

معسكر النماذج اللّغويّة الكبيرة

The top corners of the slide feature decorative geometric shapes. On the left, there is a brown-to-orange gradient rectangle and a dark blue triangle. On the right, there is a blue-to-purple gradient rectangle and a dark blue triangle. The background is a solid dark blue with faint, larger geometric shapes scattered across it.

التعلم العميق و الانتشار العكسي

Deep Learning & Backpropagation

- مقدمة عن التعلم العميق
- تطبيقات التعلم العميق
- الشبكات العصبية
 - دوال التنشيط
 - شبكة عصبية ذات تغذية أمامية
 - معاملات ضبط الشبكة العصبية
- كيف تتعلم الشبكات العصبية؟

المحتوى

مقدمة عن التعلم العميق

تعلم الآلة

تعلم الآلة هو نوع من الذكاء الاصطناعي يزود أجهزة الكمبيوتر بالقدرة على **التعلم دون أن تتم برمجتها بشكل صريح**.

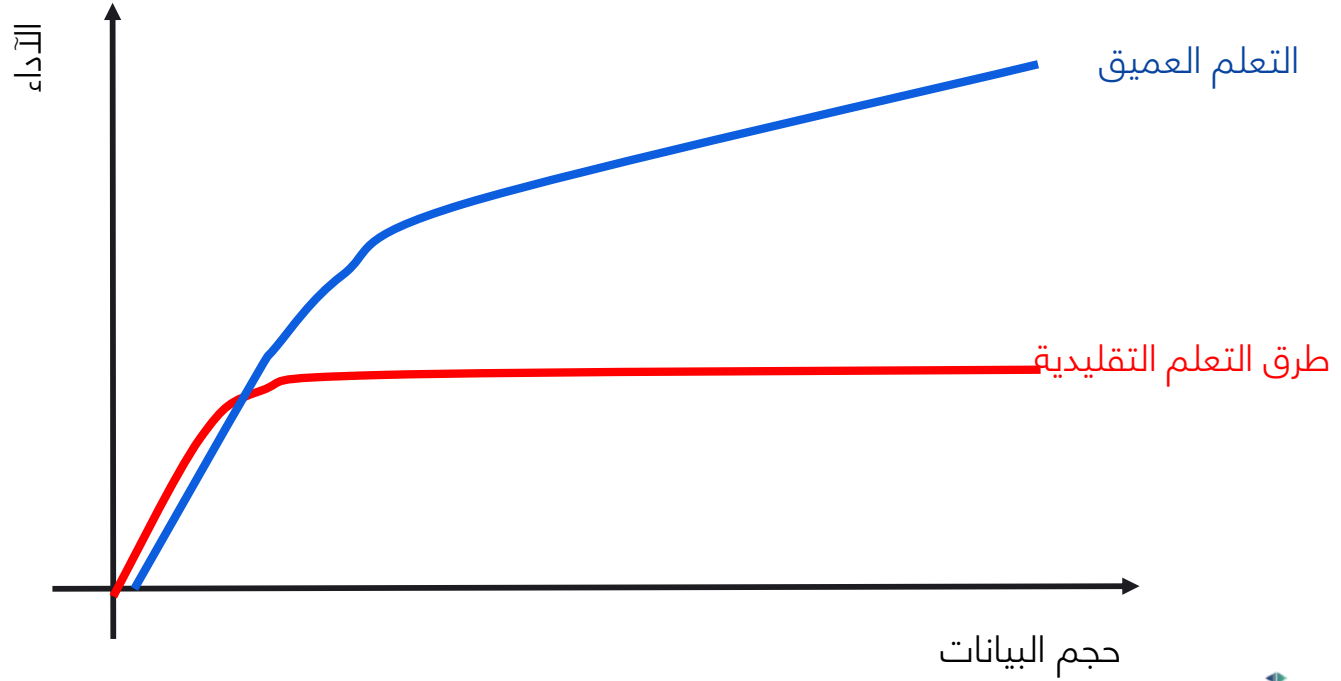


التعلم العميق

- جزء من مجال تعلم الآلة.
- فعال بشكل استثنائي في تحديد وتعلم الأنماط.
- يستخدم خوارزميات التعلم التي تستمد المعنى من البيانات باستخدام تسلسل هرمي من طبقات متعددة تحاكي الشبكات العصبية في الدماغ البشري.



مقارنة



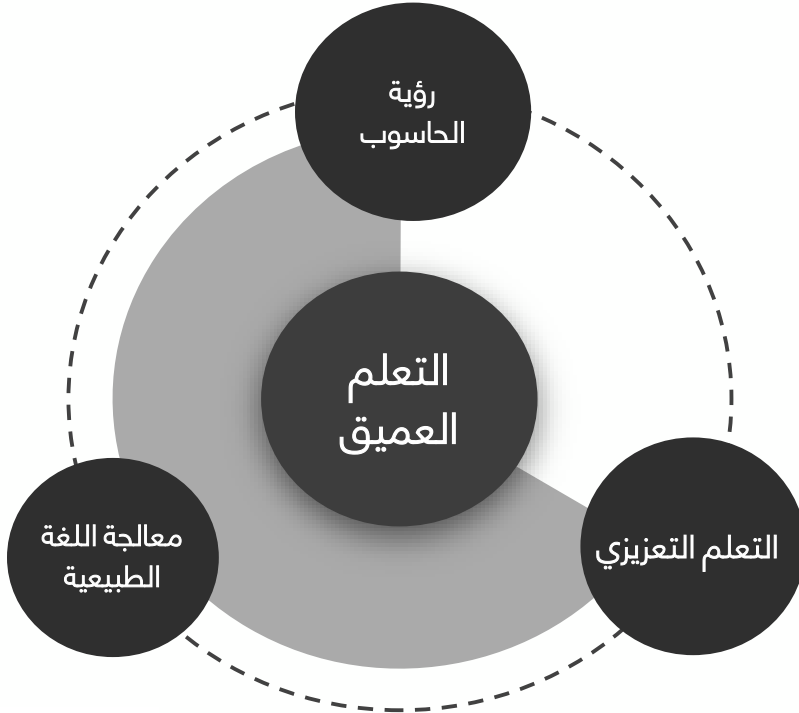
تطبيقات التعلم العميق

تطبيقات التعلم العميق

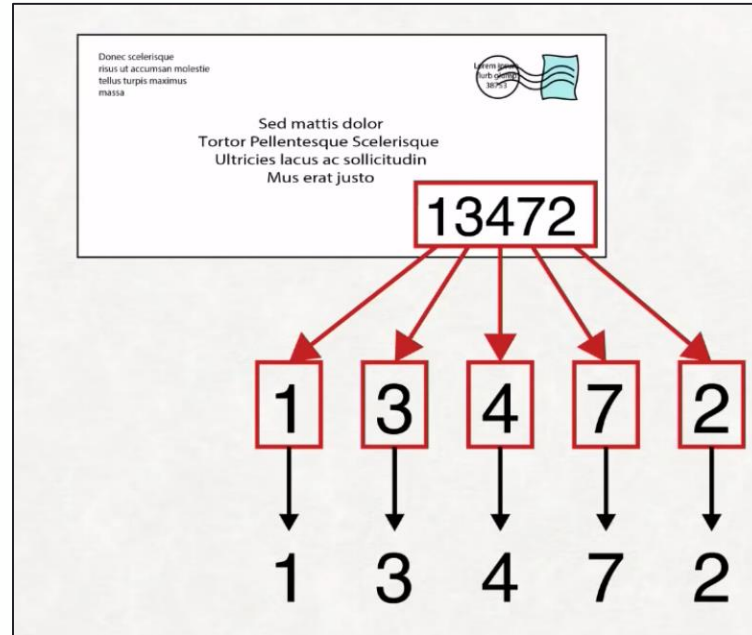
- رؤية الحاسوب: معالجة الصور ، والتعرف على الوجوه ، واكتشاف الكائنات ، وتحليلات الفيديو.

- التعلم التعزيزي العميق

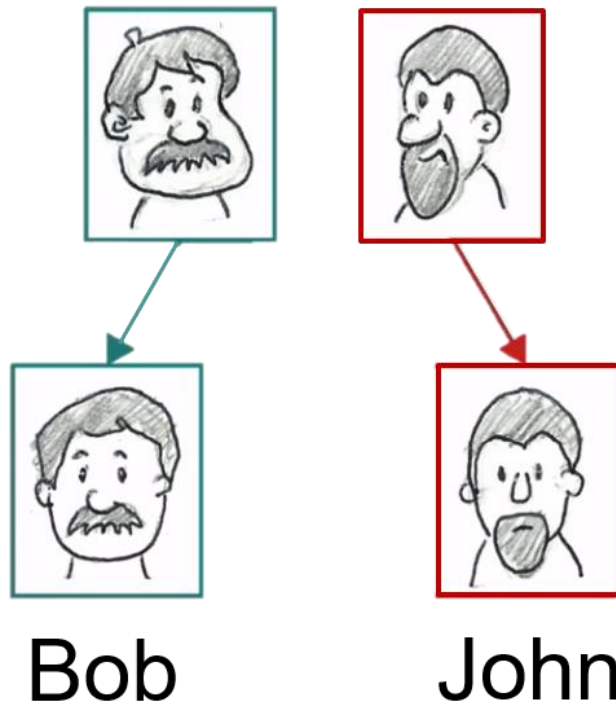
- معالجة اللغة الطبيعية: تحليل المشاعر ، التعرف على الكلام ، تحويل النص إلى كلام ، الذكاء الاصطناعي للمحادثة.



