backward



$$\mathbf{W}_4 \leftarrow \mathbf{W}_4 - \eta \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \mathbf{W}_4}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

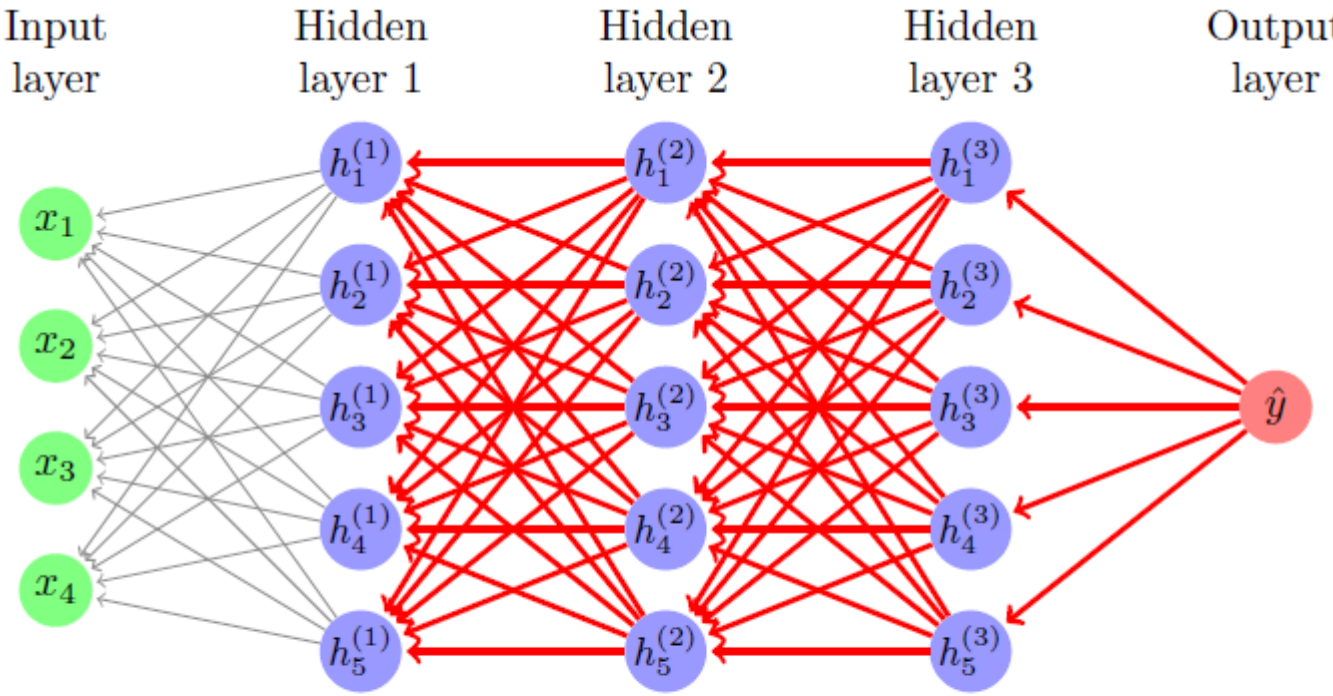
مسیر Backward



$$\mathbf{W}_3 \leftarrow \mathbf{W}_3 - \eta \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_3} \frac{\partial h_3}{\partial \mathbf{W}_3}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

