

## مقدمة في دوال التنشيط

### لماذا نحتاج إلى دوال التنشيط؟

في الشبكات العصبية لدينا عصبونات ، و كل عصبونه تستقبل مدخلات (inputs) و تجري عليها عمليات جمع موزنه ثم تمررها بدالة التنشيط (activation functions) و بدورها تحولها إلى مخرج ، تساعد دوال التنشيط الشبكة العصبية على تعلم العلاقات والأنماط المعقدة في البيانات. والسؤال الآن هو ماذا لو كنا لا نستخدم أي دالة تنشيط ونسمح للخلايا العصبية بإعطاء المجموع الموزن للمدخلات كما هو كمخرج. في هذه الحالة ، ستكون العملية الحسابية صعبه جدًا نظرًا لأن المجموع الموزون للمدخلات ليس له أي نطاق واعتمادًا على المدخلات يمكن أن يأخذ أي قيمة. ومن ثم ، فإن أحد الاستخدامات المهمة لدوال التنشيط هو إبقاء الناتج مقيدًا بنطاق معين. و استخدام آخر لدالة التنشيط هو إضافة اللاخطية في البيانات. نختار دائمًا دوال غير خطية دوال تنشيط.

### لماذا نختار دائمًا دوال التنشيط الغير الخطي في الشبكة العصبية؟

اللاخطية (Non-linearity) تعني أن المخرجات لا يمكن تولد من مجموعة خطية من المدخلات. اللاخطية مهمة في الشبكات العصبية لأن دوال التنشيط الخطي ليست كافية لتشكيل دوال مقارنة. إذا استخدمنا دوال التنشيط الخطية في شبكة عصبية عميقة ، فبغض النظر عن مدى عمق شبكتنا ، فستكون مساوية لشبكة عصبية خطية بدون طبقات مخفية و ذلك لأنه يمكن دمج دوال التنشيط الخطية لتشكيل دالة خطية واحدة أخرى. لذلك ، سيتم اختزال شبكتنا بالكامل إلى عصبون واحد يحتوي الدالة الخطية المدمجة كدالة تنشيط ، ولن يتمكن العصبون بمفرده من تعلم العلاقات المعقدة في البيانات. نظرًا لأن معظم مشكلات العالم الواقعي معقدة

للغاية ، فنحن بحاجة إلى دوال التنشيط غير الخطية في الشبكة العصبية. الشبكة العصبية بدون دوال التنشيط غير الخطية ستكون مجرد نموذج انحدار خطي بسيط (linear regression model).

