Class Exercises. Set 1

1. Consider the dataset = $\begin{bmatrix} 3 & 14 \\ 4 & 20 \\ 6 & 27 \\ 8 & 41 \\ 12 & 63 \\ 15 & 73 \end{bmatrix}$, where the first column represents variable *x* and the second column the variable *y*.

a. Suppose we want to fit a line $f_1(x)= y = ax + b$ to the data by linear regression. Find *a* and *b*.

b. Suppose now that we want to fit $f_2(y) = x = cy + d$, again by linear regression. Find *c* and *d*.

c. Predict the y value for x=5 using both $f_1(x)$ and $f_2(y)$.

d. What are the main differences between both models in terms of assumptions (= when should you prefer one over the other)?

2. Implement in Python a function `[weights] = regression(train data X, train data Y)` that outputs the coefficients of the linear regression on the data. Test it to compute the polynomial regression of degree 9 with the data set `Sin(x)` from PRMLBishop. Obtain figure 1.4.d from PRMLBishop.

3. In linear regression, we are given a training set $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$.

Assuming a parametric model of the form: $y_i = w^t x_i + \epsilon_i$, where ϵ_i are noise terms from a given distribution, linear regression seeks to find the parameter vector w that provides the best of fit of the above regression model.

Assuming that ϵ_i are Independent and Identically Distributed and sampled from the same zero mean Gaussian $N(0,\sigma^2)$ distribution we have verified that the maximum likelihood estimate (MLE) for p(y|X) is also the least square estimate minimizing

$J_1(w) = 0.5 \sum_{i=1}^{N}(y_i - w^t x_i)^2$

Assume now that ϵ_i are independent and identically distributed according to a Laplace distribution with zero mean. That is each ϵ_i follows Laplace(0; b) = $\frac{1}{2b}e^{-\frac{|\epsilon_i|}{b}}$.

a. Write down the formula for calculating the MLE of w.
b. Provide the loss function $J_2(w)$ whose minimization is equivalent to finding the MLE of w under the above noise model.
c. What is the advantage of this model compared to the standard Gaussian assumption?