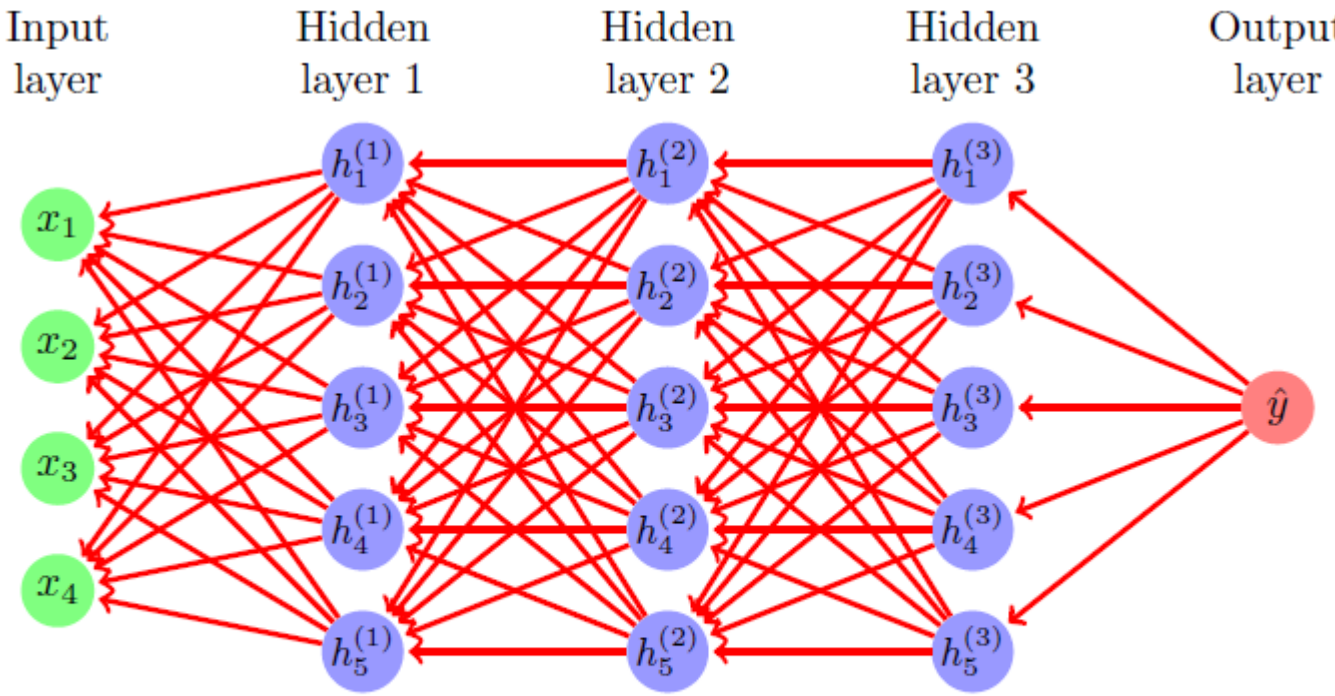
مسیر Backward



$$\mathbf{W}_2 \leftarrow \mathbf{W}_2 - \eta \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial \mathbf{W}_2}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