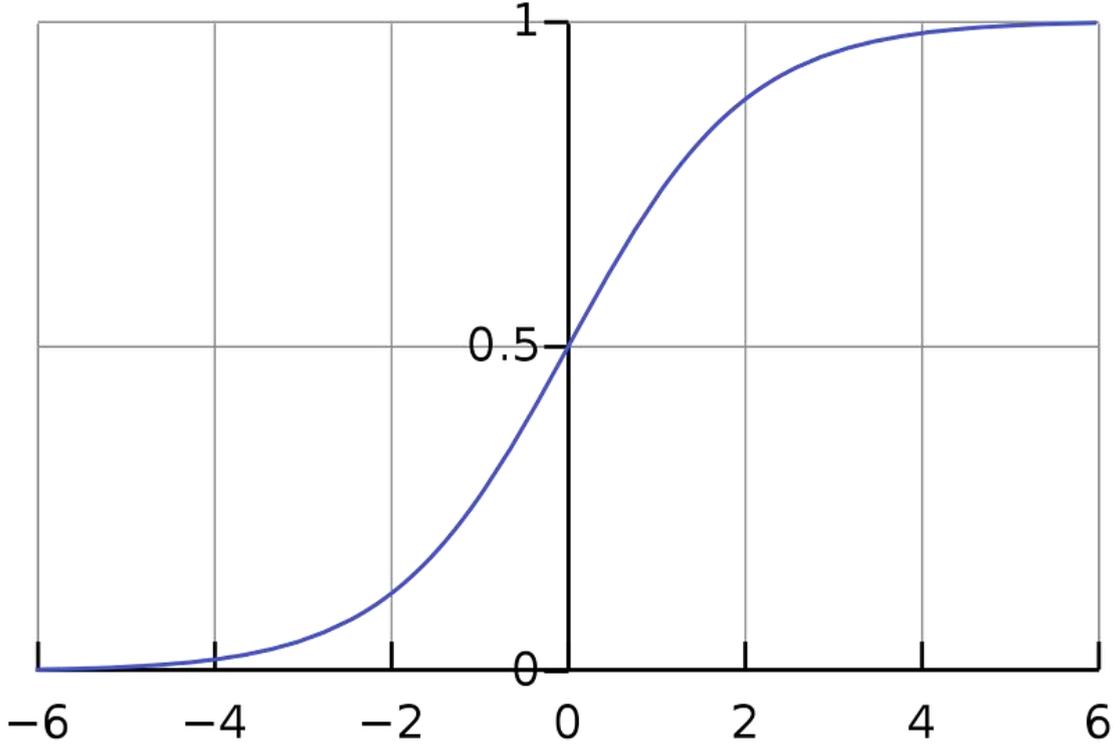
دوال التنشيط

### 1. دالة السيغمويد (Sigmoid)

السيغمويد هي دالة رياضية على شكل حرف "S" لها صيغتها-

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

هنا رسم بياني لدالة السيغمويد. الرغم من أن دالة السيغمويد شائعة جدًا ، إلا أنها لا تستخدم كثيرًا للأسباب التالية-



الايجابيات :

1. دالة السيغمويد مستمرة و قابلة للإشتقاق.
2. المخرجات محدوده بين 0 و 1
3. تعطي تنبؤات واضحة بالنسبة لعمليات التصنيف الثنائي.

السلبيات :

1. يمكن أن تسبب مشكلة تلاشي المشتقات.
2. لا تتمركز حول الصفر.
3. مكلفة حسابيا

مثال:

## 2. دالة سوفت ماكس (softmax):

دالة سوفت ماكس هي تعميم دالة السيغمويد على لتناسب عمليات التصنيف المتعددة . يتم استخدامها بشكل شائع في الطبقة النهائية من عمليات التصنيف المتعددة . إنها تأخذ متجهًا "k" للرقم الحقيقي ثم تسويته (normalizes) في توزيع احتمالي يتكون من "k" من احتمالات تتناسب مع مضاعفات الرقم المُدخل. قبل تطبيق سوفت ماكس ، يمكن أن تكون بعض مكونات المتجه سلبية ، أو أكبر من واحد ، وقد لا تساوي 1 ولكن بعد تطبيق سوفت ماكس سيكون كل مكون في النطاق من 0 إلى 1 حاصل جمعهم يساوي 1 ، وبالتالي يمكن تفسيرها على أنها الاحتمالات.

The standard (unit) softmax function  $\sigma : \mathbb{R}^K \rightarrow \mathbb{R}^K$  is defined by the formula

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \text{ and } \mathbf{z} = (z_1, \dots, z_K) \in \mathbb{R}^K$$

الايجابيات:

يمكن استخدامه عمليات التصنيف المتعددة .وبالتالي يتم استخدامه في طبقة الإخراج للشبكات العصبية.

السلبيات :

إنها مكلفة من الناحية الحسابية حيث يتعين علينا حساب الكثير من المصطلحات الأسية.

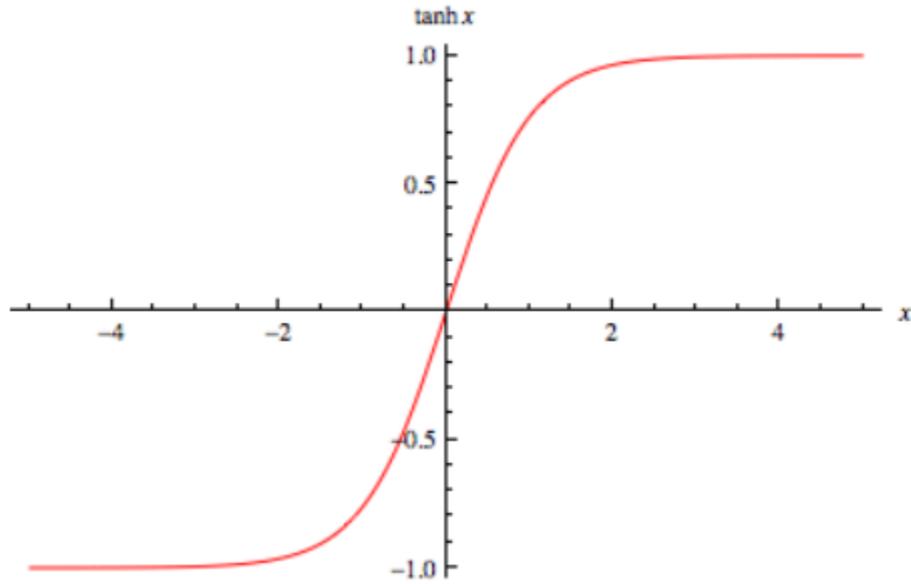
## 3. دالة الظل الزائدي (tanh):

