**نام مدرس:** دکتر محمدی

**دستیاران آموزشی مرتبط:** خانم محمودی-خانم انوری

**مهلت تحویل:** ۱۴۰۱/۱۲/۱۳

**تمرین سری اول**

**درس یادگیری عمیق**

۱- فرض کنید خروجی یک لایه خطی (linear layer) بردار output و خروجی مطلوب برای یک نمونه بردار y می‌باشد.(۲۰ نمره)

Output = [4, 6, -10, -20]

Y = [1, 0, 0, 0]

أ. اگر از تابع فعال‌ساز Softmax و تابع ضرر Cross Entropy استفاده کنیم، مقدار تابع ضرر برای این نمونه و گرایان آن نسبت به output را محاسبه کنید. مراحل محاسبات را به طور کامل یادداشت کنید.

$$Softmax\ (x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}$$

$$Cross\ Entropy\ (y_i, \hat{y}) = -\sum_{i=1}^{n} y_i\ \log \hat{y}$$

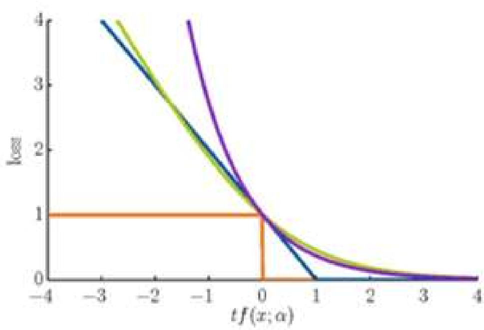
ب. اگر تابع فعال‌ساز sigmoid و تابع ضرر Binary Cross Entropy باشد، بار دیگر مقدار تابع ضرر برای این نمونه و گرادیان آن نسبت به output را به دست بیاورید.

$$Sigmoid\ (x_i) = \frac{1}{1 +\ exp^{-x_i}}$$

$$Binary\ Cross\ Entropy\ (y_i, \hat{y}) = -\sum_{i=1}^{n} y_i\ \log \hat{y}_i + (1 - y_i)\ \log(1 - \hat{y}_i)$$

برای محاسبه مقادیر عددی می‌توانید از کد پایتون استفاده کنید اما مراحل را به طور کامل بنویسید.

۲- از بین نمودار های رسم شده مشخص کنید کدام نمودار مربوط به تابع ضرر hinge و کدام مربوط به تابع ضرر Cross-entropy می‌باشد. استدلال خود را همراه با اثبات فرمول ریاضی بیان کنید (دو نمودار اضافی هستند). (۱۰ نمره)



سید محمد عرفان سلوکی منزه
۹۰۱۷۲۲۱۹۹

$$\text{Output} = O = [4, 6, -10, -20]$$
$$y = [1, 0, 0, 0]$$

$$\hat{y} = \text{softmax}(O) \quad , \quad \text{softmax}(O_i) = \frac{e^{O_i}}{\sum_{j=1}^{n} e^{O_j}} \quad , \quad n = 4$$

$$\square = e^{4} + e^{6} + e^{-10} + e^{-20}$$

الف

$$\left.\begin{aligned} \hat{y}_1 &= \frac{e^{4}}{\square} \\ \hat{y}_2 &= \frac{e^{6}}{\square} \\ \hat{y}_3 &= \frac{e^{-10}}{\square} \\ \hat{y}_4 &= \frac{e^{-20}}{\square} \end{aligned}\right\} \longrightarrow \hat{y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4]$$

$$\simeq [0.1192, 0.8807, 9.9 \times 10^{-8}, 4.5 \times 10^{-12}]$$

$$CE(y, \hat{y}) = -\sum_{i=1}^{n} y_i \log \hat{y}_i$$

$$CE(y, \hat{y}) = -[1 \times \log 0.1192 + 0 + 0 + 0]$$

$$= 2.1269$$

---

$$\frac{\partial}{\partial O_i} CE(y, \hat{y}) = \frac{\partial}{\partial O_i} - \sum_{i=1}^{n} y_i \log \frac{e^{O_i}}{\sum_{j=1}^{n} e^{O_j}}$$

$$= [\sum_{i=1}^{n} y_i [\underbrace{\log e^{O_i}}_{O_i} - \log \sum_{j=1}^{n} e^{O_j}]]$$

$$-\left[\sum_{i=1}^{n} y_i o_i - \sum_{i=1}^{n} y_i \log \sum_{j=1}^{n} e^{o_j}\right]$$

$$\frac{\partial}{\partial o_i}\left[\sum_{i=1}^{n} y_i \log \sum_{j=1}^{n} e^{o_j} - \sum_{i=1}^{n} y_i o_i\right]$$

$$= y_i \frac{e^{o_i}}{\sum_{j=1}^{n} e^{o_j}} - y_i = y_i\,(\mathrm{softmax}(o_i) - 1)$$