مسیر Backward

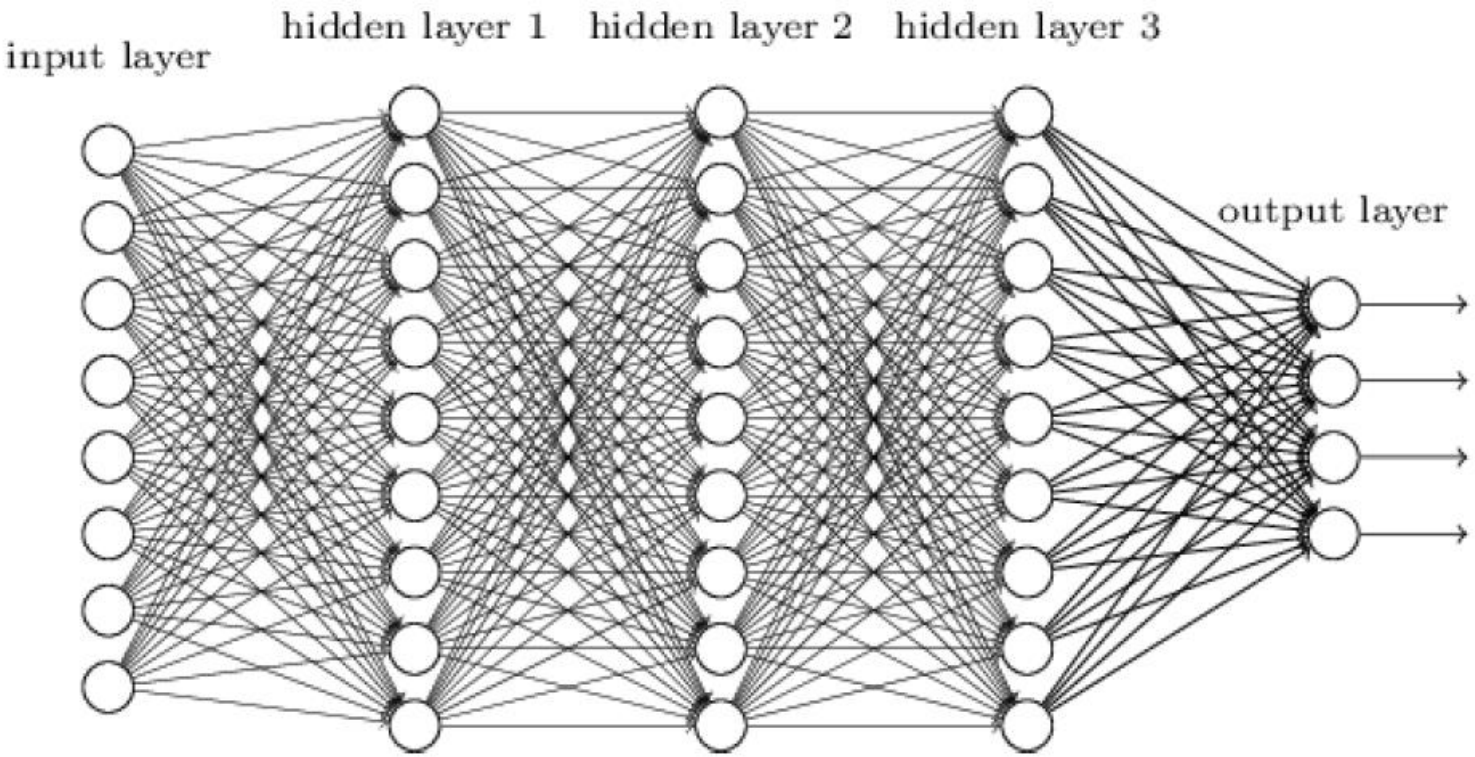


$$\mathbf{W}_1 \leftarrow \mathbf{W}_1 - \eta \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_3} \frac{\partial h_3}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial \mathbf{W}_1}$$

تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

شبکه‌های عصبی عمیق

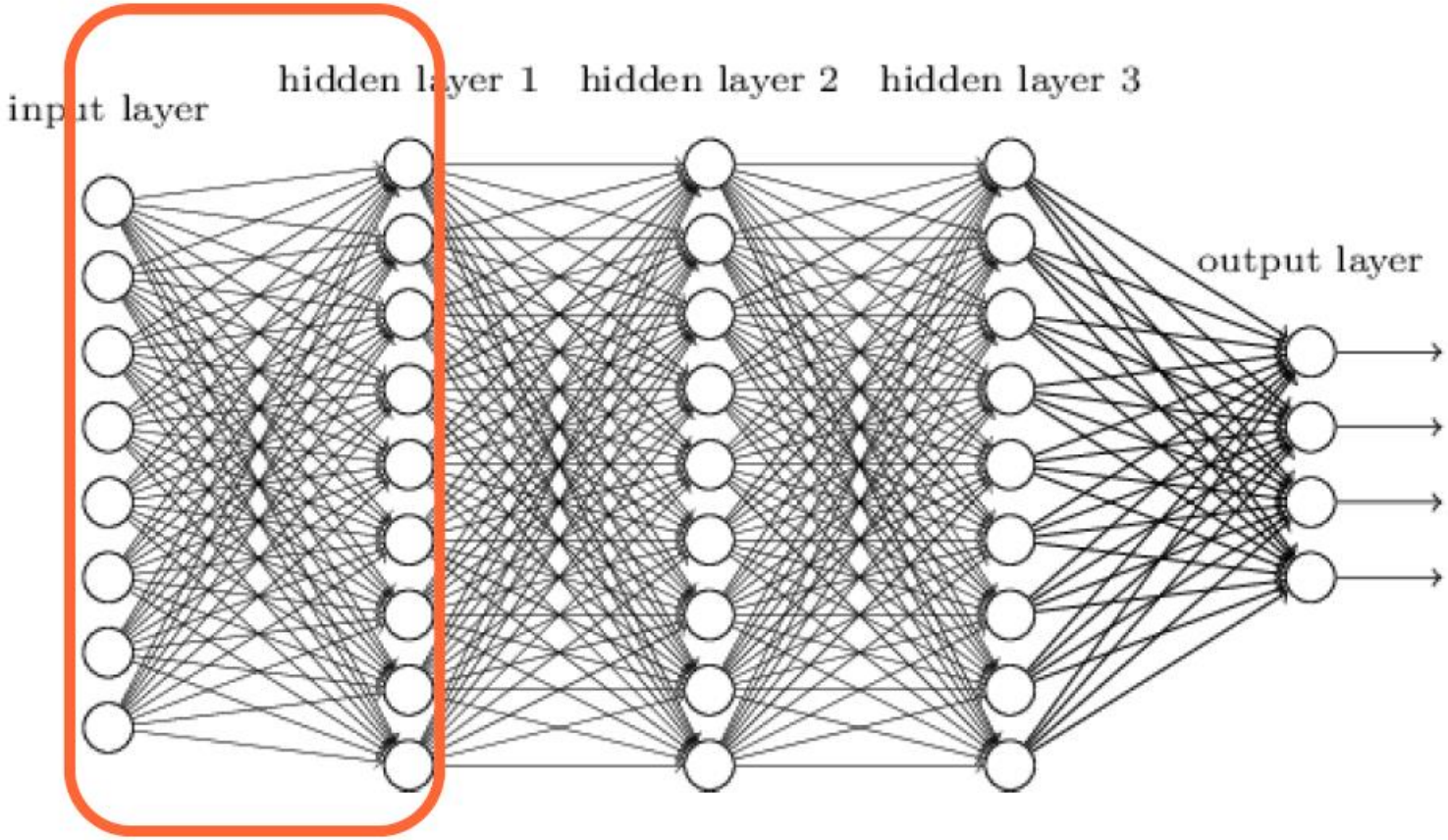
- شبکه‌های عصبی عمیق
  - داده‌های حجیم با بعد بالا – افزایش تعداد لایه‌ها



