سوال ۲:

ب) برای حل این مسئله و یافتن مقدار بهینه w، تابع هزینه داده‌شده را تحلیل می‌کنیم. تابع هزینه به صورت زیر است:

$$L(w) = \sum_{i=1}^{n} (w^T x_i - y_i)^2 + \lambda \|w\|_2^2$$

مرحله ۱: بازنویسی تابع هزینه

ابتدا می‌توانیم تابع هزینه $L(w)$ را به صورت ماتریسی بازنویسی کنیم. فرض کنیم:

- X ماتریس ویژگی‌ها به ابعاد $n \times d$ باشد که هر سطر آن برابر x_i^T است.
- y بردار پاسخ‌ها باشد که به صورت $y = [y_1, y_2, \ldots, y_n]^T$ است.

در این صورت، قسمت اول تابع هزینه را می‌توان به صورت ماتریسی به شکل زیر نوشت:

$$\sum_{i=1}^{n} (w^T x_i - y_i)^2 = \|Xw - y\|_2^2 = (Xw - y)^T (Xw - y)$$

بنابراین تابع هزینه به شکل زیر بازنویسی می‌شود:

$$L(w) = (Xw - y)^T (Xw - y) + \lambda w^T w$$

مرحله ۲: محاسبه گرادیان نسبت به w

برای یافتن مقدار بهینه w، گرادیان تابع هزینه $L(w)$ را نسبت به w محاسبه کرده و آن را برابر صفر قرار می‌دهیم:

$$\nabla_w L(w) = \nabla_w \left((Xw - y)^T (Xw - y) + \lambda w^T w \right)$$

گرادیان قسمت اول به صورت زیر است:

$$\nabla_w (Xw - y)^T (Xw - y) = 2X^T (Xw - y)$$

و گرادیان قسمت دوم به صورت زیر است:

$$\nabla_w (\lambda w^T w) = 2\lambda w$$

بنابراین، گرادیان کلی تابع هزینه برابر است با:

$$\nabla_w L(w) = 2X^T (Xw - y) + 2\lambda w$$

مرحله ۳: قرار دادن گرادیان برابر صفر

برای بهینه‌سازی، گرادیان را برابر صفر قرار می‌دهیم:

$$2X^T (Xw - y) + 2\lambda w = 0$$

این معادله را می‌توان به صورت زیر ساده کرد:

$$X^T X w - X^T y + \lambda w = 0$$

$$(X^T X + \lambda I) w = X^T y$$

در اینجا I ماتریس همانی به ابعاد $d \times d$ است.
مرحله ۴: حل معادله برای w
اگر $X^T X + \lambda I$ معکوس‌پذیر باشد، می‌توانیم مقدار بهینه w را به دست آوریم:

$$w = (X^T X + \lambda I)^{-1} X^T y$$