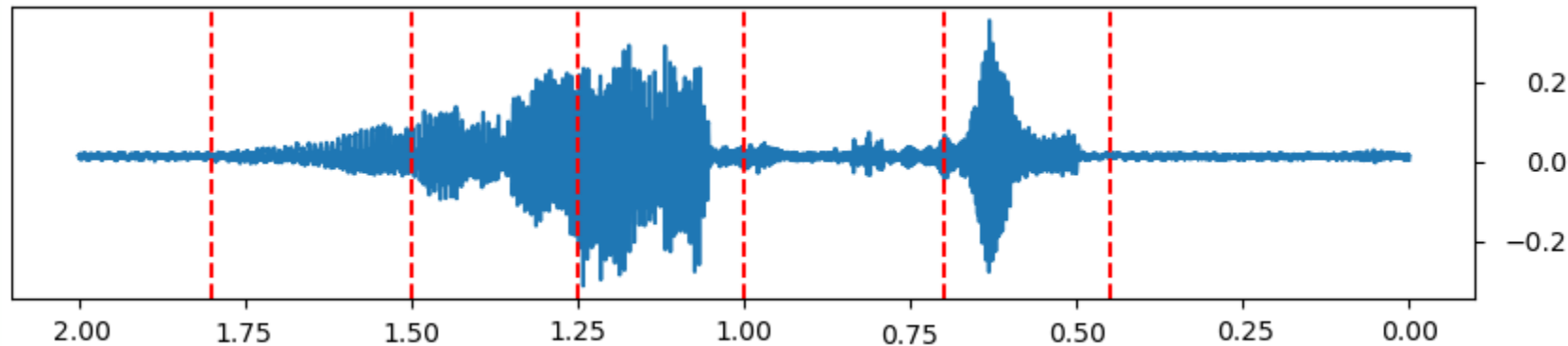
التعرف الآلي على الحروف Optical Character Recognition



التعرف على الوجوه

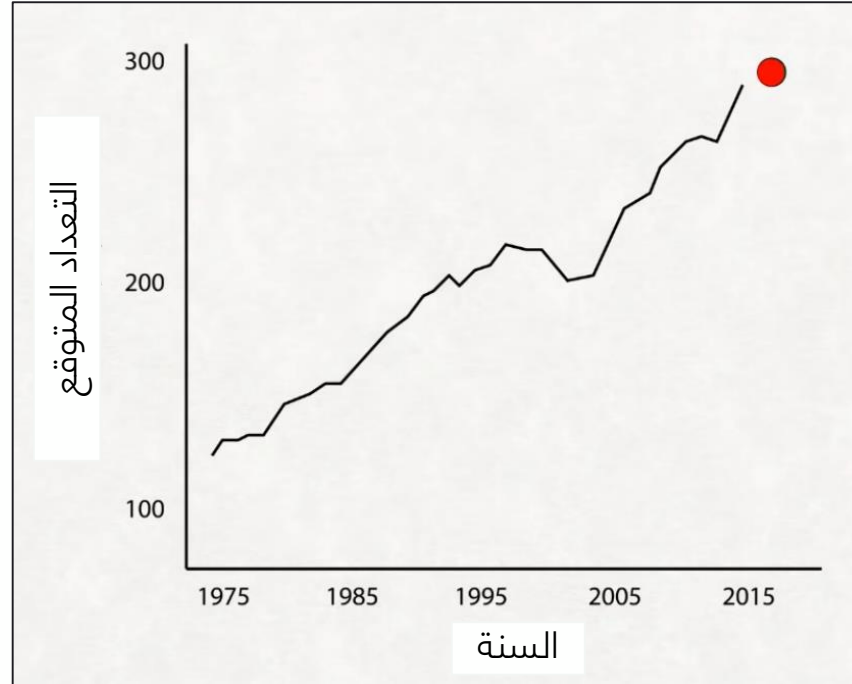


المساعد الافتراضي



لَمْ	عَيِّ	فَ	قُ	رُ
عَيْلَمْ			رُقُ	
رُقْفَعَيْلَمْ				

التحليل الانحداري للسلاسل الزمنية



تنظيف الصور من التشويش

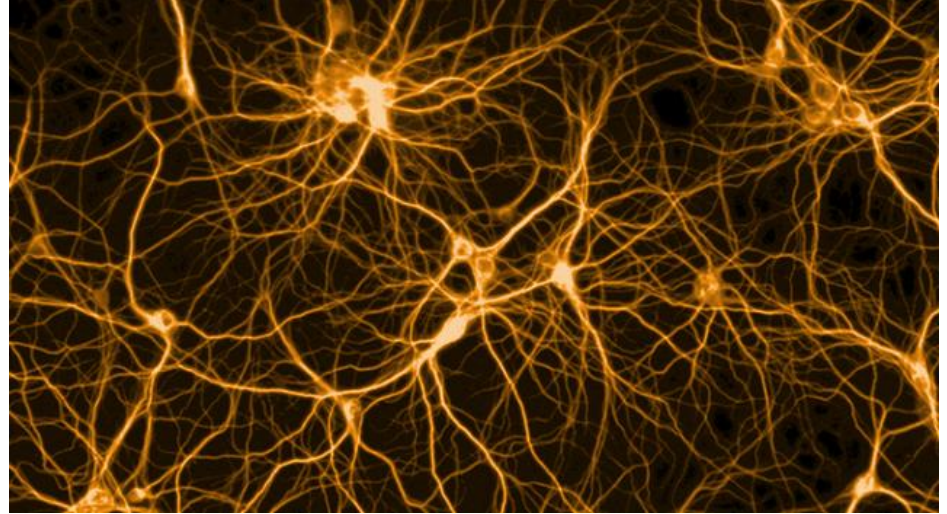


The top left corner of the slide features several abstract geometric shapes. These include a large orange-to-red gradient rectangle, a smaller dark blue triangle, and a larger dark blue hexagon. Other smaller shapes in shades of blue and green are scattered across the top edge of the slide.

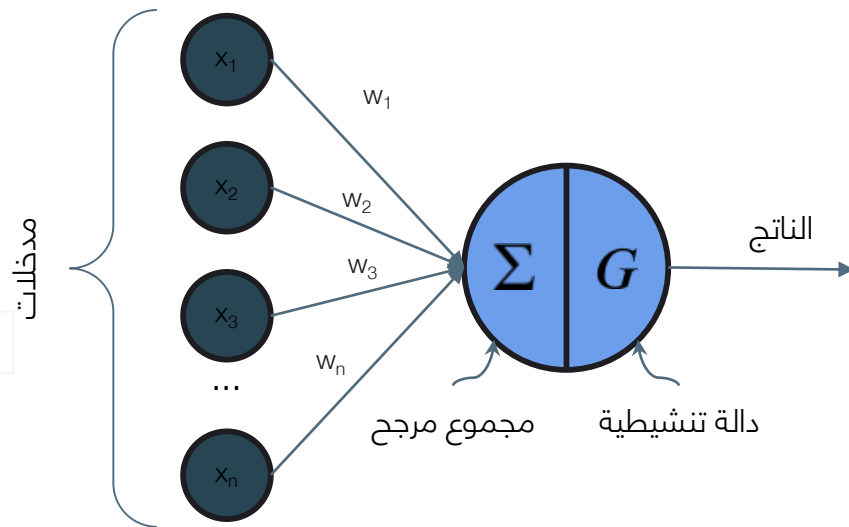
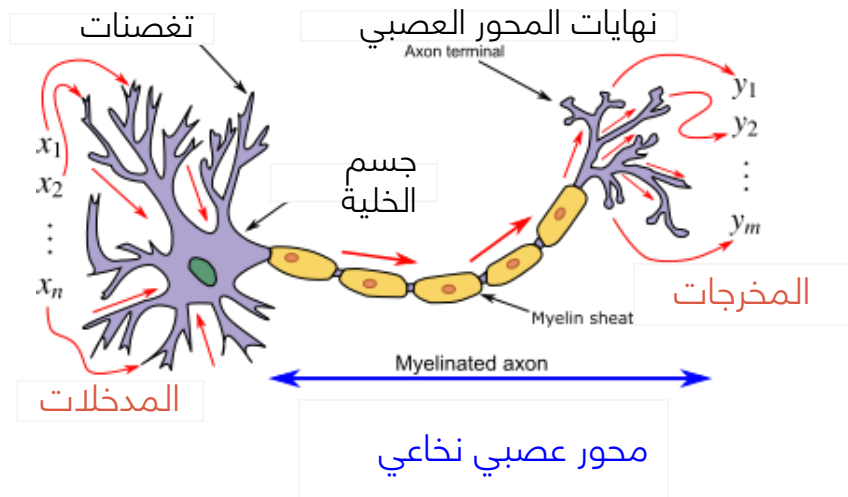
الشبكات العصبية Neural Networks

تشبيه

يملك دماغنا الكثير من الخلايا العصبية المتصلة ببعضها البعض وتمثل قوة الروابط بين الخلايا العصبية معرفة طويلة المدى.

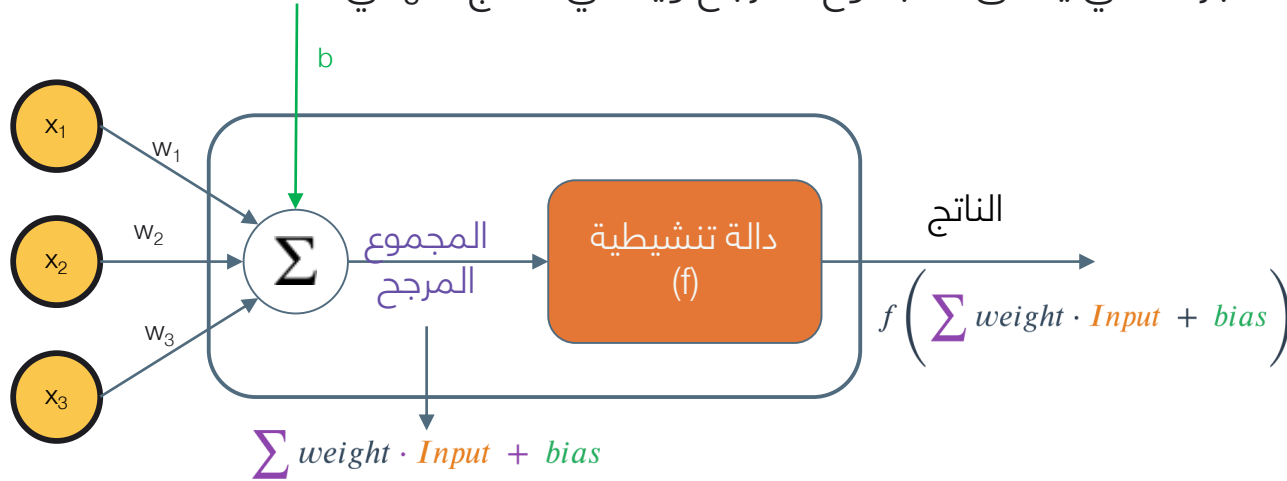


تشبيه



هيكل الشبكة العصبية

- يتكون الجزء الداخلي من الخلايا العصبية الاصطناعية من جزأين أساسيين:
- الجزء الأول يحتسب المجموع المرجح للمدخلات أي الناتج الصافي
 - الجزء الثاني يتلقى المجموع المرجح ويعطي الناتج النهائي.



The top corners of the slide feature decorative geometric shapes. On the left, there is a brown-to-orange gradient rectangle and a dark blue triangle. On the right, there is a dark blue triangle and a blue-to-purple gradient rectangle. In the center top, there are several dark blue polygons, including a hexagon and a diamond.

