## Exercise 1

How do you interpret the weights $w$ and offset $b$ in a least squares regression analysis?

The weights are the coeffcients of the linear regression function. They represent the strenght and direction of features to the target variable. The offset is the base line of the function. For example if the predicted variable is house price and the features are size and distance to city, then a high (absolute) weight for size means that the size of a house is highly important when determining its price. Genearally houses are not free so the bias/offset might be chosen to the minimum house price.

## Exercise 2

We have seen in class that the gradient of the least squares regression function is

$$\nabla L = X^\top (Xw - y)$$

Verify the correctness of this gradient using standard calculus from high school, i.e., no matrix calculus.

Specify Scalar LSE as $f(w) = \frac{1}{2}(x \cdot w - y)^2$

$$\boxed{\begin{aligned}&\text{Chain rule}\\ &f(x) = g(h(x))\\ &f'(x) = h'(x) \cdot g'(h(x))\end{aligned}}$$

$$g(x) = \frac{1}{2} h(x)^2$$

$$g'(x) = 2 \cdot \frac{1}{2} h(x)$$

$$= h(x)$$

$$h(x) = x \cdot w - y$$

$$h'(x) = x$$

$$f'(x) = x \cdot (x \cdot w - y)$$

## Exercise 3

Prove that the system of linear equations for least squares regression always has a solution. How many solutions can it have? Please list all cases that can happen.

System of Linear e q uations: $(X^T X) w = X^T y$

-> has a solution if $X^T X$ is invertible. This is the case when the rank of X (number of lineorly indepent columns/rows) is equal to the number of columns in X.

Unique Solution: $X^T X$ is invertible i.e. there are no linear dependencies amon X. Then there exists only one w which yields the smallest error

Infinite Solutions: $X^T X$ is not inveltible, meaning there are linear relations within X. Thus the weights for those relations can be combined in infinite ways and still return the optimal solution.