

در حد فرض

تاریخ تست سری اول

سوال 11

$$f(w_r, w_1, w_0) = (w_r x_1^2 + w_1 x_0 + w_0 - y_0)^2 + (w_r x_1^2 + w_1 x_1 + w_0 - y_1)^2 + \dots$$

$$\Rightarrow \nabla_w f(w_r, w_1, w_0) = \begin{bmatrix} \frac{\partial f}{\partial w_0} \\ \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_r} \end{bmatrix} = \begin{bmatrix} 2 \sum_i (\hat{y}^i - y^i) \\ 2 \sum_i (\hat{y}^i - y^i) x_i \\ 2 \sum_i (\hat{y}^i - y^i) (x_i)^2 \end{bmatrix}$$

* توجه: فرض کنیم dataset n داده دارد.

$$* w = \begin{bmatrix} w_0 \\ w_1 \\ w_r \end{bmatrix}$$

ب)

a gradient descent rule

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \nabla_w f(w_r, w_1, w_0)$$

ساده شده

$$\begin{bmatrix} w_0^{(t+1)} \\ w_1^{(t+1)} \\ w_r^{(t+1)} \end{bmatrix} \leftarrow \begin{bmatrix} w_0^{(t)} \\ w_1^{(t)} \\ w_r^{(t)} \end{bmatrix} - \alpha \begin{bmatrix} 2 \sum_i (\hat{y}^i - y^i) \\ 2 \sum_i (\hat{y}^i - y^i) x_i \\ 2 \sum_i (\hat{y}^i - y^i) (x_i)^2 \end{bmatrix}$$

* با افزایش α این امکان وجود دارد که W به $miss$ لغز و گداز شود.

دلیل این موضوع $step\ size$ های بزرگ در هر مرحله از $gradient\ descent$ است.

* اگر هم α را خیلی کوچک در نظر بگیریم، باید به یاد داشته باشیم که $gradient\ descent$ یک فرآیند

تکرار کردن تعداد $iteration$ ها دارد، به عبارتی که این تعداد به سقف خود

برسد ما W بهینه را پیدا کنیم و به سرعت گداز شدن کاهش پیدا کند.

* پس به نیت زیاد بزرگ و زیاد کوچک باشو