1 Exercise 1

We consider a finite set of data points $\{(y_0(x_0), x_0), \ldots, (y_{n-1}(x_{n-1}), x_{n-1})$ and the following two vectors:

$$x^\top = (x_0 \ \ldots \ x_{n-1})$$
$$y^\top = (y_0, \ \ldots \ y_{n-1})$$

Furtheremore we assume that there exist a continious function f and a normal distributed error $\epsilon_i \propto N(0, \sigma^2)$ sutch that $\forall i$

$$y(x_i) = f(x_i) + \epsilon_i \tag{1}$$

We now approximate $f(x_i)$ with a polynomial

$$f(x_i) \approx \sum_{j=0}^{p} (x_i)^j \beta_j := \tilde{y} \tag{2}$$

s.t

$$y(x_i) = \tilde{y}(x_i) + \epsilon_i + \zeta_i \tag{3}$$

were ζ_i is the error of the approximation. This also can be written in a vectorial form

$$y(x) = A(x)\beta + \epsilon + \zeta \qquad \text{with} \quad a_{ij} = (y_i)^j \tag{4}$$

were $\epsilon \in \mathbb{R}^n$, $\beta \in \mathbb{R}^p$ with $p \in \mathbb{N}$ and $A(x) \in \mathbb{R}^{n \times p}$ is called the feature Matrix. In the following we assume that we have only one error $\epsilon_i \propto N(0, \sigma^2)$

$$y = \tilde{y} + \epsilon \tag{5}$$

We now define

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2 = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \sum_{j=0}^{p-1} A_i^{\ j} \beta_j) \tag{6}$$

to dertermine the quality of our approximation as a function of β_i. The optimal fit is given by the minima of that function:

$$\partial_{\beta_j}(C(\beta))\, e_j = \frac{2}{n}(y^i - A_i^{\ k}\beta_k) A_i^{\ j} e_j \tag{7}$$

were e_j correspond to the eukledian basis vectors and the Einstein summetion rule is used. Again one can express this in Matrix form:

$$\begin{pmatrix} \partial_{\beta_0} C(\beta) \\ \vdots \\ \partial_{\beta_{p-1}} C(\beta) \end{pmatrix} = \begin{pmatrix} (y_0 - A_0^{\ k}\beta_k) A_0^{\ 0} + \cdots + (y_{n-1} - A_{n-1}^{\ \ k}\beta_k) A_{n-1}^{\ \ 0} \\ \vdots \\ (y_0 - A_0^{\ k}\beta_k) A_0^{\ p-1} + \cdots + (y_{n-1} - A_{n-1}^{\ \ k}\beta_k) A_{n-1}^{\ \ p-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{8}$$

By defininining the following vector

$$a_i = A_k{}^i e^k \tag{9}$$

this simplyfies to

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} (a_0, y) - (a_0, A\beta) \\ \vdots \\ (a_{p-1}, y) - (a_{p-1}, A\beta) \end{pmatrix} = A^\top (y - A\beta) \tag{10}$$

were $(\cdot, \cdot)$ corresponds to the standard scalarproduct. The parameters of the approximation with the smallest error ϵ are thus given by

$$\tilde{\beta} = (A^\top A)^{-1} A^\top y \tag{11}$$

We now look at some statistical quantities. For that note that the following two relations with X_1 and X_2 beeing two random variables and $\lambda \in \mathbb{R}^{n \times m}$ a constant hold:

$$\mathbb{E}[\lambda \cdot (X_1 + X_2)] = \lambda \cdot (\mathbb{E}[X_1] + \mathbb{E}[X_2]) \tag{12}$$

$$\mathrm{var}[\lambda \cdot X_1] = \lambda \lambda^\top \cdot \mathrm{var}[X_1] \tag{13}$$

With this we get the following expectation values and variances for y_i

$$\mathbb{E}(y_i) = \mathbb{E}[A_i^{\ j} \beta_j + \epsilon_i] = \mathbb{E}[A_i^{\ j} \beta_j] + \mathbb{E}[\epsilon_i] = A_i^{\ j} \beta_j \tag{14}$$

$$\mathrm{var}(y_i) = \mathbb{E}[(y_i - \mathbb{E}(y_i))^2] = \mathbb{E}[(A_i^{\ j} \beta_j + \epsilon_i - A_i^{\ j} \beta_j)^2] = \mathbb{E}[\epsilon_i^2] = \mathrm{var}(\epsilon_i) = \sigma^2 \tag{15}$$

Therefore $y \propto N(A_i^{\ j} \beta_j, \sigma^2)$ and thus y follows a normal distribution $N(A\beta, \sigma^2)$. The expectation value and variance of the ideal parameters $\tilde{\beta}$ are then given by

$$\mathbb{R}[\tilde{\beta}] = \mathbb{E}[(A^\top A)^{-1} A^\top y] = (A^\top A)^{-1} A^\top \mathbb{E}[y] = (A^\top A)^{-1} (A^\top A)\, \beta = \beta \tag{16}$$

$$\mathrm{var}[\tilde{\beta}] = \mathrm{var}[(A^\top A)^{-1} A^\top y] = (A^\top A)^{-1} A^\top \big((A^\top A)^{-1} A^\top\big)^\top \mathrm{var}[y^2] \tag{17}$$

$$= (A^\top A)^{-1} A^\top A \big((A^\top A)^\top\big)^{-1} \mathrm{var}[y^2] \tag{18}$$

$$= (A^\top A)^{-1} \mathrm{var}[y^2] = \sigma^2 (A^\top A)^{-1} \tag{19}$$