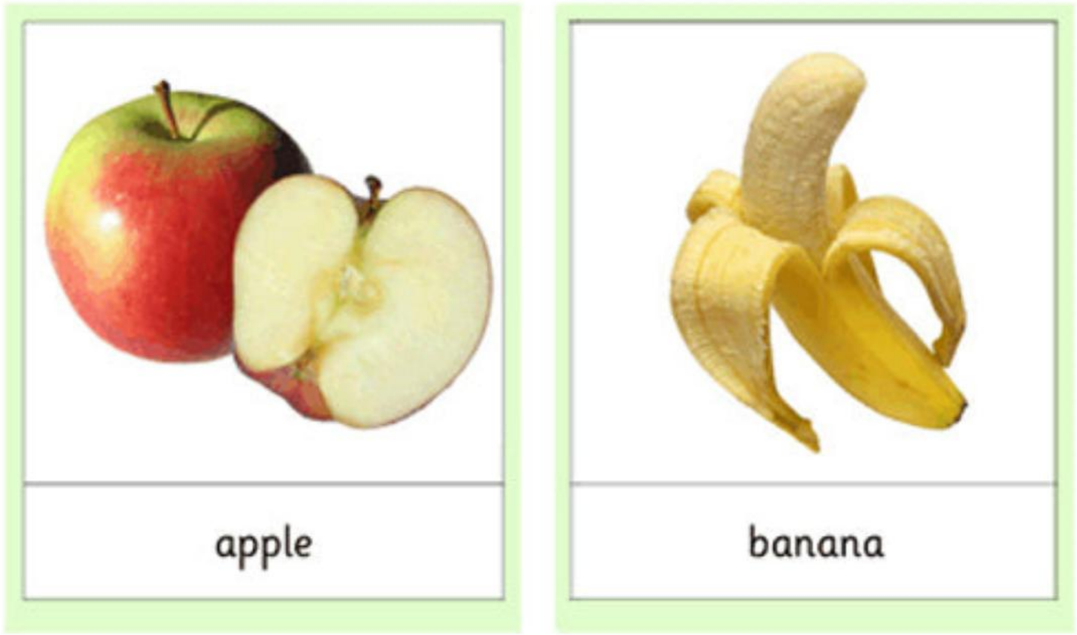
Multi-Class Classification

Lecture 7 - Part 1

شرح المفاهيم الأساسية وتطبيقاتها في الـ Neural Networks

ما هي Multi-Class Classification؟



- عندما يكون لدينا أكثر من **Class** واحد ممكن للتصنيف.
- أمثلة عملية:
  - تصنيف الفاكهة:
    {apple, banana, orange}
  - تصنيف الملابس: **{t-shirt, jeans, dress}**
- في الوضع الطبيعي، كل عينة تنتمي لـ **Class واحد فقط**، وهذا ما يسمى بـ **Single-Label**.

الفرق بين Hard Label و Soft Label

Hard Label

يتم تمثيله كـ **One-Hot Vector**.

```
y = [0, 1, 0]
```

الرقم 1 للكلاس الصحيح والباقي أصفار. يُستخدم غالباً في **Supervised Learning** التقليدي.