الظل الزائدي أو باختصار "tanh" لديه المعادلة التالية

$$\tanh x = \frac{\sinh x}{\cosh x}$$

$$\tanh x = \frac{e^x - e^{-x}}{2} \div \frac{e^x + e^{-x}}{2} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

وهي تشبه إلى حد كبير دالة السيغمويد. متركزه حول صفر وتتراوح مداها بين -1 و 1+.



الايجابيات:

1. إنها مستمرة و قابلة للإشتقاق في كل مكان.
2. تتمحور حول صفر.
3. سيحد المخرجات في نطاق من -1 إلى 1+.

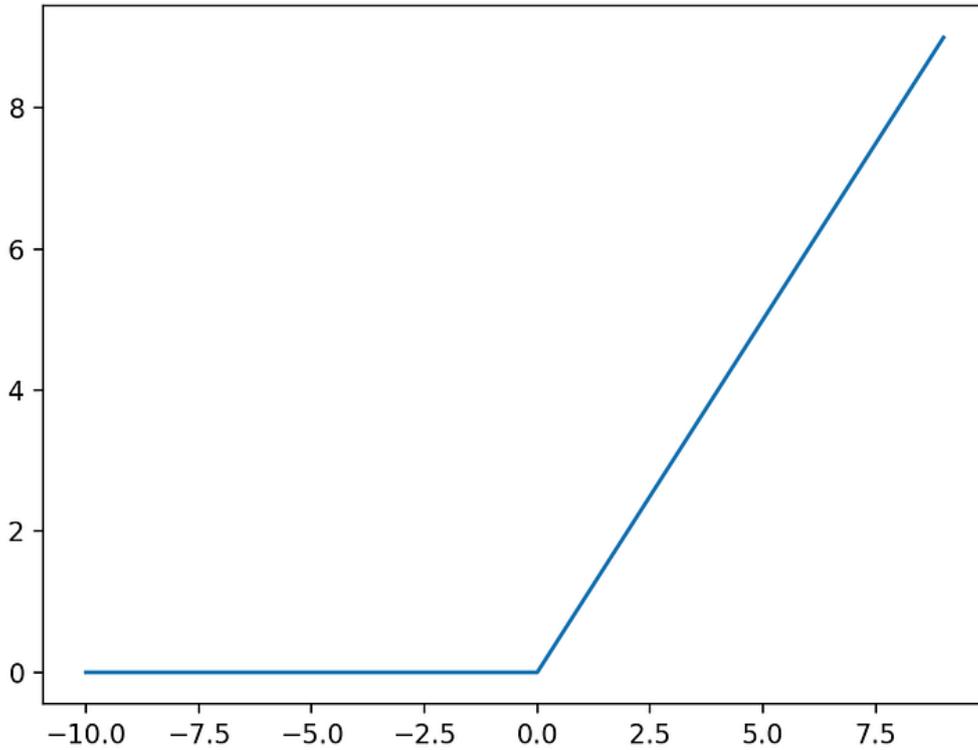
سلبيات:

1. يمكن أن تسبب مشكلة تلاشي المشتقات.
2. مكلفة حسابيا.

#### 4. دالة ريلو (Relu)

الوحدة الخطية المصححة ( rectified linear unit ) غالبًا ما ريلو فقط صورة للنشر

$$f(x) = x^+ = \max(0, x),$$



الايجابيات:

1. سهل الحساب.
2. لا تسبب مشكلة تلاشي المشتقات
3. نظرًا لعدم تنشيط جميع العصبونات ، فإن هذا يخلق تناثر (sparsity) في الشبكة ، وبالتالي تتصف الدالة بالسرعة و الفعالية.
- 4.

