Inleveropgaven maandag 22 februari

Opgave 0.1. Voor de functie $f(x) = \sqrt{\|x\|}$ op het interval $[-1, 1]$.

a) Vind het tweedegraads polynoom $p_2(x) = a_0 + a_1 x + a_2 x^2$ dat de functie f het best benadert op het interval $[-1, 1]$.

b) Bereken $\|p_2\|^2$ en $\|f\|^2$.

c) Wat zegt de Stelling van Pythagoras over $\|p_2\|$, $\|f - p_2\|$ en $\|f\|$? Bereken de afstand $\Delta = \|f - p_2\|$ tussen f en het tweedegraads polynoom dat f het best benadert.

Opgave 0.2 (Lineaire Regressie). We zoeken de functie $f(x) = \alpha + \beta x$ die de meetwaarden $(x_0, y_0), \ldots (x_{N-1}, y_{N-1})$ het best benadert, in de zin dat de som van de kwadraten van de gemaakte fout minimaal is. We zoeken dus α en β zo dat

$$\Delta^2 := \sum_{i=0}^{N-1} (f(x_i) - y_i)^2. \tag{1}$$

minimaal is. Definiëer de vectoren $\vec{v}$, $\vec{b}_0$ en $\vec{b}_1$ in $\mathbb{R}^N$ door

$$\vec{v} = (y_0, \ldots, y_{N-1}) \quad \vec{b}_0 = (1, \ldots, 1) \quad \vec{b}_1 = (x_0, \ldots, x_{N-1})\,.$$

a) Laat zien dat $\Delta = \|\alpha\vec{b}_0 + \beta\vec{b}_1 - \vec{v}\|$. Deze uitdrukking wordt dus geminimaliseerd als $\alpha\vec{b}_0 + \beta\vec{b}_1$ de projectie $P(\vec{v})$ van $\vec{v}$ op de deelruimte opgespannen door $\vec{b}_0$ en $\vec{b}_1$ is.

b) Orthogonaliseer $\vec{b}_0, \vec{b}_1$ tot $\vec{b}'_0, \vec{b}'_1$.

c) Wat zijn de basiscoëfficiënten λ_0 en λ_1 in $P(\vec{v}) = \lambda_0\vec{b}'_0 + \lambda_1\vec{b}'_1$? Druk je antwoord uit in de gemiddelden

$$\overline{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i, \quad \overline{y} = \frac{1}{N}\sum_{i=0}^{N-1} y_i$$

en de (co)variantie

$$\mathrm{Cov}(x, y) = \frac{1}{N}\sum_{i=0}^{N-1} (x_i - \overline{x})(y_i - \overline{y}), \quad \mathrm{Var}(x) = \frac{1}{N}\sum_{i=0}^{N-1} (x_i - \overline{x})^2\,.$$

d) Druk de projectie uit in de oorspronkelijke basis, en bepaal α en β in de uitdrukking $P(\vec{v}) = \alpha\vec{b}_0 + \beta\vec{b}_1$.

Het resultaat is wellicht de moeite om te onthouden:

Propositie 0.1. *De waarden van α en β waarvoor de functie $f = \alpha + \beta x$ de meetpunten $(x_0, y_0), \ldots (x_{N-1}, y_{N-1})$ het best benadert in de zin dat de kwadratische fout (1) minimaal is, zijn*

$$\beta = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}, \quad \alpha = \overline{y} - \beta\overline{x}\,.$$