Soft Label

يتم تمثيله كـ **Distribution** (توزيع احتمالات).

```
y = [0.1, 0.7, 0.2]
```

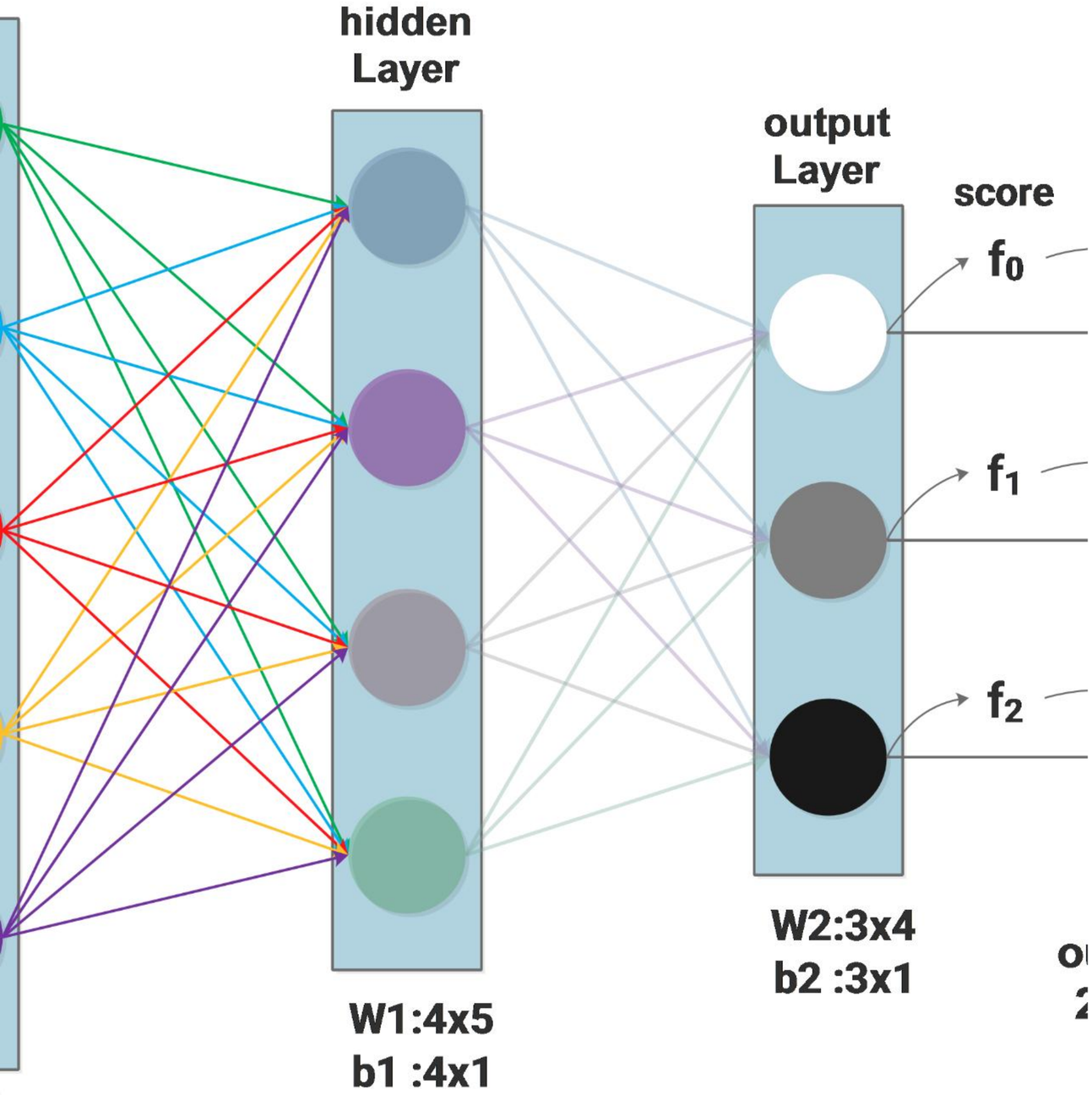
يعطي درجة ثقة لكل كلاس. مفيد في **Label Smoothing** و **Knowledge Distillation**.

دالة Softmax

تقوم بتحويل الـ **Vector of Logits (z)** إلى احتمالات حقيقية.

$$p_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}$$

- المخرجات دائمًا موجبة.
- مجموع الاحتمالات يساوي 1.
- الدالة الأسية **exponential** تضخم الفروقات بين القيم.