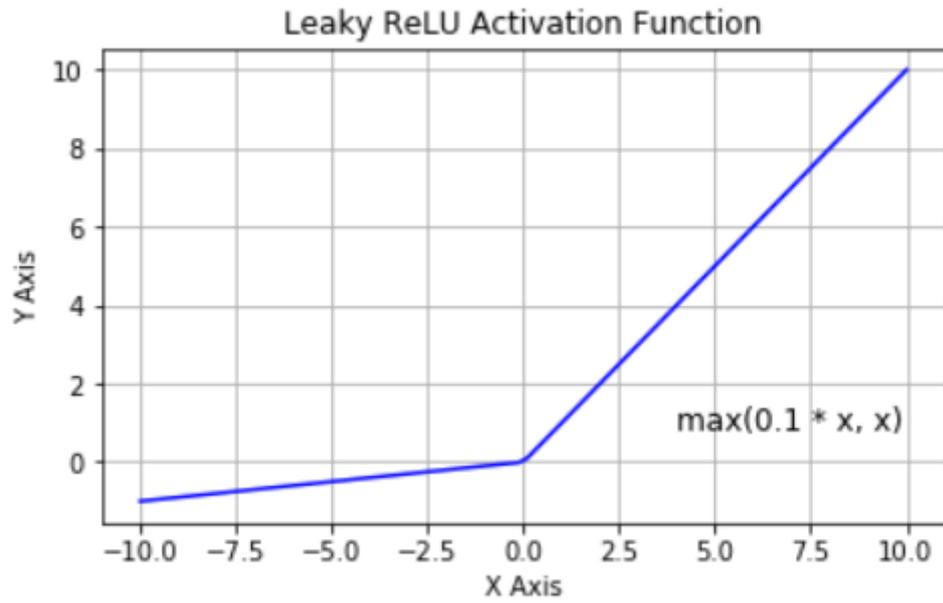
سلبية:

1. تسبب مشكلة إنفجار المشتقات
2. لا تتمركز حول الصفر.
3. يمكن أن تقتل بعض العصبونات إلى الأبد لأنها تعطي دائمًا صفر للقيم السلبية.

## 5. دالة ريلو المُسرب (Leaky Relu)

دالة ريلو المُسرب هو تحسين دالة ريلو. فمن عيوب دالة ريلو هو أنها يمكن أن تقتل بعض العصبونات في كل تكرار ، وهذا ما يعرف باسم حالة موت ريلو. يمكن أن يتغلب التسريب على هذه المشكلة ، بدلاً من إعطاء صفر للقيم السلبية ، ستعطي رقم سالب صغيرًا نسبيًا ، وبالتالي لن تقتل أي عصبونه.

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{otherwise.} \end{cases}$$



الايجابيات:

1. سهولة الحساب.
2. لا يسبب مشكلة تلاشي المشتقات
3. لا يسبب مشكلة موت ريلو

سلبيات:

1. تسبب مشكلة إنفجار المشتقات
2. لا تتمركز حول الصفر.

## 6. دالة ريلو المعيرة (Parameterized Relu)

في ريلو المعيرة ، بدلاً من تحديد معدل ثابت للمحور السلبي ، يتم تمريره كمعيار جديدة قابلة للتدريب تتعلمه الشبكة من تلقاء نفسها لتحقيق تقارب أسرع.

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ ax & \text{otherwise.} \end{cases}$$

الايجابيات:

1. سوف تتعلم الشبكة أنسب قيمة لألفا من تلقاء نفسها.

2. لا يسبب مشكلة تلاشي الإشتقاق

سلبيات:

1. من الصعب حسابها.

2. يعتمد الأداء على المشكلة.