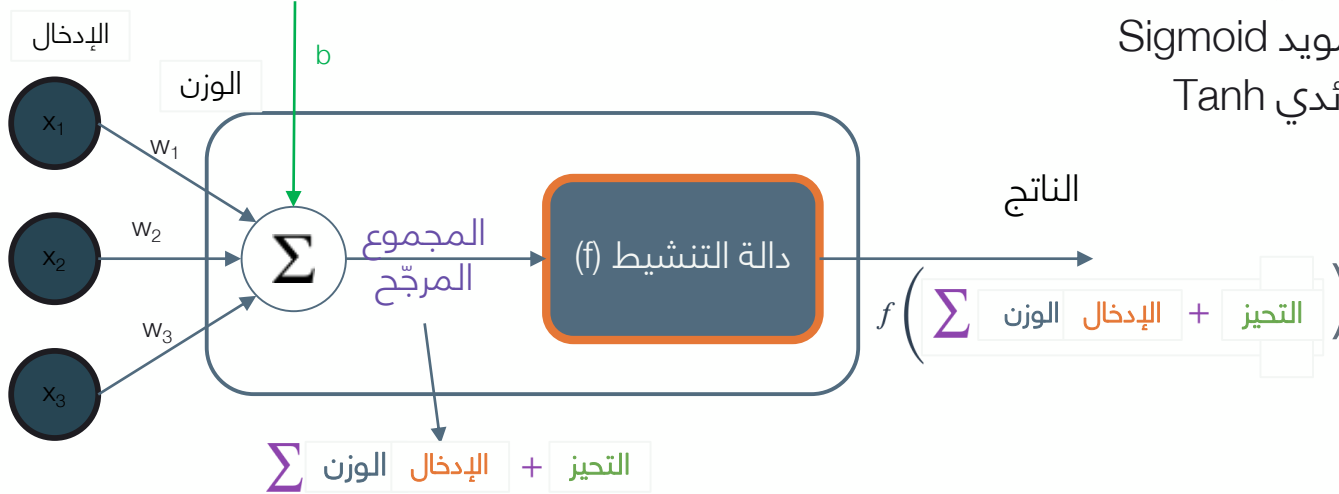
دوال التنشيط

Activation Functions

دوال التنشيط الأكثر استخداماً

فيما يلي أشهر دوال التنشيط:

- ❖ خطوة ثنائية Binary step
- ❖ وحدة خطية مصححة ReLU
- ❖ سيجمويد Sigmoid
- ❖ ظل زائدي Tanh

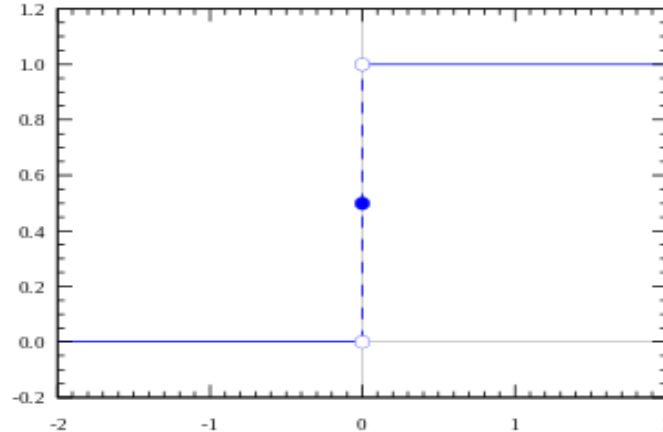


خطوة ثنائية (Binary Step)

تخبر ما إذا كانت قيمة الإدخال أعلى أو أقل من 0

$$y = \text{sign}(x)$$

قيمة الإدخال (x)

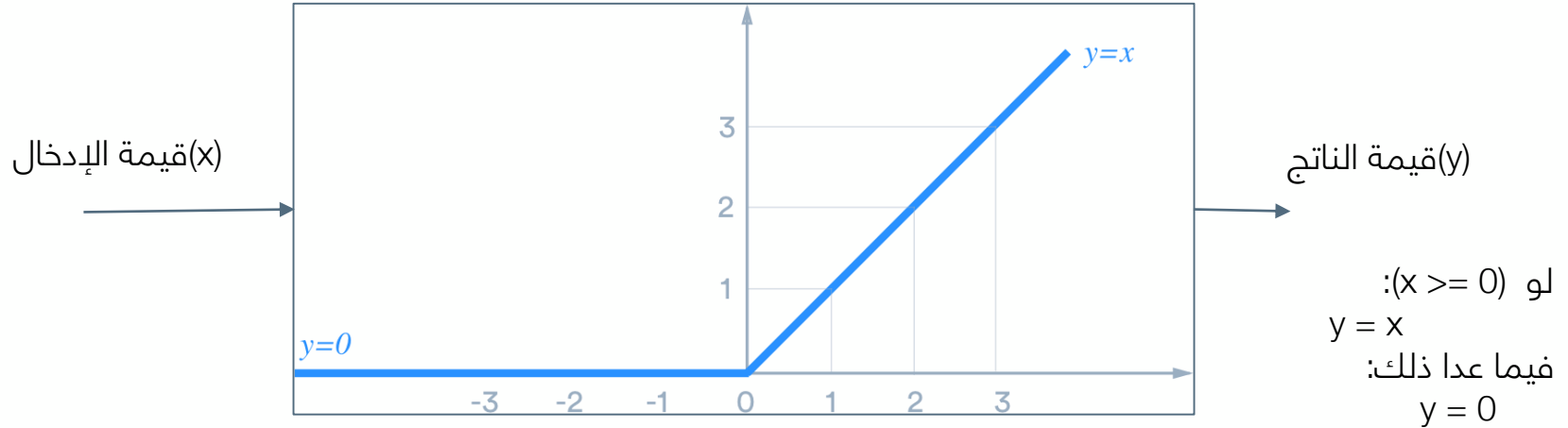


قيمة الناتج (y)

لو $(x \geq 0)$:
 $y = 1$
فيما عدا ذلك:
 $y = 0$

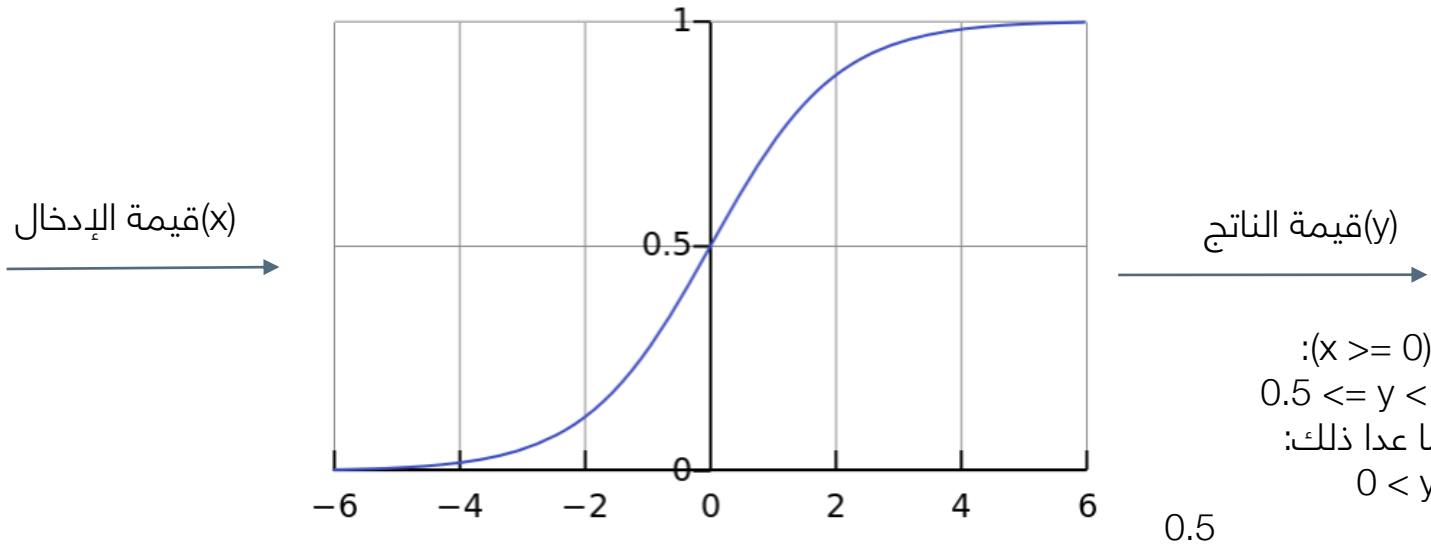
وحدة خطية مصححة (ReLU)

$$Y = \max(0, x)$$

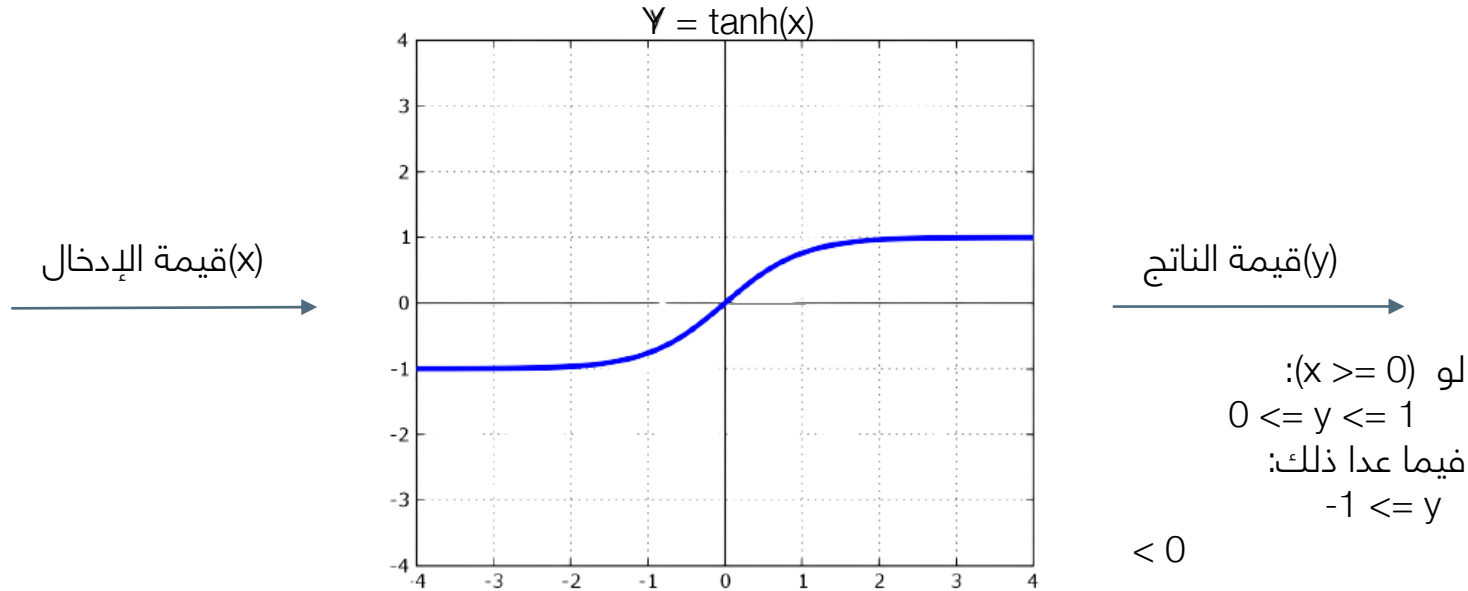


سيجمويد (Sigmoid)