محو شدگی گرادیان

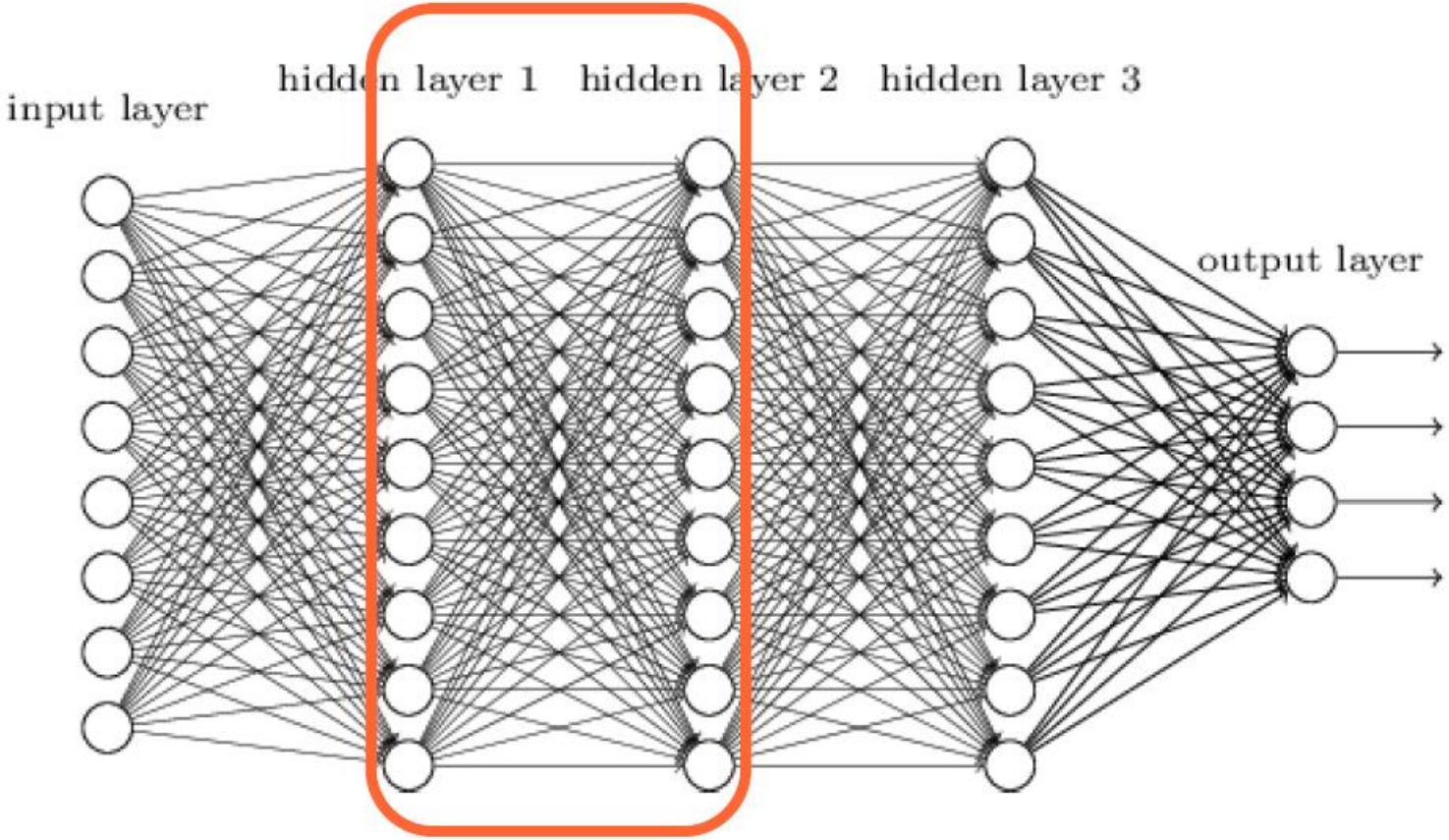
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شبکه‌های عصبی باور عمیق