$$= y_i\,(\hat{y}_i - 1)$$

$$\text{sigmoid}(o_i) = \frac{1}{1+e^{-o_i}} = \hat{y}_i$$

$$bce(y,\hat{y}) = -\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)$$

$$\hat{y}_1 = \frac{1}{1+e^{-4}}$$

$$\hat{y}_2 = \frac{1}{1+e^{-6}}$$

$$\hat{y}_3 = \frac{1}{1+e^{10}}$$

$$\hat{y}_4 = \frac{1}{1+e^{20}}$$

$$\hat{y} \simeq [0.982,\ 0.9975,\ 4.5\times10^{-5},\ 2\times10^{-9}]$$

$$bce(y,\hat{y}) = [\log 0.982 + 0] + [0 + \log(1-0.9975)] + [0 + \log(1-4.5\times10^{-5})] + [0 + \log(1-2\times10^{-9})]$$

$$\simeq 6.02$$

$$bce(y,\hat{y}) = -\sum_{i=1}^{n} y_i \ln \sigma(o_i) + (1-y_i)\ln(1-\sigma(o_i))$$

$$\frac{\partial\, bce(y,\hat{y})}{\partial o_i} = -\left[\frac{y_i\,\sigma(o_i)(1-\sigma(o_i))}{\sigma(o_i)} + \frac{-(1-y_i)\,\sigma(o_i)\,(1-\sigma(o_i))}{1-\sigma(o_i)}\right]$$

$$= -\left(y_i(1-\sigma(o_i)) + (y_i-1)\times\sigma(o_i)\right) = -y_i + \sigma(o_i)$$

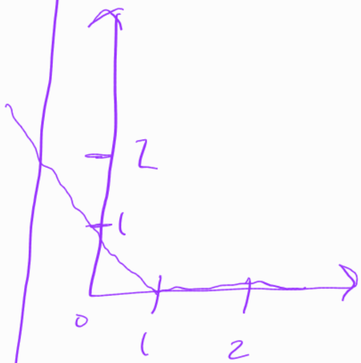
با دادن $\Rightarrow$ $\simeq [-1.72\times10^{-2},\ 9.97\times10^{-1},\ 4.5\times10^{-5},\ 2\times10^{-9}]$

hinge loss → $\max(0, 1 - t \cdot h(x))$

در سمت +g مشخصات نمودار خطی با شیب منفی → عرض از مبدا = 1

در سمت -g " " " " شیب صفر و عرض از مبدا صفر

نمودار آبی رنگ




$\max(0, 1 - t \cdot 0) = 1$ → عرض از مبدا

$\forall x, x \geq 1 \quad hinge(x) = 0$

---

cross-entropy loss → $-\sum_{i=1}^{n} y_i \log(\hat{y}_i)$

only apply loss for $y_i = 1$ → $ce \to -\log(\hat{y}_i)$

$\hat{y}_i = \sigma(h(x)) \longrightarrow ce \to -\log\frac{1}{1+e^{-h(x)}} = \log(1+e^{-f(x)})$

کافی است مقدار 3 تابع آبی، سبز و بنفش را در $x=-1$ حساب کنیم.

در $x=-1$ → آبی > بنفش و آبی > سبز

$hinge(-1) = 2$

$ce(-1) = 1.89$ → نمودار آبی

۳- در این سوال قرار است یک لایه خطی با زبان Pytorch طراحی کنیم. به نوت بوک داده شده مراجعه کرده و بخش‌های مشخص شده برای سوال سوم را تکمیل کنید. (۲۵ نمره)

أ. نمودار تغییرات دقت و تابع ضرر برای فرآیند training را رسم کرده و تحلیل کنید.

ب. به ازای هر کدام از کلاس‌ها (کلاس صفر تا ۹) وزن‌های مربوط به آن کلاس را ترسیم کرده (به صورت یک تصویر) و تحلیل خود را از آن بیان کنید.

۴- یک استارتاپ تاکسی اینترنتی، برای تخمین زمان رسیدن تاکسی به مقصد، ۴ متغیر مسافت، سرعت متوسط، میزان ترافیک و شرایط آب و هوایی را لحاظ کرده و از فرمول زیر استفاده می‌کند: (۲۷ نمره)

$$Time\_of\_arrival = \frac{Distance}{speed} * w0 + traffic * w1 + weather\_conditions * w2 + b + \varepsilon$$

أ) فرض کنید نویز از توزیع گوسی به صورت مقابل پیروی می‌کند: $\varepsilon \sim N(\mu = 0, \sigma^2)$

بهترین میزان را برای پارامتر های w0, b از طریق Maximum Likelihood تخمین بزنید. (نوشتن تمامی مراحل بر روی کاغذ برای همه‌ی پارمترها الزامیست)

ب) در نوت‌بوک پیوست شده، داده‌های EAT.csv را یک‌بار با تابع آماده Linear Regression و یک‌بار نیز با نوشتن کد رگرسیون از پایه، آموزش دهید (با فرمول بهینه ساز GD) و نتایج را از طریق توابع ضرر MSE ,MAE مقایسه کنید.

