سوال ۴:

الف) برای پیدا کردن پارامتر w، از تابع هزینه مجموع مجذور خطا استفاده می‌کنیم، که به صورت زیر تعریف می‌شود:

$$L(w) = \sum_{i=1}^{n} (y_i - \exp(wx_i))^2$$

در این رابطه:

- y_i مقدار واقعی خروجی برای نمونه i–ام در مجموعه داده است.
- $\exp(wx_i)$ مقدار پیش‌بینی شده توسط مدل برای ورودی x_i است.
- اختلاف بین y_i و $\exp(wx_i)$ بیانگر خطا برای هر نمونه است، و مجذور آن، خطای مربع شده را به ما می‌دهد.

بنابراین، تابع هزینه $L(w)$ که باید کمینه شود، مجموع خطاهای مربع شده برای تمام نمونه‌های مجموعه داده است.

ب) برای بهینه‌سازی پارامتر w با استفاده از روش کاهش گرادیان، نیاز داریم تابع هزینه $L(w)$ را نسبت به w مشتق گرفته و از گرادیان به‌دست‌آمده برای به‌روزرسانی w استفاده کنیم.

تابع هزینه $L(w)$ به صورت زیر تعریف شده است:

$$L(w) = \sum_{i=1}^{n} (y_i - \exp(wx_i))^2$$

گام اول: محاسبه گرادیان تابع هزینه نسبت به w

برای استفاده از کاهش گرادیان، ابتدا گرادیان تابع هزینه $L(w)$ را نسبت به w محاسبه می‌کنیم:

$$\frac{\partial L(w)}{\partial w} = \sum_{i=1}^{n} 2\,(y_i - \exp(wx_i)) \cdot (-x_i \exp(wx_i))$$

که با ساده‌سازی به شکل زیر می‌رسد:

$$\frac{\partial L(w)}{\partial w} = -2 \sum_{i=1}^{n} (y_i - \exp(wx_i)) \cdot x_i \exp(wx_i)$$

گام دوم: به‌روزرسانی w با استفاده از کاهش گرادیان

در هر گام از الگوریتم کاهش گرادیان، پارامتر w را با استفاده از رابطه زیر به‌روزرسانی می‌کنیم:

$$w_{t+1} = w_t - \alpha \frac{\partial L(w)}{\partial w}\bigg|_{w=w_t}$$

که در آن:

- w_t مقدار w در تکرار tام است.
- α نرخ یادگیری (گام بهینه‌سازی) است که مقدار آن باید به اندازه کافی کوچک انتخاب شود تا الگوریتم به درستی همگرا شود.
- $\frac{\partial L(w)}{\partial w}\big|_{w=w_t}$ مقدار گرادیان تابع هزینه نسبت به w در نقطه $w = w_t$ است.

بنابراین، رابطه به‌روزرسانی w به صورت زیر خواهد بود:

$$w_{t+1} = w_t + 2\alpha \sum_{i=1}^{n} (y_i - \exp(w_t x_i)) \cdot x_i \exp(w_t x_i)$$

این رابطه به ما اجازه می‌دهد تا با استفاده از مقدار فعلی w و نرخ یادگیری α، مقدار w را در هر گام کاهش گرادیان به‌روزرسانی کنیم تا به مقدار بهینه همگرا شود.

ج) برای یافتن رابطه‌ای که مقدار بهینه پارامتر w باید در آن صدق کند، لازم است مشتق تابع هزینه $L(w)$ را نسبت به w صفر کنیم.

برای کمینه‌سازی تابع هزینه، گرادیان آن نسبت به w را محاسبه کرده و برابر صفر قرار می‌دهیم:

$$\frac{\partial L(w)}{\partial w} = -2 \sum_{i=1}^{n} (y_i - \exp(w x_i)) \cdot x_i \exp(w x_i) = 0$$

این رابطه را می‌توان به‌صورت زیر بازنویسی کرد:

$$\sum_{i=1}^{n} (y_i - \exp(w x_i))\, x_i \exp(w x_i) = 0$$

ساده‌سازی بیشتر این رابطه:

$$\sum_{i=1}^{n} y_i x_i \exp(w x_i) = \sum_{i=1}^{n} x_i \exp(2 w x_i)$$

این معادله دقیقاً با گزینه ج مطابقت دارد:

$$\sum_{i=1}^{n} x_i \exp(2 w x_i) = \sum_{i=1}^{n} x_i y_i \exp(w x_i)$$

بنابراین، مقدار بهینه w باید در رابطه ج صدق کند.