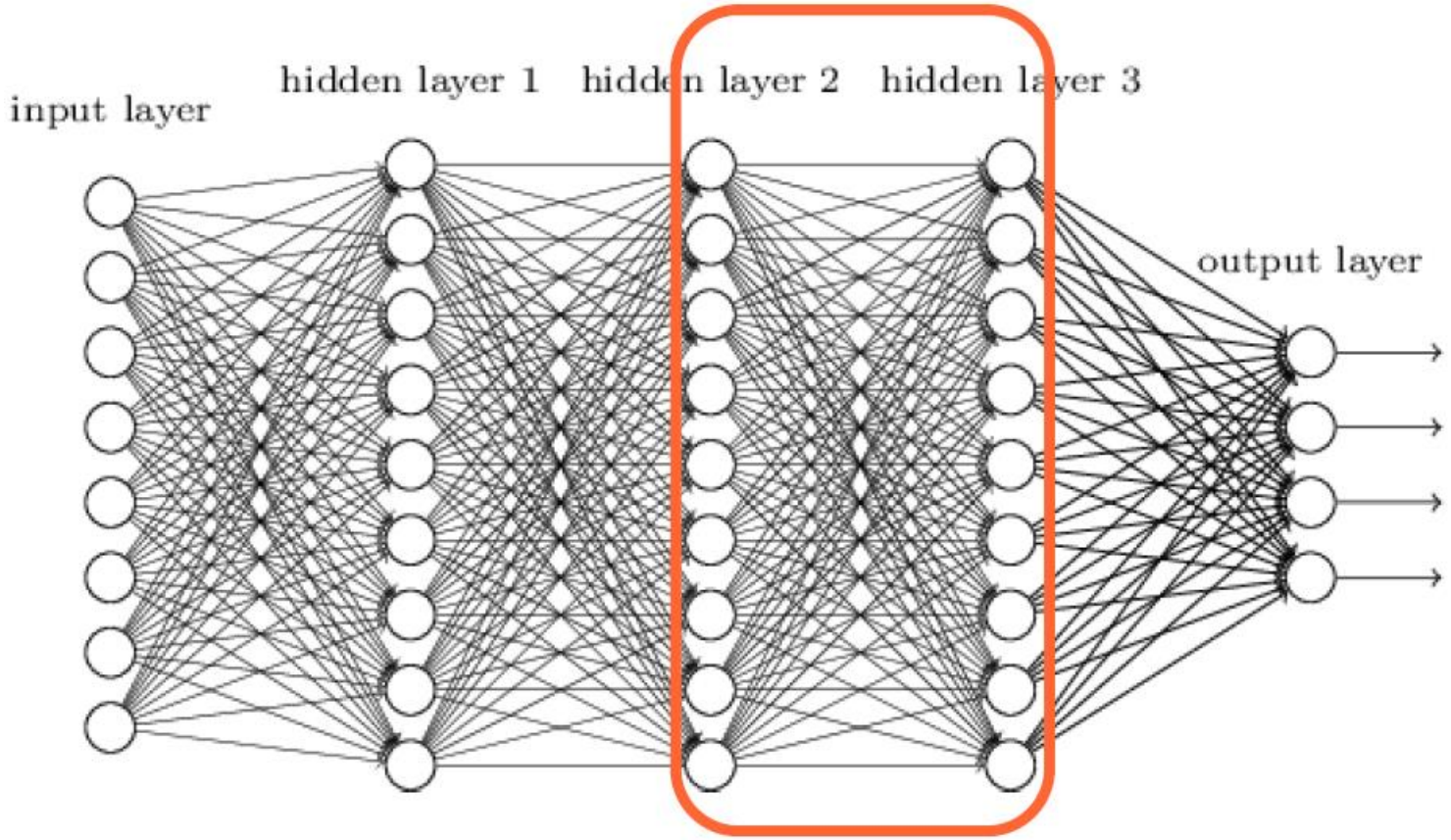
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
گروه دایکه | dayche.com