$$y = \frac{1}{1 + e^{-x}}$$

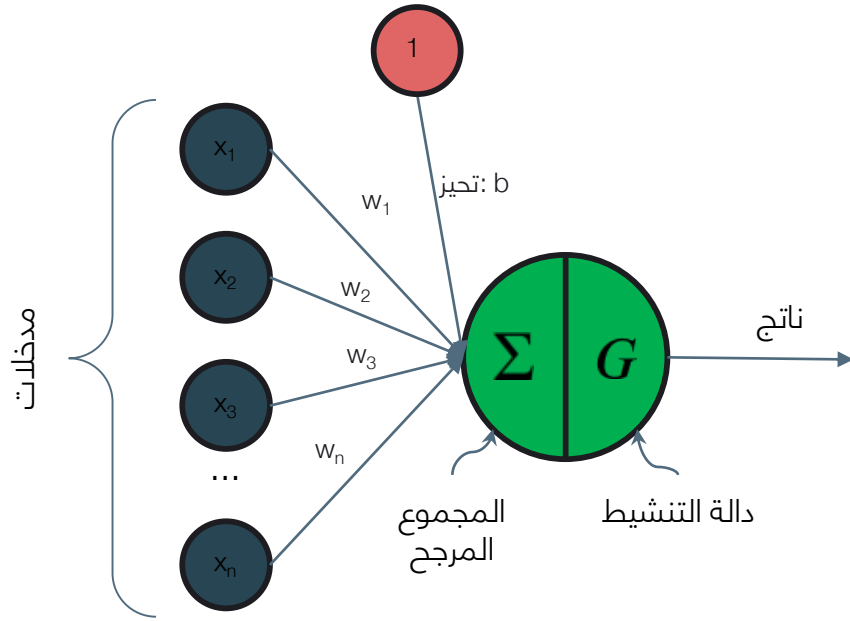


الظل الزائدي (\tanh)



شبكة عصبية ذات تغذية أمامية

بيرسبترون



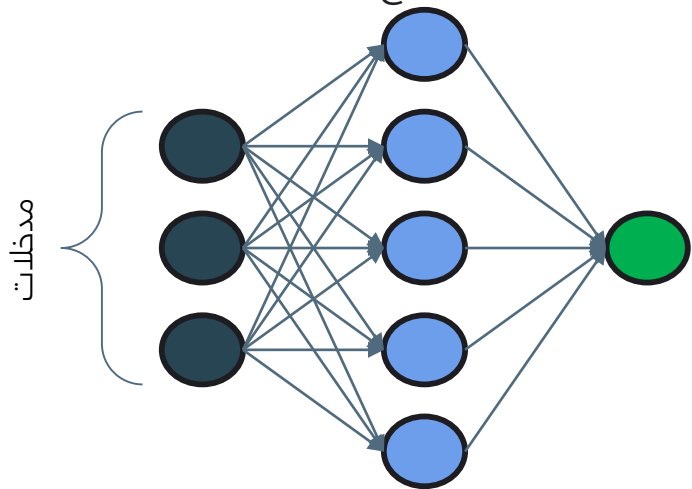
- خلية عصبية ذات ناتج واحد
- دالة التنشيط: خطوة ثنائية (Binary Step)
- تستخدم للتصنيف بين الفئات القابلة للفصل خطيًا.

$$y = \text{sign} \left(\sum_{i=1}^n w_i x_i + b \right)$$

المعاملات الخاضعة للتدريب

شبكة ذات طبقة مخفية واحدة

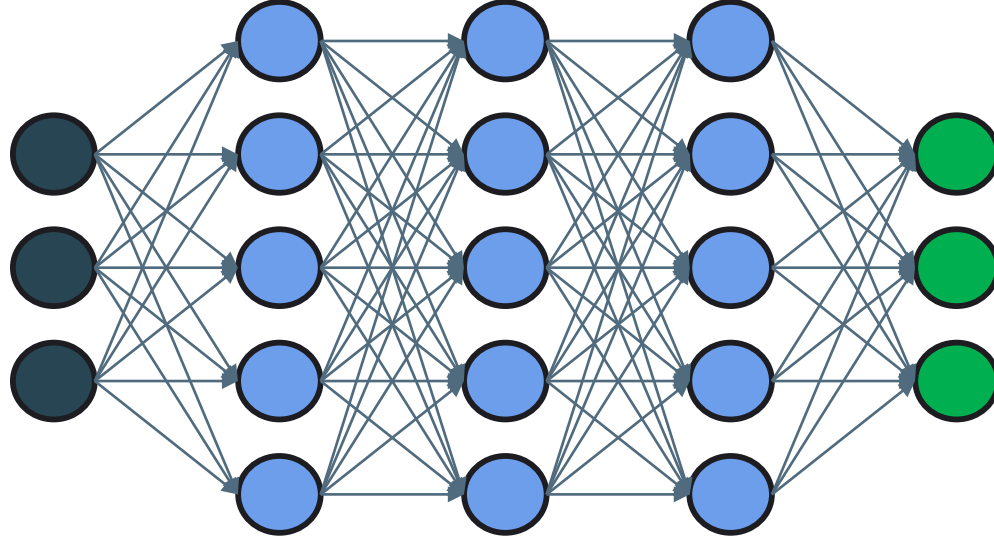
- لدينا طبقة واحدة تسمى "مخفية" بين طبقات الإدخال والإخراج.



- يحسب المجموع المرجح لمخرجات الطبقة السابقة.
- يتم تمرير المجموع إلى دالة التنشيط للحصول على الناتج النهائي.

- يحسب المجموع المرجح للمدخلات.
- يمرر المجموع إلى وظيفة التنشيط للحصول على الإخراج ليتم تمريره كمدخل للطبقة التالية.

الشبكات العصبية العميقة



وحدات الإدخال

وحدات مخفية

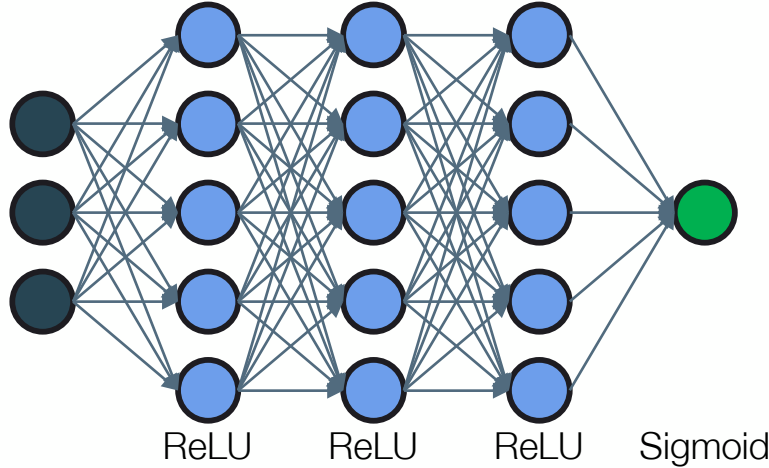
وحدات الإخراج

The top corners of the slide feature abstract geometric shapes. On the left, there is a brown and orange parallelogram. On the right, there is a blue and purple parallelogram. The background is a solid dark blue with several faint, light blue geometric shapes (triangles and diamonds) scattered across it.

معاملات ضبط الشبكات العصبية

Neural Network Hyperparameters

أنواع الدوال التنشيطية

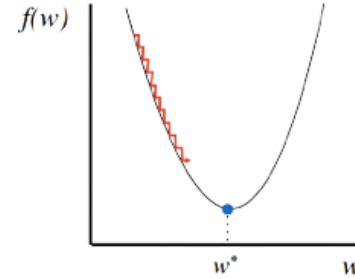


- لا توجد قاعدة أساسية
- يميل المطورون إلى استخدام ReLU في طبقاتهم المخفية.
- يجب أن يكون تنشيط طبقة الإخراج متسقًا مع نوع المشكلة.
- لا يتم استخدام الخطوة الثنائية كثيرًا لأنها تحتوي على انقطاع عند القيمة 0 وهذا يؤدي إلى مشاكل في المشتقات المحسوبة أثناء مرحلة التعلم.

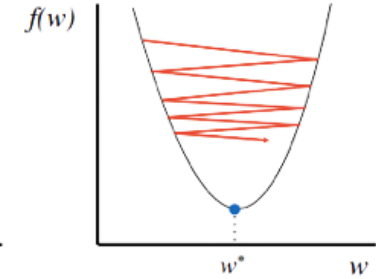
معدل التعلم

- يعد اختيار معدل التعلم الصحيح خطوة أساسية للوصول إلى التقارب.

