



تمرین ششم - بخش اول

شماره دانشجویی: ۹۸۱۰۱۰۷۴

محمدجواد هزاره

سوال ۱

(آ) برای مشتقات جزئی داریم:

$$\begin{aligned}
 \frac{\partial f}{\partial w_j} &= \frac{\partial}{\partial w_j} \left(\sum_i (\hat{y}^i - y^i)^2 \right) \\
 &= \sum_i \frac{\partial}{\partial w_j} (\hat{y}^i - y^i)^2 \\
 &= \sum_i 2(\hat{y}^i - y^i) \frac{\partial}{\partial w_j} \hat{y}^i \\
 &= \sum_i 2(\hat{y}^i - y^i) (x^{(i)})^j
 \end{aligned}$$

بنابراین گرادیان f به صورت زیر خواهد بود:

$$\nabla f = \begin{pmatrix} 2 \sum_i \epsilon_i \\ 2 \sum_i \epsilon_i x^{(i)} \\ 2 \sum_i \epsilon_i (x^{(i)})^2 \end{pmatrix}, \quad \epsilon_i = \hat{y}^i - y^i$$

(ب) اگر داشته باشیم $w = (w_2, w_1, w_0)^T$ ، آنگاه برای Gradient Descent داریم:

$$w^{t+1} = w^t - \alpha \nabla f$$

با کم کردن α سرعت همگرا شدن w کاهش پیدا می‌کند. با زیاد کردن α نیز تا یک حدی، سرعت همگرا شدن زیاد می‌شود اما از این حد به بعد، به کلی همگرایی را از دست می‌دهیم و w^t واگرا می‌شود.