شبکه‌های عصبی باور عمیق



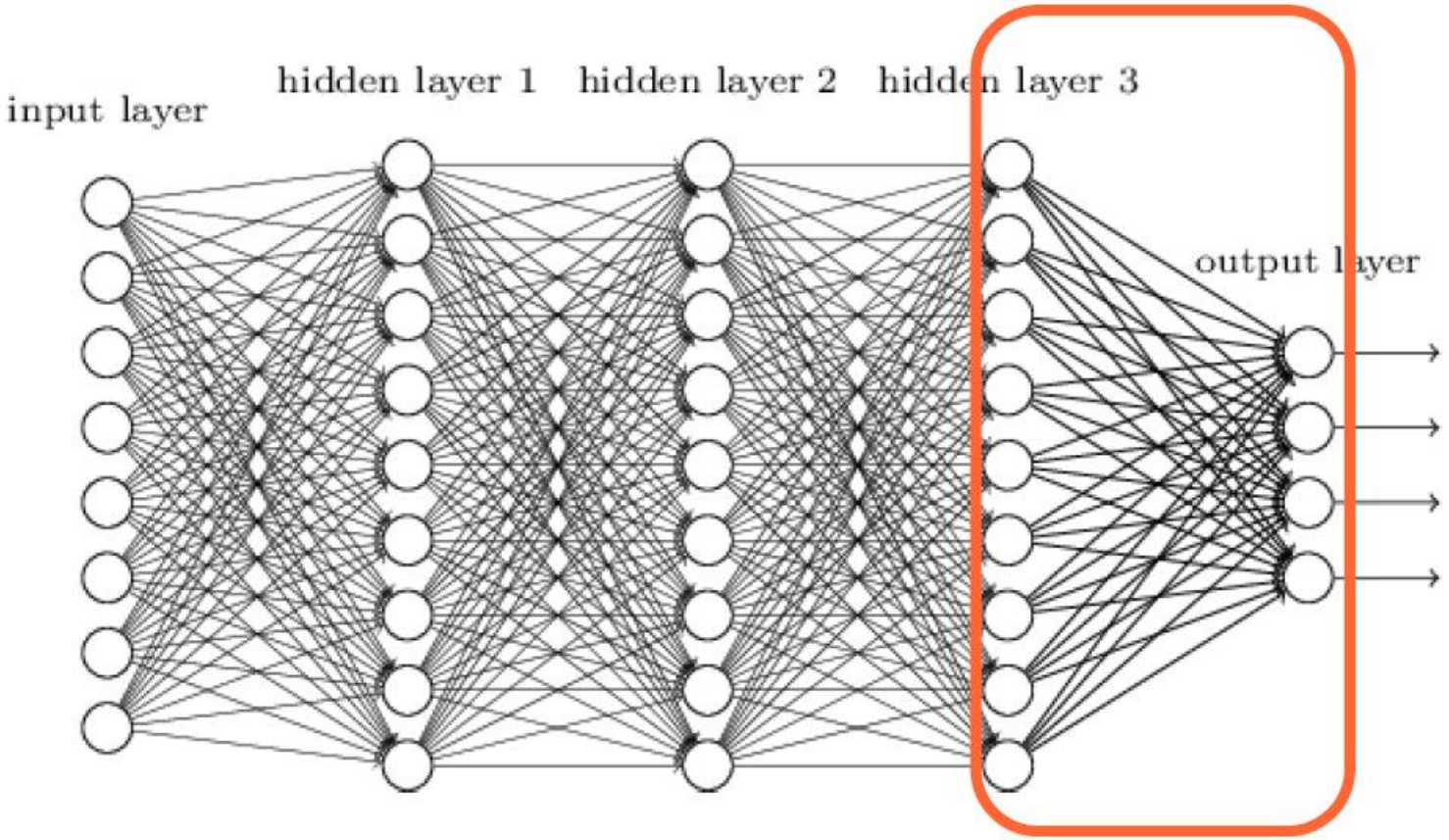
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شبکه‌های عصبی باور عمیق