$$w := w - \alpha \frac{\partial J}{\partial w}$$



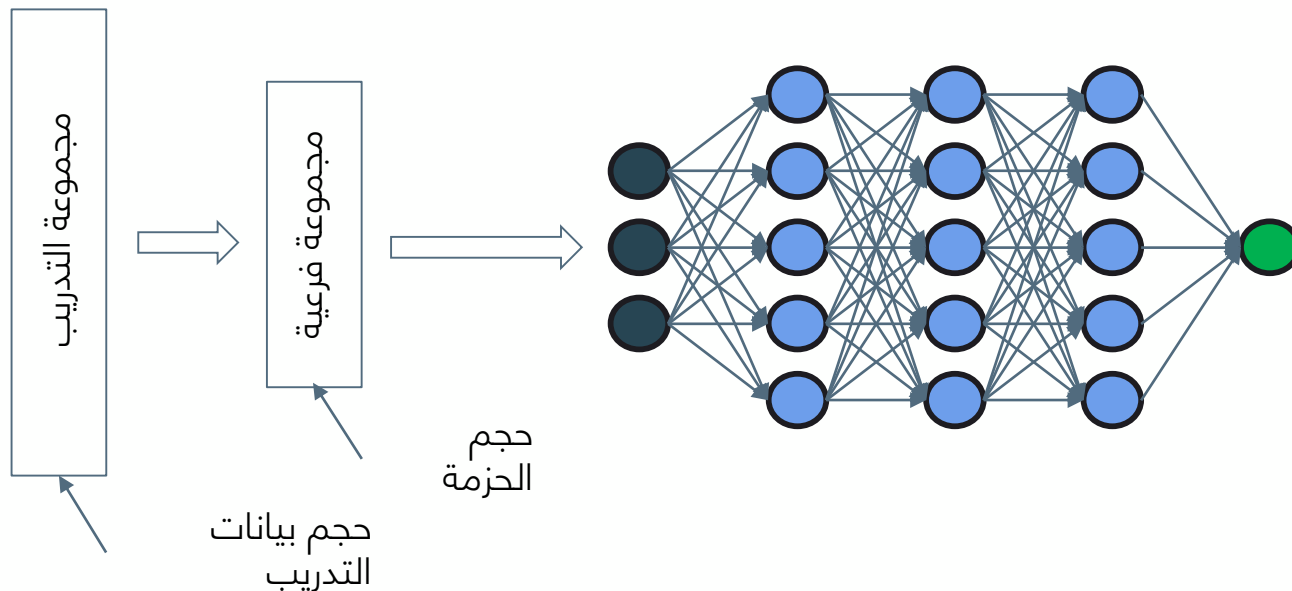
صغير جدا: تقارب
بطيء



كبير جدا: تجاوز وتباعد
محتمل

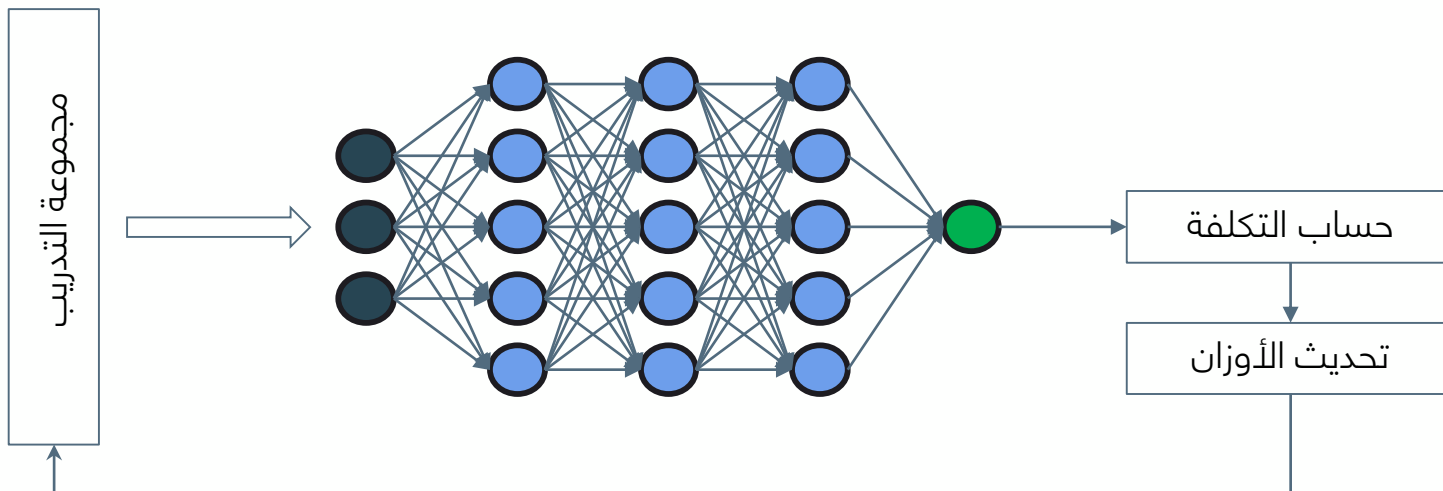
حجم الحزمة

كمية البيانات التي تغذيها للشبكة قبل إجراء تحديث في الأوزان.

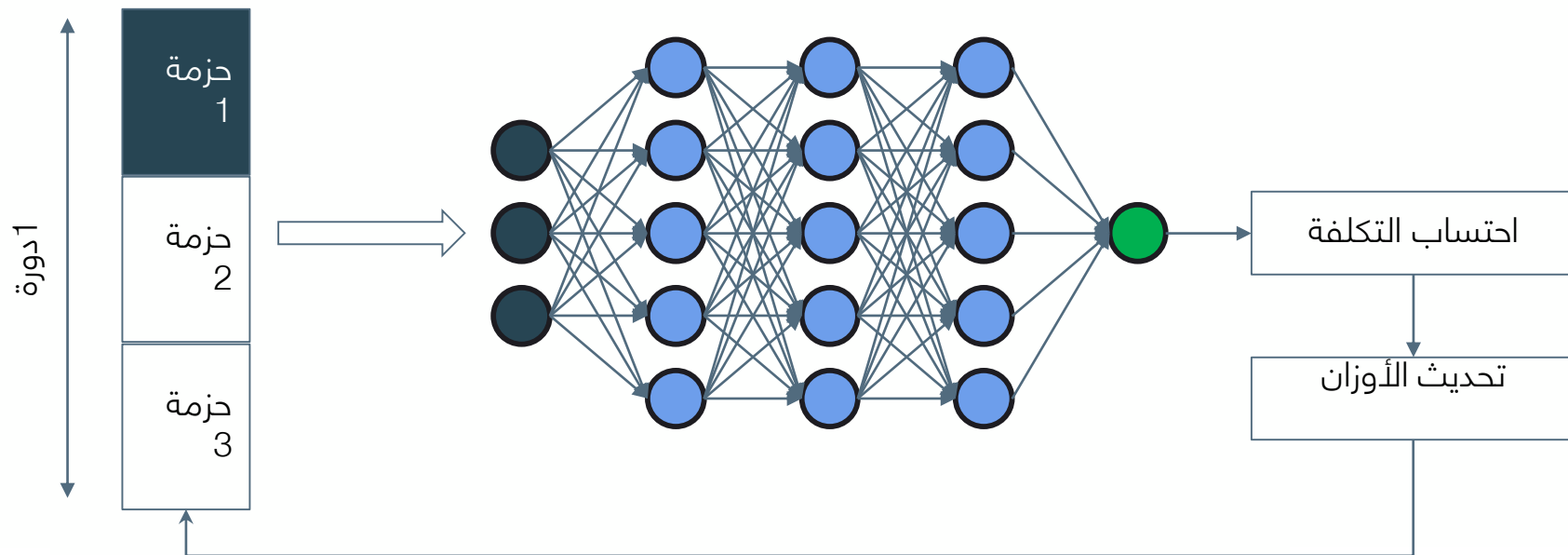


الدورات

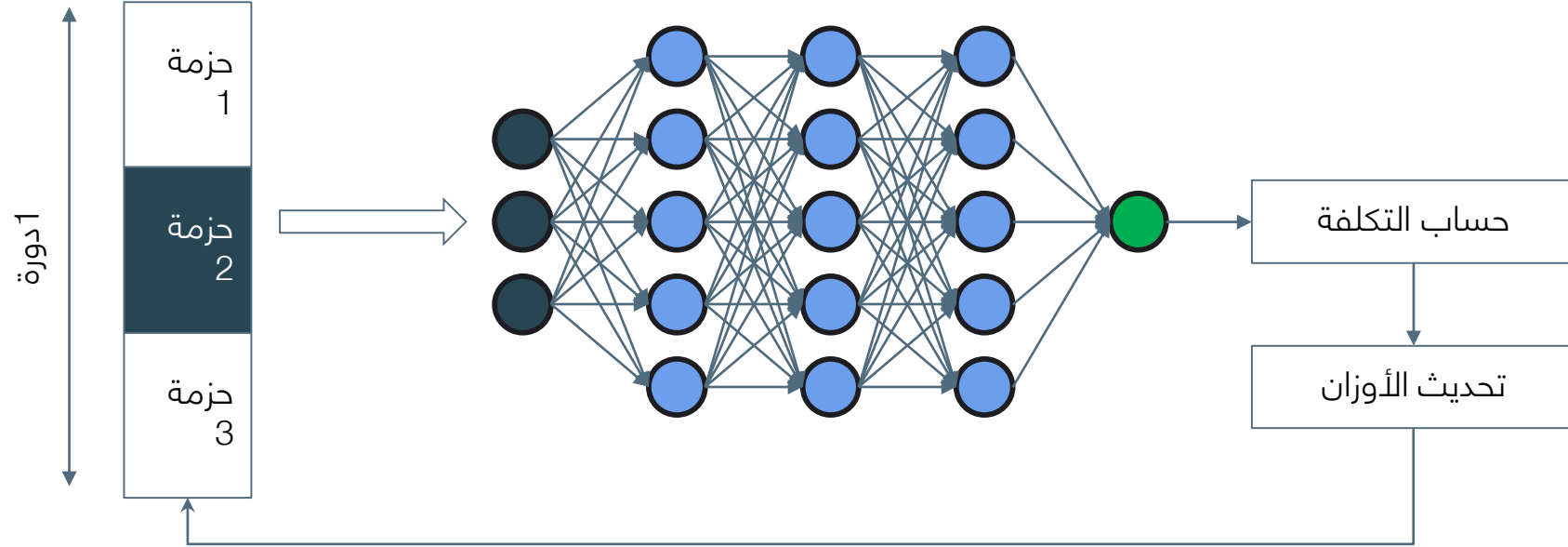
عدد المرات التي نغذي فيها جميع أمثلة مجموعة التدريب لشبكتنا.



