Fitting a line — Minh Quoc Nguyen
290130

Call N data points
$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(i)}, y^{(i)}), \dots, (x^{(N)}, y^{(N)})$

Finding a and b for fitting the line
$$y = ax + b$$
to the data points, that means minimizing predefined error.

Mean squared error:
$$\ell(a,b) = \frac{1}{N}\sum_{i=1}^{N}\left(\cancel{x}y^{(i)} - y(x_k^{(i)})\right)^2$$

Taking partial derivatives:
$$\frac{\partial \ell}{\partial a} = \frac{1}{N}\sum_{i=1}^{N} 2\left(y^{(i)} - (ax^{(i)} + b)\right)(-x^{(i)})$$
$$= \frac{2}{N}\sum_{i=1}^{N}\left(ax^{(i)} + b - y^{(i)}\right)x^{(i)}$$
$$\frac{\partial \ell}{\partial b} = \frac{1}{N}\sum_{i=1}^{N} 2\left(y^{(i)} - (ax^{(i)} + b)\right).1$$

~~Symplyfing the~~

Then
$$0 = \frac{\partial \ell}{\partial a} \Rightarrow 0 = \frac{2}{N}\left[a\sum_{i=1}^{N}(x^{(i)})^2 + b\sum_{i=1}^{N}x^{(i)} - \sum_{i=1}^{N}y^{(i)}x^{(i)}\right]$$
$$\Leftrightarrow 0 = a\overline{x^2} + b\overline{x} - \overline{xy} \quad (1)$$
where $\overline{x^2} = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)})^2$, $\overline{x} = \frac{1}{N}\sum_{i=1}^{N}x^{(i)}$, $\overline{xy} = \frac{1}{N}\sum_{i=1}^{N}y^{(i)}x^{(i)}$

$$0 = \frac{\partial \ell}{\partial b} \Rightarrow 0 = \frac{2}{N}\left[\sum_{i=1}^{N}y^{(i)} - a\sum_{i=1}^{N}x^{(i)} + b\sum_{i=1}^{N}1\right]$$
$$\Leftrightarrow 0 = \overline{y} - a\overline{x} + b \quad (2)$$

From (1), (2):
$$\begin{cases} a\overline{x^2} + b\overline{x} = \overline{xy} \\ a\overline{x} + b = \overline{y} \end{cases}$$
(Suppose the system have unique solution $\begin{vmatrix}\overline{x^2} & \overline{x} \\ \overline{x} & 1\end{vmatrix} \neq 0$)

$$\Rightarrow \begin{cases} a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} \\[2ex] b = \dfrac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2} \end{cases}$$

If now each data points have m attributes:
$(x^{(i)}, y^{(i)}) \rightarrow (x_1^{(i)}, x_2^{(i)}, \dots x_m^{(i)}, y^{(i)})$

The predected value becomes:
$$y = w_0 + [w_1, \dots w_m]\begin{bmatrix}x_1 \\ x_2 \\ \vdots \\ x_m\end{bmatrix}$$
$$= w_0 + w^T x.$$

Also taking the derivative: for $1 \le j \le m$.
$$\frac{\partial \ell}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}2\left(y(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$
$$= \frac{2}{N}\sum_{i=1}^{N}\left[w_0 + w^T x^{(i)} - y^{(i)}\right]x_j^{(i)}$$
$$= \frac{2}{N}\sum_{i=1}^{N}\left[w'^T x'^{(i)} - y^{(i)}\right]x_j^{(i)}$$
where $w'^T = \begin{bmatrix}1 \\ w_1 \\ w_2 \\ \vdots \\ w_m\end{bmatrix}$, $x'^{(i)} = \begin{bmatrix}1 \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_m^{(i)}\end{bmatrix}$

$$\frac{\partial \ell}{\partial w_0} = \frac{1}{N}\sum_{i=1}^{N}2\left(y(x^{(i)}) - y^{(i)}\right)1$$
$$= \frac{2}{N}\sum_{i=1}^{N}\left[w'^T x'^{(i)} - y^{(i)}\right]$$

Denote
$$X = \begin{bmatrix}1 & \text{---} & x^{(1)} & \text{---} \\ 1 & \text{---} & x^{(2)} & \text{---} \\ \vdots & & \vdots & \\ 1 & \text{---} & x^{(N)} & \text{---}\end{bmatrix}$$
X size $N \times (m+1)$

$$Y = \begin{bmatrix}y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)}\end{bmatrix} \qquad Y \in \mathbb{R}^N$$

Then

$$\begin{bmatrix} \frac{\partial l}{\partial w_0} \\ \frac{\partial l}{\partial w_1} \\ \vdots \\ \frac{\partial l}{\partial w_m} \end{bmatrix} = X^T(Xw' - Y)$$

$$\Rightarrow \nabla l(w) = 0 \Leftrightarrow X^T(Xw' - Y) = 0$$

$$\Leftrightarrow X^TXw' - X^TY = 0$$

$$\Leftrightarrow w' = (X^TX)^{-1}(X^TY)$$

(Suppose $(X^TX)^{\cancel{-1}}$ is invertible)