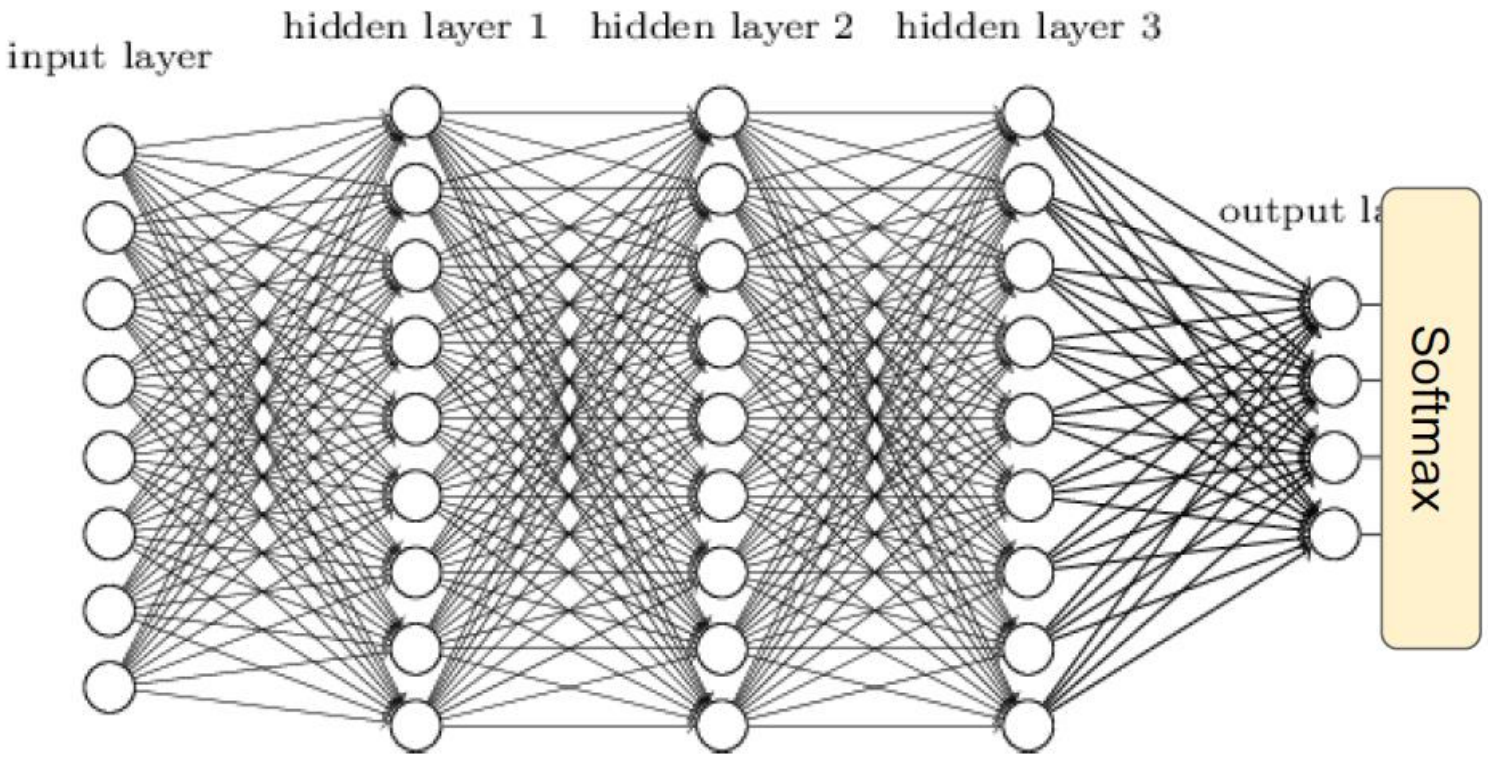
تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شبکه‌های عصبی باور عمیق



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه

شبکه‌های عصبی باور عمیق



تولید محتوا: وحید محمدزاده ایوقی
daychegroup
daychegroup
dayche.com | گروه دایکه