مثال:

في Tensorflow ، يتم تنفيذ ريلة المعيره كطبقة مخصصة. مثال -

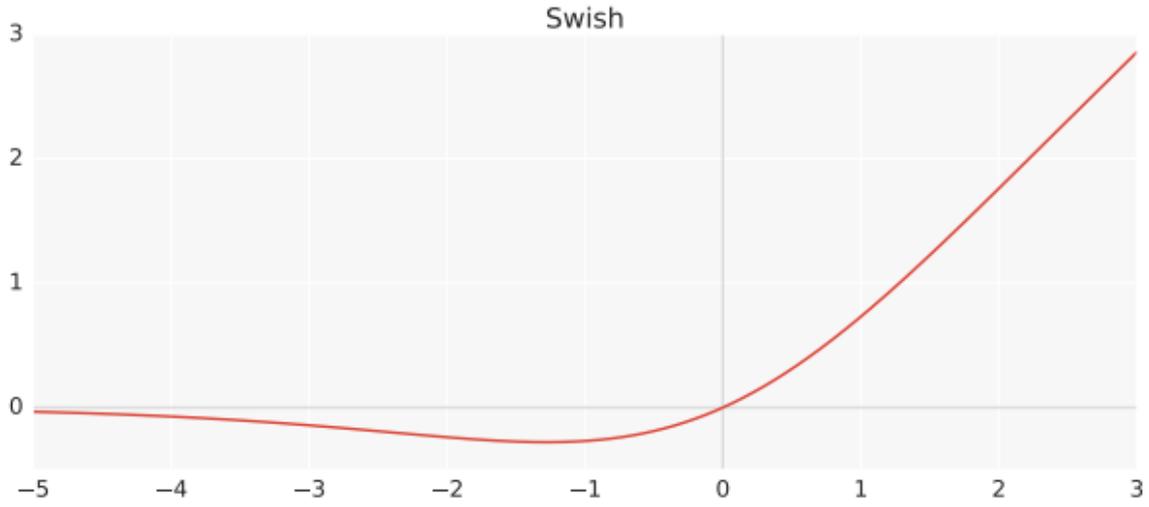
```
keras.layers.advanced_activations.PReLU(weights=None,alpha_initializer="zero")
```

```
<keras.layers.advanced_activations.PReLU at 0x2298517c608>
```

7. دالة سويش (Swish):

يتم الحصول على دالة سويش بضرب  $x$  في دالة السغمويد.

$$\text{swish}(x) := x \times \text{sigmoid}(\beta x) = \frac{x}{1 + e^{-\beta x}}.$$



تم اقتراح دالة سويش من قبل فريق غوغل براين (Google's Brain) . حيث تظهر تجاربهم أن سويش تميل إلى العمل بشكل أسرع من ريلو في الشبكات العميقة

الايجابيات:

1. لا تسبب مشكلة موت ريلو.
2. أثبت أنها أفضل من ريلو بشيء بسيط.

سلبيات:

1. مكلفة حسابيا.

**8. دالة إيلو (ELU)**

الوحدة الخطية الأسية (Exponential Linear Unit) هي شكل آخر من أشكال ريلو و التي تحاول جعل التنشيط أقرب إلى الصفر مما يسرع التعلم. وقد أظهرت دقة تصنيف أفضل من ريلو. تحتوي إيلو على قيم سلبية تدفع متوسط عمليات التنشيط إلى ما يقرب من الصفر.

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ a(e^x - 1) & \text{otherwise,} \end{cases}$$

الايجابيات:

1. لا تسبب مشكلة موت ريلو.

سلبيات:

1. مكلفة حسابيا.

2. تسبب مشكلة انفجار المشتقة

3. يجب تحديد قيمة ألفا.

```
tf.keras.layers.ELU(alpha=1.0)
```

```
<tensorflow.python.keras.layers.advanced_activations.ELU at 0x22985181f08>
```

## 9. سوفت بلس و سوفت ساين (Softplus and Softsign)

دالة سوفت بلس هي -

$$f(x) = \ln(1 + e^x),$$

مشتقتها هي دالة السيغمويد .

أما سوفت ساين هي -

$$\text{softsign}(x) = x / (\text{abs}(x) + 1)$$

لا يتم استخدام سوفت بلس و سوفت ساين كثيرا وبشكل عام ، يُفضل ريلوومتغيراته عليها.

الايجابيات:

1. لا تسبب مشكلة تلاشي الإشتقاق

سلبيات:

1. مكلفة حسابيا.

2. أبطأ من ريلو.