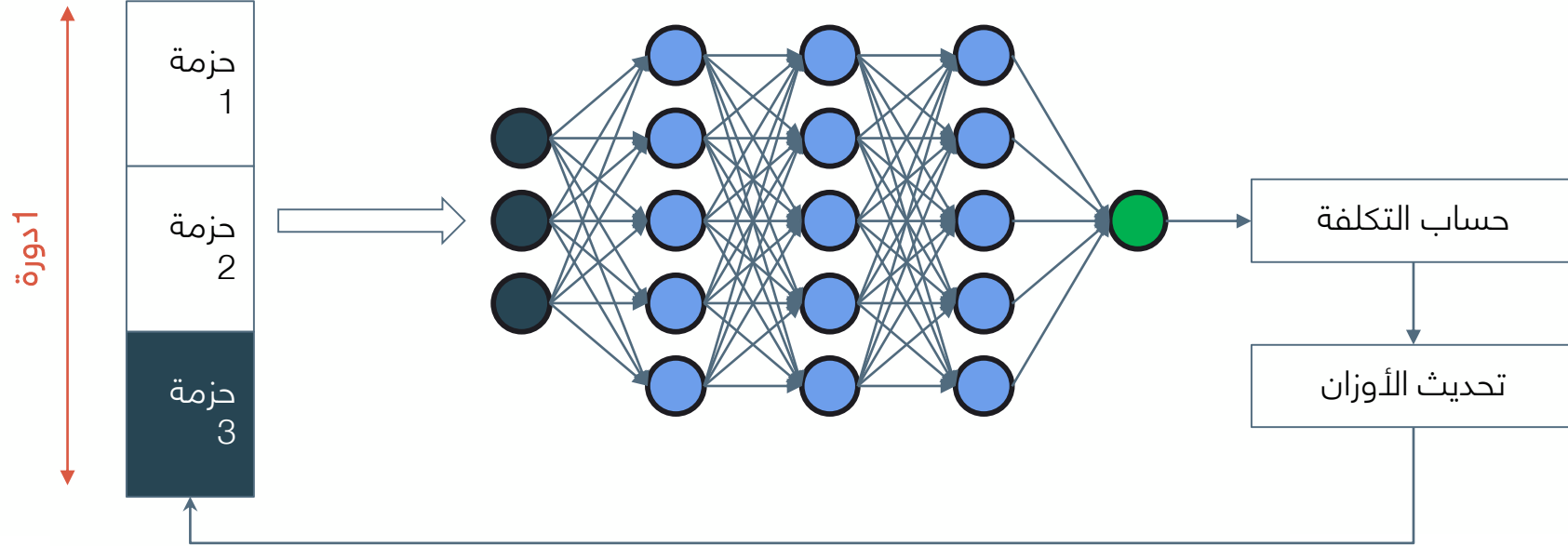
حجم الحزمة مقابل الدورات



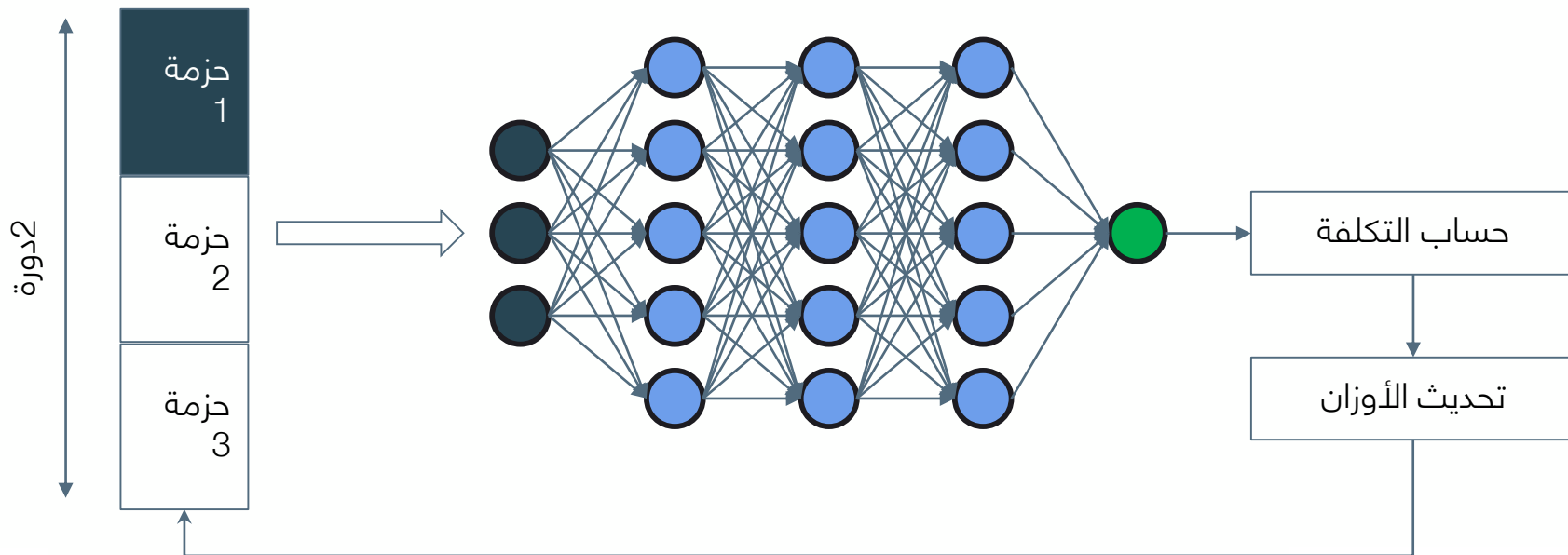
حجم الحزمة مقابل الدورات




حجم الحزمة مقابل الدورات



حجم الحزمة مقابل الدورات





التدريب العملي:

بناء شبكتك العصبية الأولى

Hands-on: Build your first NN

كيف تتعلم الشبكات العصبية؟

هيكل البيانات

نأخذ حالة عامة حيث تتكون بيانات الإدخال من:

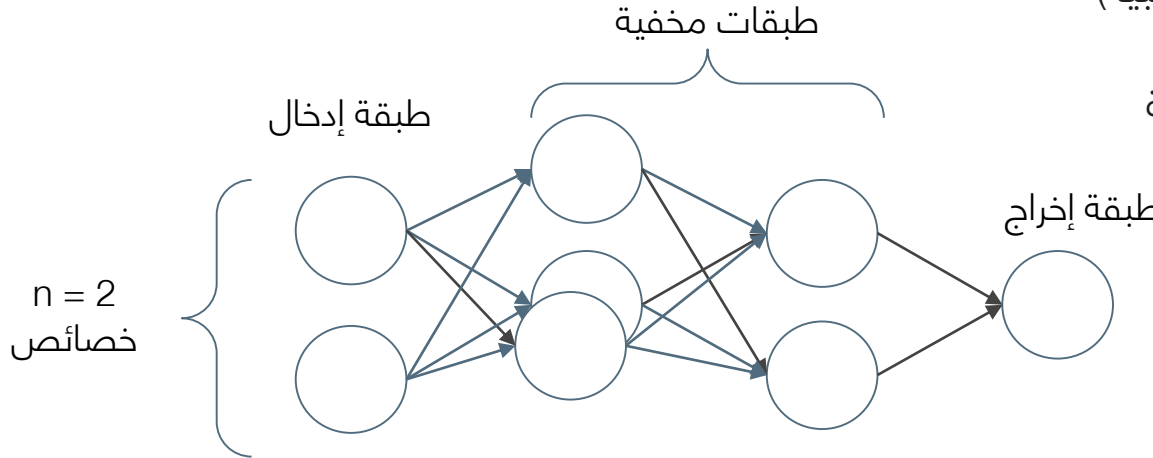
- عدد m من الأمثلة
- كل منهم يحتوي على n من الخصائص
- ناتج واحد فقط

		m = 5 الأمثلة					
n = 2 خصائص	{	x0	0.2	0.2	0.2	0.3	0.2
		x1	0.1	0.3	0.1	0.4	0.1
ناتج		y	1	0	1	0	1

هيكل الشبكة

في الحالة العامة التي اتخذناها ،
تتألف الشبكة العصبية من:

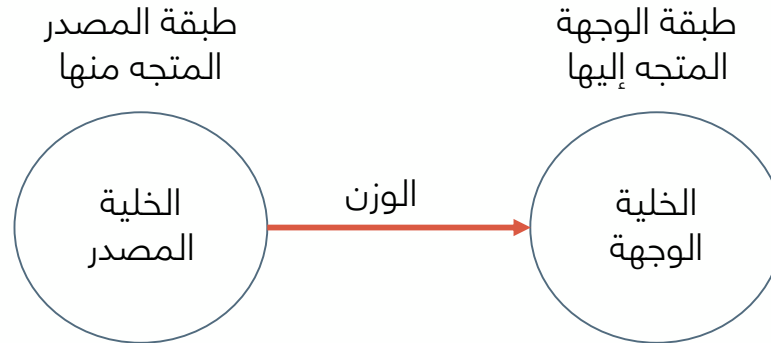
- طبقة إدخال (من n خلية عصبية)
- طبقات مخفية
- طبقة إخراج (من خلية عصبية واحدة)



الأوزان

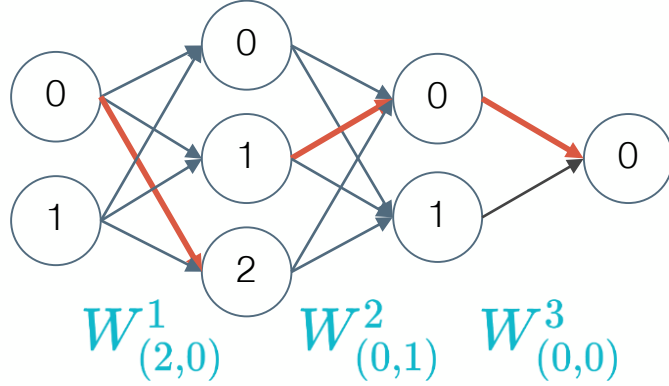
الوزن $W_{(i,j)}^l$ يتم تعريفه بثلاث خصائص:

- / فهرس طبقة الوجهة (المتجه إليها)
- / فهرس الخلية العصبية بالطبقة
- / فهرس الخلية العصبية بالطبقة $l-1$ (خلية المصدر)



الأوزان

طبقة 0	طبقة 1	طبقة 2	طبقة 3
--------	--------	--------	--------



الوزن $W^l_{(i,j)}$ يتم تعريفه بثلاث خصائص:

- / فهرس طبقة الوجهة (المتجه إليها)

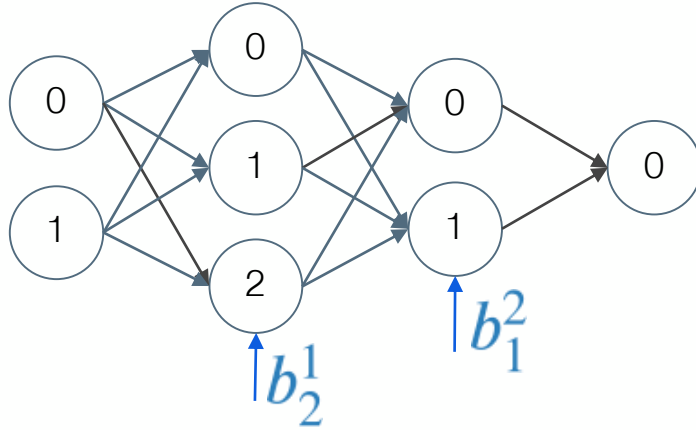
- / فهرس الخلية العصبية بالطبقة

- / فهرس الخلية العصبية بالطبقة (1- / خلية المصدر)

