

تمرین ۱-۴

Subject
Date

بسم الله الرحمن الرحيم
98170498

$$f(w_r, w_i, w_o) = \sum_i (w_r x_i^r + w_i x_i + w_o - y_i)^2$$

$$w_o \rightarrow \frac{\partial f}{\partial w_o} = \frac{\partial (\sum_i (w_r x_i^r + w_i x_i + w_o - y_i)^2)}{\partial w_o} = 2 \sum_i (w_r x_i^r + w_i x_i + w_o - y_i)$$

$$w_i \rightarrow \frac{\partial f}{\partial w_i} = 2 \sum_i x_i (w_r x_i^r + w_i x_i + w_o - y_i)$$

$$w_r \rightarrow \frac{\partial f}{\partial w_r} = 2 \sum_i x_i^r (w_r x_i^r + w_i x_i + w_o - y_i)$$

gradient descent

$$w(t+1) = w(t) - \alpha \nabla f$$

$$w(t+1) = \begin{aligned} & \rightarrow w_o(t) - \alpha \sum_i (w_r x_i^r + w_i x_i + w_o - y_i) \\ & \rightarrow w_i(t) - \alpha \sum_i x_i (w_r x_i^r + w_i x_i + w_o - y_i) \\ & \rightarrow w_r(t) - \alpha \sum_i x_i^r (w_r x_i^r + w_i x_i + w_o - y_i) \end{aligned}$$

اگر α خیلی بزرگ باشد، سرعت یادگیری کم می شود و در زمان بهینه رسیدن به حالتی که زمان کمی طول می کشد و اگر α خیلی کوچک باشد، زمان بهینه رسیدن به حالتی که زمان کمی طول می کشد و اگر α خیلی بزرگ باشد، سرعت یادگیری کم می شود و در زمان بهینه رسیدن به حالتی که زمان کمی طول می کشد.