Exercise 4.1 - Maximum Likelihood Estimate (MLE) (1+1 points)

a) Show how a linear regression procedure can be justified as an MLE procedure, assuming that mean squared error is used as a metric. Recall that

$$MSE = \frac{1}{m}\sum_{i=1}^{m} \left\|\hat{y}^{(i)} - y^{(i)}\right\|^2 .$$

Justify and motivate the assumptions you make along the way. This particular deduction is not covered in the lecture. Consult the book to gain further understanding.

b) Given an i.i.d. sample $X_1, \ldots, X_n$ from a Poisson distribution with parameter λ, find the MLE of the parameter λ. Recall that

$$\Pr(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} .$$

a) If we are given a probabilistic model of the data of the form

$$p(y|x,\theta) = \mathcal{N}(w^T x, \sigma^2)$$

where $\theta = (w, \sigma^2)$ and we assume σ^2 to be fixed, and $w \in \mathbb{R}^{d+1}$ where d is the dimension of the data vectors $x \in \mathbb{R}^d$ and w contains an extra entry simulating the bias of the model

We can write the MLE estimator of θ as

$$\hat{\theta} = (\hat{w}, \hat{\sigma}^2) = \text{argmin} - \sum_{i=1}^{m} \log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{1}{2\sigma^2}(y_i - w^T x)^2\right)\right]$$

$$= \text{argmin} \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - w^T x)^2 + \frac{m}{2}\cdot\log(2\pi\sigma^2)$$

$$= \text{argmin} \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - w^T x)^2$$

which implies that a weight that minimizes the MSE also minimizes the MLE.

b) We seek to estimate λ via the maximum-likelihood method.

$$\lambda_{mle} = \operatorname{argmin} - \log P(D|\lambda)$$

where $D = \{X_1, ..., X_n\}$ is our given data. We have

$$\begin{aligned}
\underset{\lambda \in (0,\infty)}{\operatorname{argmin}} -\log P(D|\lambda) &= \underset{\lambda \in (0,\infty)}{\operatorname{argmin}} - \sum_{i=1}^{n} \log \frac{\lambda^{x_i} e^{-\lambda}}{x!} \\
&= \underset{\lambda \in (0,\infty)}{\operatorname{argmin}} - \sum_{i=1}^{n} \log (\lambda^{x_i} e^{-\lambda}) - \log (x_i!) \\
&= \underset{\lambda \in (0,\infty)}{\operatorname{argmin}} - \sum_{i=1}^{n} \log (\lambda^{x_i} e^{-\lambda}) \\
&= \underset{\lambda \in (0,\infty)}{\operatorname{argmin}} - \sum_{i=1}^{n} \left[\log (\lambda^{x_i}) - \lambda\right] \\
&= \underset{\lambda \in (0,\infty)}{\operatorname{argmin}} \underbrace{- \sum_{i=1}^{n} \left[\log (\lambda) x_i - \lambda\right]}_{:= f(\lambda)}
\end{aligned}$$

Setting the derivative of $f(\lambda)$ to zero yields

$$- \sum_{i=1}^{n} \left[\frac{1}{\lambda} x_i - 1\right] = 0 \iff \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Thus

$$\lambda_{mle} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$\blacksquare$