عادة ما يتم تهيئة قيم الأوزان الأولية بشكل عشوائي

التحيز

طبقة 0	طبقة 1	طبقة 2	طبقة 3
--------	--------	--------	--------



عادة ما يتم تهيئة قيم التحيز الأولية بشكل عشوائي

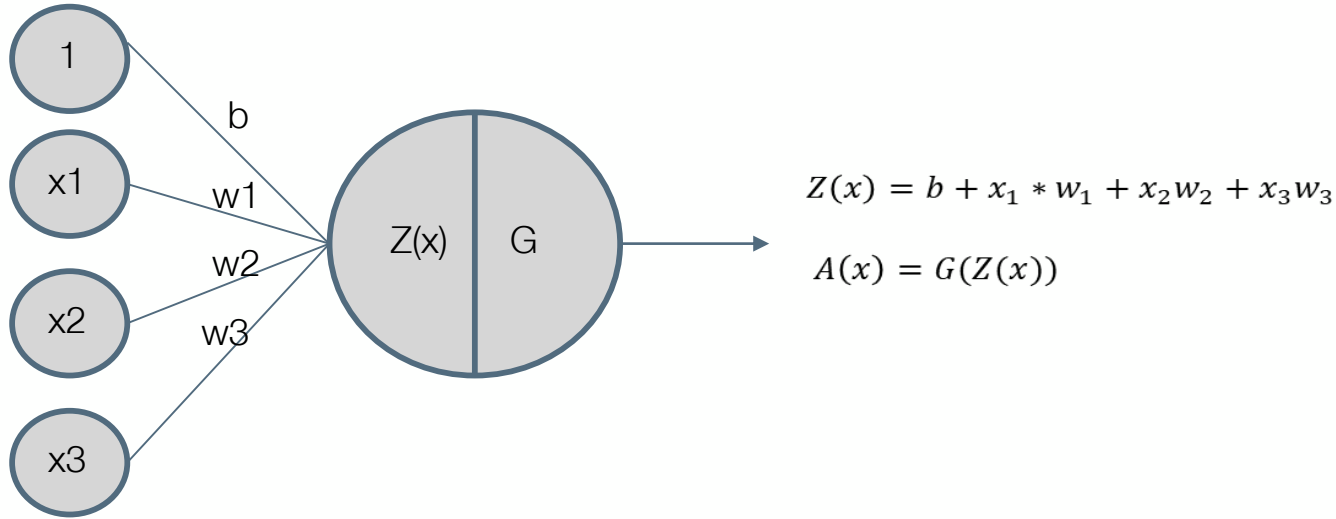
- التحيز هو سمة من سمات الخلية العصبية.
- يتم تمثيل التحيز بالرمز b_i^l :
 - i هو فهرس الخلية الوجهة.
 - l هو فهرس الطبقة الوجهة.

Abstract geometric shapes in shades of blue, purple, and orange are positioned in the top corners of the slide.

الانتشار الأمامي Feed Forward

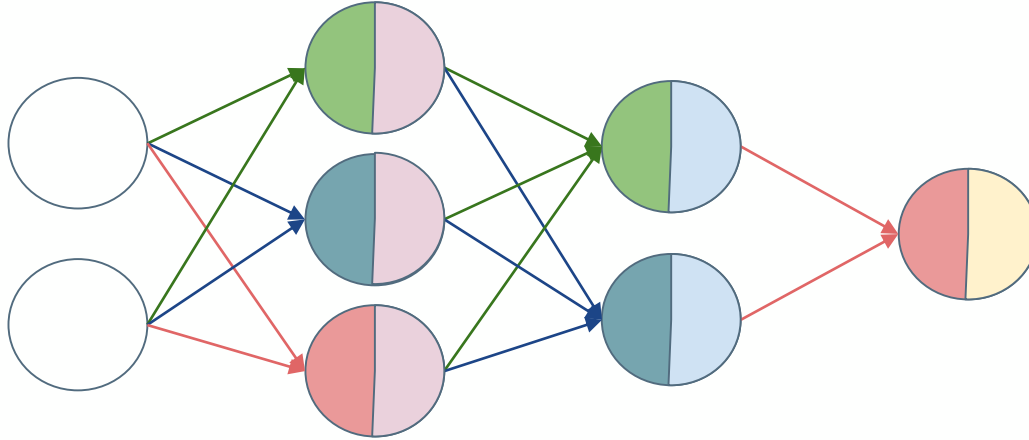
الانتشار الأمامي

تذكر الدالة التي تعمل عليها خلية عصبية واحدة لها مدخلات متعددة.



الانتشار الأمامي

- سيتم الآن تنفيذ هذه الدالة لكل خلية عصبية في كل طبقة من طبقات الشبكة على النحو التالي:
- تعمل الخلايا العصبية في كل طبقة في وقت واحد.
 - دوال التنشيط هي نفسها في كل طبقة.

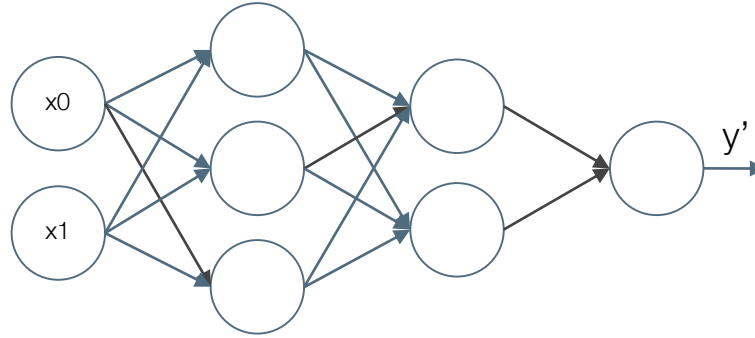


The top left corner of the slide features several overlapping geometric shapes, including a large orange-to-red gradient rectangle, a smaller dark blue triangle, and a larger dark blue hexagon. Other smaller geometric shapes are scattered across the top edge.

دالة التكلفة Cost Function

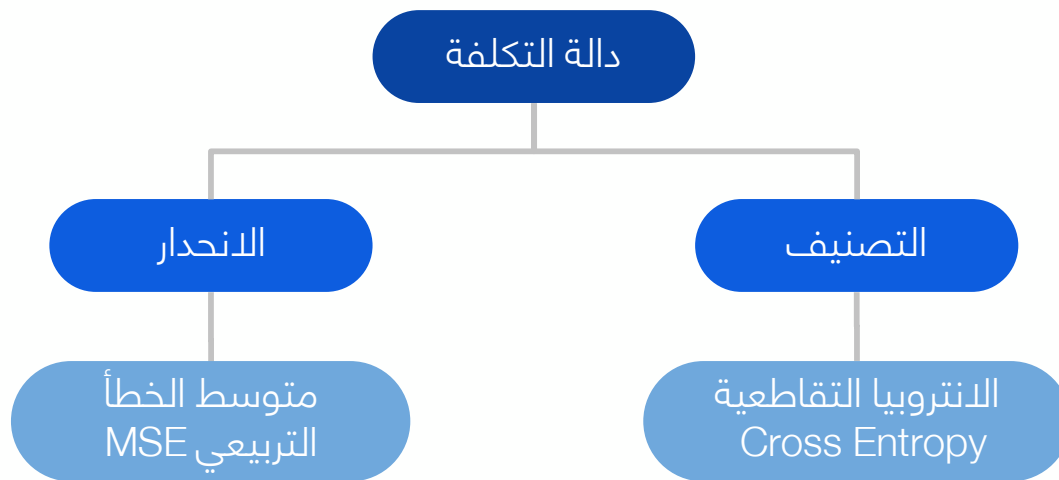
دالة الخسارة Loss Function

- بعد اكتمال الانتشار الأمامي ، نحصل على القيمة المتوقعة y' المقابلة للمدخلات التي قدمناها.
- نحدد دالة الخسارة على أنها **مقياس الخطأ** بين القيمة المتوقعة (y') والقيمة الصحيحة (y) لمدخل معين.



دالة التكلفة Cost Function

تخبرك دالة التكلفة بنسبة "ارتكاب الأخطاء" من خلال احتساب متوسط دوال الخسارة المحسوبة على جميع أمثلة التدريب الفردية للنموذج.



الانتشار العكسي

تحاول هذه الخوارزمية تقليل دالة التكلفة لعن طريق ضبط الأوزان.
المعطيات:

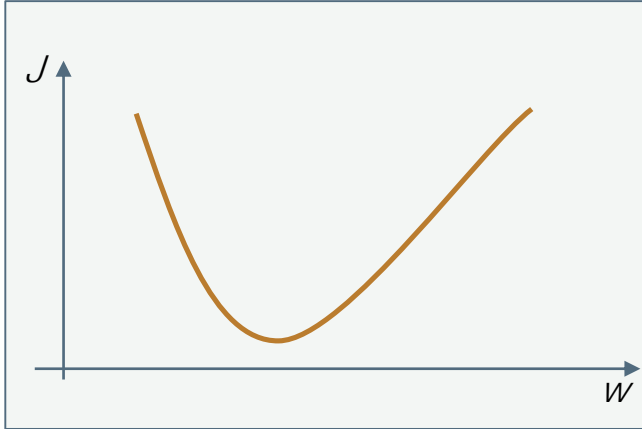
- دالة التكلفة :

$$J = f(w)$$

- مشتقة دالة التكلفة بالنسبة للأوزان:

$$\frac{dJ}{dw} = \frac{df(w)}{dw}$$

- معدل التعلم α



النزول الاشتقاقي

سنستخدم خوارزمية النزول الاشتقاقي لإيجاد الحد الأدنى:

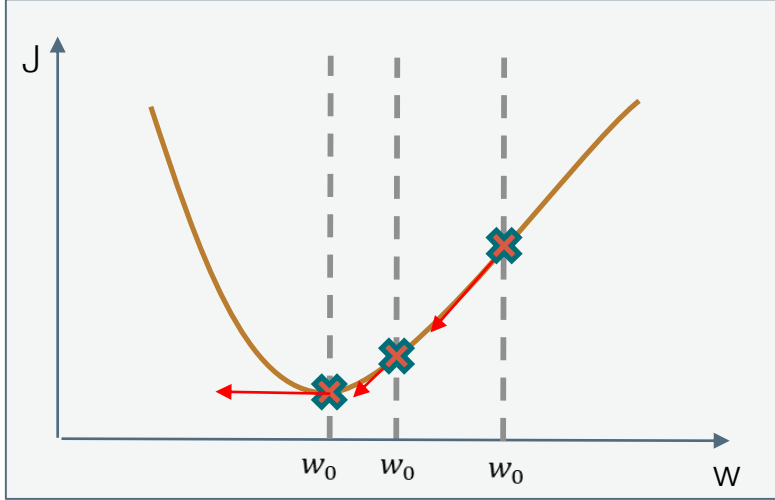
- اختر نقطة البداية w_0 بشكل عشوائي.