۵. در این بخش می‌خواهیم وزن‌های آموزش‌دیده برای ویژگی‌های مختلف و میزان وابستگی آنها به مجموعه داده را بررسی کنیم. برای این منظور، ۱۰۰ آزمایش انجام دهید که در هر کدام یک زیربخش ۳۰۰تایی به صورت تصادفی از داده‌ها را تهیه کنید و پارامترهای مدل قسمت قبل را با آنها آموزش بدهید. سپس، میانگین و انحراف معیار هر کدام از پارامترهای مدل در این ۱۰۰ آزمایش را ترسیم کنید و نتیجه حاصل را تحلیل کنید. بر اساس آمار سازمان بهداشت جهانی، سالانه ۱۲ میلیون مرگ در سراسر جهان به دلیل بیماری‌های قلبی رخ می‌دهد. برای پیش‌بینی خطر ابتلا به بیماری‌های قلبی، مجموعه داده‌ای از خطرناک‌ترین عوامل بسترسازی این بیماری جمع‌آوری شده است(heart_disease.csv). در این مجموعه داده، ستون TenYearCHD نشان می‌دهد که آیا خطر ابتلا به بیماری، فرد را در ۱۰ سال آینده تهدید می‌کند یا خیر. (۱۸ نمره)

أ. فرض کنید در داده فوق از هر ۴۲۳۸ نفر ، ۶۴۴ نفر در خطر ابتلا به بیماری قلبی هستند. حال، اگر یک فرد به صورت تصادفی انتخاب شود، چقدر احتمال دارد به بیماری قلبی مبتلا شود؟

$\frac{644}{4238} = 0.152 \rightarrow 15.2\%$

(4)

$$\text{time of arrival} = \frac{\text{dis}}{\text{speed}} w_0 + \text{traffic } w_1 + \text{cond } w_2 + b + \varepsilon$$

$$= f_0 w_0 + f_1 w_1 + f_2 w_2 + b + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

پس $\Rightarrow$ $P(y \mid f_0, f_1, f_2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - f_0 w_0 - f_1 w_1 - f_2 w_2 - b)^2}{2\sigma^2}}$

$$P(y \mid X) = \prod_{i=1}^{n} P(y^i \mid f_0^i, f_1^i, f_2^i) \qquad X = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \end{bmatrix}$$

$\to$ max this or min $-\log P(y \mid X)$

$$-\log P(y \mid X) = -\sum_{i=1}^{n} \log \left[ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\sim\sim^2}{2\sigma^2}} \right]$$

$$\longrightarrow = + \sum_{i=1}^{n} \underbrace{+\frac{1}{2}\log(2\pi\sigma^2)}_{\text{ترم ثابت}} + \underbrace{\frac{\sim\sim^2}{2\sigma^2}}_{\text{ترم MSE}}$$

$$\underbrace{w^*, b^*}_{w} = \underset{w, b}{\text{argmin}} \sum_{i=1}^{n} \frac{(y - wx - b)^2}{2\sigma^2} \rightarrow w^* = \underset{w}{\text{argmin}} \| y - Xw \|^2$$

$\to w^* = [b, w^*]$, $X = \begin{bmatrix} 1 \\ f_0 \\ f_1 \\ f_2 \end{bmatrix}$

$$\nabla_w \| y - Xw \|^2 = -2X^T (y - Xw) \Rightarrow w^* = (X^T X)^{-1} X^T y$$

$$= X^T y$$

| | | |
|---|---|---|
| **نام مدرس:** دکتر محمدی<br>**دستیاران آموزشی مرتبط:** خانم محمودی-خانم انوری<br>**مهلت تحویل:** ۱۴۰۱/۱۲/۱۳ | **تمرین سری اول**<br>**درس یادگیری عمیق** | دانشگاه علم و صنعت ایران |

ب. فرض کنید هر یک از افراد در داده فوق، یک نمونه تصادفی از توزیع برنولی با پارامتر p باشند. تخمین p با استفاده از رویکرد Maximum Likelihood را برای این مجموعه داده محاسبه کنید. (نوشتن تمامی مراحل بر روی کاغذ الزامیست)

ت. با توجه به اطلاعات فوق، یک مدل Logistic Regression روی داده‌ها بر حسب ستون TenYearCHD آموزش دهید و سپس پیش‌بینی آن را ارزیابی نمایید. همچنین تفاوت Linear Regression و Logistic Regression را بیان نمایید.

**لطفا سند قوانین انجام و تحویل تمرین های درس را مطالعه و موارد خواسته شده را رعایت فرمایید**

موفق و سلامت باشید

توزیع برنولی ← $p^y(1-p)^{1-y}$

$$p(y|x) = \prod_{i=1}^{n} p(y^i|x^i)$$

↳ max this or $\cancel{max}$ min $-\log p(y|x)$

$$= -\sum_{i=1}^{n} y_i \log p + (1-y_i)\log(1-p)$$

$$= -\log p \sum_{i=1}^{n} y_i + \log(1-p)\sum_{i=1}^{n}(1-y_i)$$

$$= -n\bar{y}\log p - n(1-\bar{y})\log(1-p)$$

$$\frac{\partial -\log P(y|x)}{\partial p} = -\frac{n\bar{y}}{p} - \frac{-n(1-\bar{y})}{1-p} = 0$$

$$\rightarrow \frac{n\bar{y}}{p} = \frac{n(1-\bar{y})}{1-p} \rightarrow p = \bar{y}$$