دالة الخسارة Categorical Cross-Entropy

الشكل العام

تستخدم مع **Softmax** لقياس الخطأ:

$$L = -\sum_{k=1}^{K} y_k \cdot \log(p_k)$$

حيث **y** هو التوزيع الحقيقي و **p** هو الاحتمال المتوقع.

في حالة Hard Label

إذا كان الكلاس الصحيح هو **c**، فإن المعادلة تتبسط إلى:

$$L = -\log(p_c)$$

- إذا كان النموذج متأكدًا ($p \approx 1$) ← الخسارة ≈ 0.
- إذا كان مخطئًا ($p \approx 0$) ← الخسارة كبيرة جدًا.

طرق حساب الـ Error

Evaluation Metric

Accuracy

نسبة التوقعات الصحيحة من إجمالي العينات. تُستخدم لتقييم أداء النموذج نهائياً.

Error Rate = 1 - Accuracy

Training Loss

Cross-Entropy Loss

تُستخدم لتدريب النموذج وحساب الـ Gradients لتحديث الأوزان (Weights).

مقارنة: Logits vs Softmax Probabilities

توضيح كيف تقوم Softmax بتحويل القيم (Logits) إلى احتمالات مجموعها 1.

لاحظ أن مجموع الاحتمالات (الجانب الأيسر) يساوي 1 (100%)

Input a set of training examples

For each training example x**:** Set the corresponding input activation $a^{x,1}$, and perform the following steps:

- **Feedforward:** For each $l = 2, 3, \ldots, L$ compute $z^{x,l} = w^l a^{x,l-1} + b^l$ and $a^{x,l} = \sigma(z^{x,l})$.
- **Output error** $\delta^{x,L}$**:** Compute the vector $\delta^{x,L} = \nabla_a C_x \odot \sigma'(z^{x,L})$.
- **Backpropagate the error:** For each $l = L-1, L-2, \ldots, 2$ compute $\delta^{x,l} = ((w^{l+1})^T \delta^{x,l+1}) \odot \sigma'(z^{x,l})$.

Gradient descent: For each $l = L, L-1, \ldots, 2$ update th weights according to the rule $w^l \to w^l - \frac{\eta}{m} \sum_x \delta^{x,l}(a^{x,l-1})^T$, the biases according to the rule $b^l \to b^l - \frac{\eta}{m} \sum_x \delta^{x,l}$.

تحديث الأوزان
Weight Update

الـ Gradient بالنسبة للـ Logits هو الفرق بين التوقع والحقيقة:

$$\frac{\partial L}{\partial z} = p - y$$

- إذا كان **p > y**: النموذج بالغ في الاحتمال ← نقلل الوزن.
- إذا كان **p < y**: النموذج قلل الاحتمال ← نزيد الوزن.

يتم استخدام هذا المبدأ في **Gradient Descent**

تطبيقات لمهندس الذكاء الاصطناعي

- **Image Classification:** تصنيف الصور الطبية، المركبات، إلخ.
- **NLP Intent Detection:** فهم نية المستخدم في الشات بوت.
- **Knowledge Distillation:** استخدام الـ Soft Labels لتدريب نماذج أصغر وأسرع.

أسئلة؟

شكراً لاستماعكم

Lecture 7 - Part 1: Multi-Class Classification

Image Sources

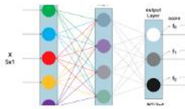
http://absorbentminds.co.uk/cdn/shop/products/xl_15_121.jpg?v=1667820307

Source: absorbentminds.co.uk

---

https://www.wangxinliu.com/images/machine_learning/2layer_nn.png

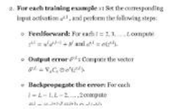
Source: www.wangxinliu.com



---

https://i.sstatic.net/goLCZ.png

Source: cs.stackexchange.com



-https://cdn-media-1.freecodecamp.org/images/h5iAZ9X9kElc-pY9HEnd8jrPjSpDWoJQJeK

Source: www.freecodecamp.org