10-606: Mathematical Foundations for Machine Learning **Fall 2021**

Recitation 2: Deriving 1D Linear Regression with Offset

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

In this worksheet, we will consider the linear model, $y = mx + b$, where $y, m, x, b \in \mathbb{R}$. In what follows, you will go through the steps to derive the values of m and b that minimize MSE for a given arbitrary dataset, $D = \{(x_i, y_i)\}_{i=1}^{N}$.

Step 1. Write out the MSE, $J(m, b; D)$.

Solution.

$$J(m, b; D) = \frac{1}{N}\sum_{i=1}^{N}(y_i - mx_i - b)^2$$

Step 2. Compute the partial derivatives $\frac{\partial J}{\partial m}(m, b; D)$ and $\frac{\partial J}{\partial b}(m, b; D)$. You may want to first expand $J(m, b; D)$ first (like we did in the non-offset case) to make things easier for you.

Solution.

We note that the term in the sum expanded is:

$$(y_i - mx_i - b)^2 = y_i^2 + m^2x_i^2 + b^2 - 2y_ix_im - 2y_ib + 2mx_ib$$

So the MSE is:

$$\begin{aligned}
J(m, b; D) &= \frac{1}{N}\sum_{i=1}^{N}\left[y_i^2 + m^2x_i^2 + b^2 - 2y_ix_im - 2y_ib + 2mx_ib\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i + \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)m^2 + b^2 - \left(\frac{2}{N}\sum_{i=1}^{N} y_ix_i\right)m - \left(\frac{2}{N}\sum_{i=1}^{N} y_i\right)b + \left(\frac{2}{N}\sum_{i=1}^{N} x_i\right)mb
\end{aligned}$$

From this point it is much easier to compute the partial derivatives.

$$\frac{\partial J}{\partial m}(m, b; D) = \left(\frac{2}{N}\sum_{i=1}^{N} x_i^2\right)m + \left(\frac{2}{N}\sum_{i=1}^{N} x_i\right)b - \frac{2}{N}\sum_{i=1}^{N} y_ix_i$$

$$\frac{\partial J}{\partial b}(m, b; D) = \left(\frac{2}{N}\sum_{i=1}^{N} x_i\right)m + 2b - \frac{2}{N}\sum_{i=1}^{N} y_i$$

Step 3. Find m_0, b_0 such that $\frac{\partial J}{\partial m}(m_0, b_0; D) = 0$ and $\frac{\partial J}{\partial b}(m_0, b_0; D) = 0$. In this case, m_0, b_0 is the global minimizer, i.e. $y = m_0 x + b_0$ is the line that achieves the lowest MSE.

Solution.

Starting with setting the $\frac{\partial J}{\partial b}(m, b; D) = 0$ and solving for b:

$$\frac{\partial J}{\partial b}(m, b; D) = \left(\frac{2}{N}\sum_{i=1}^{N} x_i\right) m + 2b - \frac{2}{N}\sum_{i=1}^{N} y_i = 0$$

$$b = \frac{1}{N}\left(\sum_{i=1}^{N} y_i - m\sum_{i=1}^{N} x_i\right)$$

Substituting this into $\frac{\partial J}{\partial m}(m, b; D)$ and setting to 0 we get:

$$\frac{\partial J}{\partial m}(m, b; D) = \left(\frac{2}{N}\sum_{i=1}^{N} x_i^2\right) m + \left(\frac{2}{N}\sum_{i=1}^{N} x_i\right)\left[\frac{1}{N}\left(\sum_{i=1}^{N} y_i - m\sum_{i=1}^{N} x_i\right)\right] - \frac{2}{N}\sum_{i=1}^{N} y_i x_i = 0$$

$$m\left[\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right] = \sum_{i=1}^{N} x_i y_i - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} yi\right)$$

$$m = \frac{\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} yi\right)}{\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2}$$