

سوال 1) اثبات فرمول Log Loss در مسائل Multi-Class و استفاده از تابع SoftMax

انترپوی کراس (Cross-Entropy) بین دو توزیع احتمال  $p$  و  $q$  به صورت زیر تعریف می شود:

$$H(p, q) = - \sum (p(x) * \log(q(x)))$$

فرض کنید یک نمونه با  $K$  کلاس داریم. برای این نمونه، مدل ما یک بردار احتمال  $p = (p_1, p_2, \dots, p_K)$  تولید می کند که در آن  $p_i$  احتمال تعلق نمونه به کلاس  $i$  است. همچنین، برچسب واقعی نمونه را با یک بردار یک کدی  $y$  نشان می دهیم که در آن عنصر مربوط به کلاس صحیح برابر با 1 و بقیه عناصر برابر با 0 هستند. در این صورت، Log Loss به صورت زیر محاسبه می شود:

$$L = - \sum (y_i * \log(p_i))$$

$$\begin{aligned} \frac{\partial L}{\partial l_n} &= - \sum_i \frac{y_i}{\hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial l_n} = \\ &= - \sum_i \frac{y_i}{\hat{y}_i} \cdot \left[ \hat{y}_n(1 - \hat{y}_n) \right]_{i=n} - \left[ \hat{y}_i \hat{y}_n \right]_{i \neq n} = \\ &\quad \text{Let's separate the two cases of } i = n \text{ and } i \neq n. \\ &= - \sum_{i=n} \frac{y_i}{\hat{y}_i} \cdot \hat{y}_n(1 - \hat{y}_n) + \sum_{i \neq n} \frac{y_i}{\hat{y}_i} \cdot \hat{y}_i \hat{y}_n = \\ &\quad \text{In these green boxes, } i = n. \text{ As a consequence, } \hat{y}_i \text{ and } \hat{y}_n \text{ are the same, and so are } y_i \text{ and } y_n. \text{ we can use them interchangeably.} \\ &= - y_n + y_n \hat{y}_n + \sum_{i \neq n} y_i \hat{y}_n = \\ &= - y_n + \sum_i y_i \hat{y}_n = \quad \text{Let's put the two cases of } i = n \text{ and } i \neq n \text{ back together.} \\ &= - y_n + 1 \cdot \hat{y}_n = \quad \text{one final twist: } y \text{ is a one hot encoded vector. So all its components are zero, except for one of them that is 1. That means that the sum in the previous step boils down to 1 by } \hat{y}_n. \\ &= \hat{y}_n - y_n \\ &\quad \text{There we go!} \end{aligned}$$

سوال 2) تشخیص Convex بودن تابع با استفاده از مشتق دوم

فرض کنید تابع  $f(x) = x^2$  را داریم. مشتق اول این تابع  $f'(x) = 2x$  و مشتق دوم آن  $f''(x) = 2$  است. از آنجایی که مشتق دوم همواره مثبت است، این تابع محدب است.

فرض کنید تابع  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  دو بار پیوسته مشتق پذیر باشد. این تابع در دامنه خود محدب است اگر و تنها اگر ماتریس هسیان (ماتریس مشتقات جزئی مرتبه دوم) آن در هر نقطه از دامنه، نیمه‌مثبت معین باشد.

- **ماتریس هسیان:** ماتریسی است که درایه‌های آن مشتقات جزئی مرتبه دوم تابع هستند.
  - **نیمه‌مثبت معین:** یک ماتریس  $A$  نیمه‌مثبت معین نامیده می‌شود اگر برای هر بردار غیر صفر  $x$ ، حاصلضرب  $x^T A x$  بزرگتر یا مساوی صفر باشد.
- در حالی که مثبت بودن مشتق دوم در تمام نقاط دامنه یک شرط **کافی** برای محدب بودن تابع است، اما شرط **ضروری** نیست. به عبارت دیگر، اگر تابعی محدب باشد، لزوماً مشتق دوم آن در همه نقاط مثبت نیست.
- توابع غیر مشتق پذیر:** بسیاری از توابع محدب وجود دارند که در برخی نقاط مشتق پذیر نیستند. برای مثال، تابع مقدار مطلق  $|x|$  در نقطه  $x=0$  مشتق پذیر نیست، اما در کل یک تابع محدب است.
- محدب بودن در نقاط عطف:** در نقاط عطف یک تابع، مشتق دوم صفر می‌شود. این بدان معناست که در این نقاط، شرط مثبت بودن مشتق دوم نقض می‌شود، اما تابع ممکن است همچنان در اطراف این نقاط محدب باشد.