- احتسب $\frac{dJ}{dw}|_{w=w_0}$

- عدل قيمة w وفقا للمعادلة:

$$w_0 := w_0 - \alpha \frac{dJ}{dw}|_{w=w_0}$$

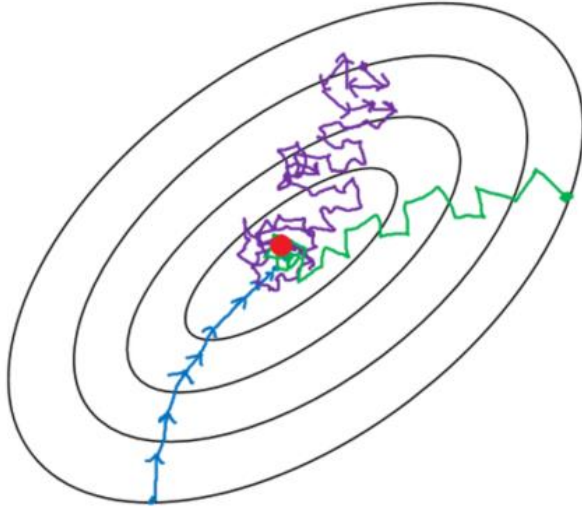
- كرر من الخطوة رقم 2



مختلفات النزول ، الاشتقاقي ،

عدد التحديثات في الدورة الكاملة	عدد الأمثلة في الخطوة الواحدة لتحديث الأوزان	مسألة التحسين
1	مجموعة البيانات كاملة	نزول اشتقاقي بحزمة كاملة Batch Gradient Descent
الحجم الكلي مقسماً على حجم الحزمة الصغيرة	مجموعة فرعية من البيانات	نزول اشتقاقي بحزم صغيرة Mini-batch Gradient Descent
عدد الأمثلة في المجموعة كاملة	مثال واحد فقط	نزول اشتقاقي عشوائي Stochastic Gradient Descent

أنواع النزول الاشتقاقي



- Batch gradient descent
- Mini-batch gradient Descent
- Stochastic gradient descent

نزل اشتقاقي بحزمة كاملة
نزل اشتقاقي بحزم صغيرة
نزل اشتقاقي عشوائي (بمثال واحد)

الانتشار العكسي

Backward Pass

الانتشار العكسي

ينطبق مبدأ النزول الاشتقاقي نفسه عندما تتعامل مع دالة (J) لمتغيرات متعددة (أوزان وتحيز)

$$\frac{\partial J}{\partial W^3} = \begin{bmatrix} \frac{\partial J}{\partial W_{00}^3} & \frac{\partial J}{\partial W_{01}^3} \end{bmatrix} \quad \frac{\partial J}{\partial b^3} = \begin{bmatrix} \frac{\partial J}{\partial b_0^3} \end{bmatrix}$$

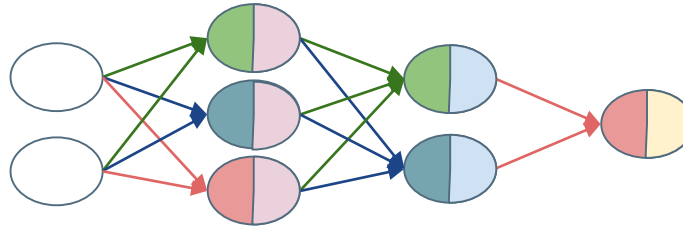
$$\frac{\partial J}{\partial W^2} = \begin{bmatrix} \frac{\partial J}{\partial W_{00}^2} & \frac{\partial J}{\partial W_{01}^2} & \frac{\partial J}{\partial W_{02}^2} \\ \frac{\partial J}{\partial W_{10}^2} & \frac{\partial J}{\partial W_{11}^2} & \frac{\partial J}{\partial W_{12}^2} \end{bmatrix} \quad \frac{\partial J}{\partial b^2} = \begin{bmatrix} \frac{\partial J}{\partial b_0^2} \\ \frac{\partial J}{\partial b_1^2} \end{bmatrix}$$

$$\frac{\partial J}{\partial W^1} = \begin{bmatrix} \frac{\partial J}{\partial W_{00}^1} & \frac{\partial J}{\partial W_{01}^1} \\ \frac{\partial J}{\partial W_{10}^1} & \frac{\partial J}{\partial W_{11}^1} \\ \frac{\partial J}{\partial W_{20}^1} & \frac{\partial J}{\partial W_{21}^1} \end{bmatrix} \quad \frac{\partial J}{\partial b^1} = \begin{bmatrix} \frac{\partial J}{\partial b_0^1} \\ \frac{\partial J}{\partial b_1^1} \\ \frac{\partial J}{\partial b_2^1} \end{bmatrix}$$

الطبقة الثالثة Layer 3

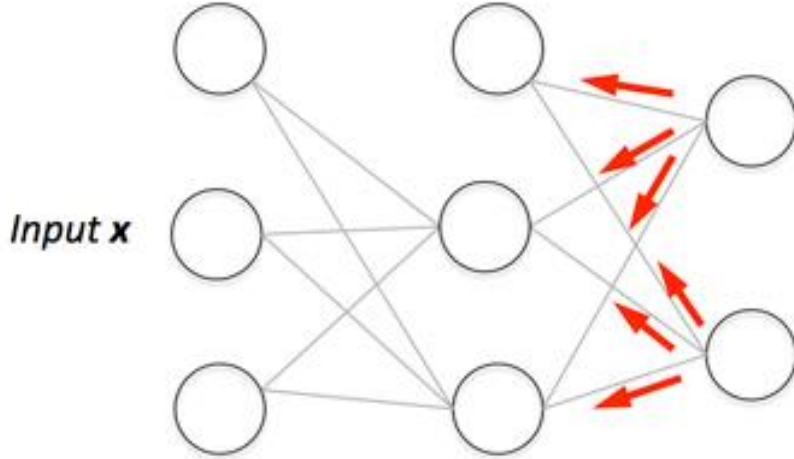
الطبقة الثانية Layer 2

الطبقة الأولى Layer 1



عملية الانتشار العكسي

- نبدأ عند طبقة الإخراج.
- نحسب مشتقات دالة الخسارة بالنسبة للأوزان.
- يتم تحديث الأوزان.
- ننتقل عكسياً للخلف إلى الطبقات المخفية.





تحديث الوزن والتحييز

Weight & Bias Update

تحديث الوزن

بعد احتساب المشتقات ، نحتسب الأوزان باستخدام النزول الاشتقاقي (حيث α : هو معدل التعلم)

$$\begin{aligned} W^1 &:= W^1 - \alpha \frac{\partial J}{\partial W^1} & b^1 &:= b^1 - \alpha \frac{\partial J}{\partial b^1} \\ W^2 &:= W^2 - \alpha \frac{\partial J}{\partial W^2} & b^2 &:= b^2 - \alpha \frac{\partial J}{\partial b^2} \\ W^3 &:= W^3 - \alpha \frac{\partial J}{\partial W^3} & b^3 &:= b^3 - \alpha \frac{\partial J}{\partial b^3} \end{aligned}$$

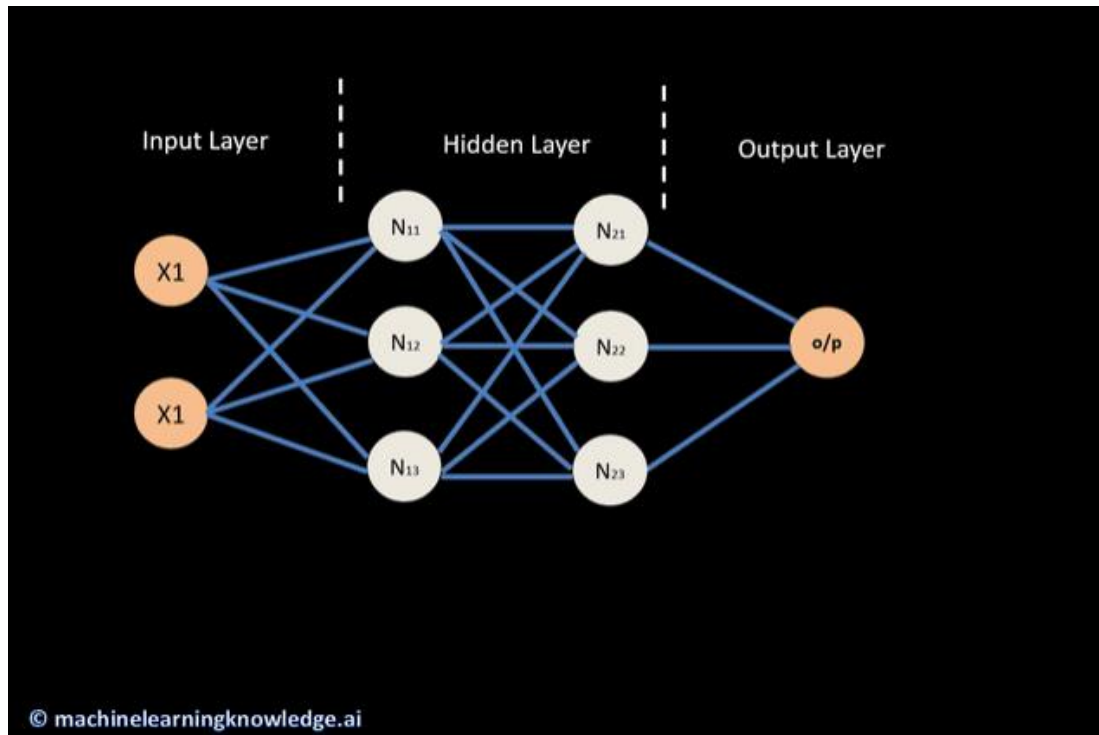
كر العملية مرة أخرى

الآن وبعد أن قمنا باحتساب الأوزان الجديدة ، نكرر عمليتي الانتشار الأمامي -العكسي تبعاً عدة مرات.

● نتوقف في حالتين:

- نصل إلى أقصى عدد من التكرارات.
- نصل إلى القيمة المثلى للأوزان.

الخلاصة



شكراً لكم

Thank you



SDAIA

الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority